ULPGC
Universidad de
Las Palmas de
Gran Canaria

Enabling Rolling Shutter Optical Camera Communication Using Artificial Intelligence:

Towards Widespread Adoption and Dual Use of Cameras as Receivers

Cristo Manuel Jurado Verdú

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy (Ph.D.)

in the doctorate program
EmITIC - Empresa, Internet y Tecnologías de las Comunicaciones

Supervised by José Alberto Rabadan Borges, and
co-supervised by Victor Guerra Yánez

Institute for Technological Development and Innovation in Communications, IDeTIC
Universidad de Las Palmas de Gran Canaria, ULPGC

Las Palmas de Gran Canaria, Canary Islands, Spain
Date: January 19, 2023

ULPGC
Universidad de
Las Palmas de
Gran Canaria
Escuela de
Doctorado

D José Alberto Rabadán Borges COORDINADOR DEL PROGRAMA DE DOCTORADO Empresa, Internet y Tecnologías de las Comunicaciones (EmITIC) DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

INFORMA,

De que la Comisión Académica del Programa de Doctorado, en su sesión de fecha .... tomó el acuerdo de dar el consentimiento para su tramitación, a la tesis doctoral titulada **"Enabling Rolling Shutter Optical Camera Communication using Artificial Intelligence: Towards Widespread Adoption and Dual Use of Cameras as Receivers"** presentada por el doctorando **D. Cristo M. Jurado Verdú** y dirigida por el **Doctor José Alberto Rabadán Borges** y codirigida por el **Doctor Victor Guerra Yánez**.

Y para que así conste, y a efectos de lo previsto en el Art.° 11 del Reglamento de Estudios de Doctorado (BOULPGC 04/03/2019) de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a...de..de dos mil...........

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
ESCUELA DE DOCTORADO

Programa de doctorado **Empresa, Internet y Tecnologías de las Comunicaciones (EmITIC)**

Título de la Tesis **Enabling Rolling Shutter Optical Camera Communication using Artificial Intelligence: Towards Widespread Adoption and Dual Use of Cameras as Receivers**

Tesis Doctoral presentada por **D. Cristo M. Jurado Verdú**

Dirigida por el **Dr. José Alberto Rabadán Borges**
Codirigida por el **Dr. Victor Guerra Yánez**

Las Palmas de Gran Canaria, a de de 20

El Director, El Codirector, El Doctorando,

(firma) (firma) (firma)

Abstract

Today's society relies on a radio-based wireless communications infrastructure that is reaching high saturation levels, in which more and more devices require high-capacity, low-latency wireless connections. To solve this problem, optical wireless communication (OWC) is positioned as an ideal candidate to provide a complementary infrastructure capable of meeting the capacity requirements of future networks. There are many branches within OWC, including free-space optical communication (FSO) based on laser transmitters, which already have successful commercial implementations. For example, this technology is used to establish intra-satellite links in the StarLink satellite internet constellation.

However, this thesis focuses on the visible light communication (VLC) branch. This technology aims to reuse the lighting systems present in offices, homes, hospitals, airports, etc., to establish new communication links, which can be used for various applications: precise indoor positioning, signage and marketing, backbone support for augmented reality, Internet access points, vehicular communications, smart cities' management, and environmental monitoring, among many others. However, although this technology has reached the readiness level required for its deployment, its penetration in markets and industry has been slowed by the lack of commercially available chipsets (based on photodiodes) and standards with a clear industrial view. Therefore, the costs derived from embedding this hardware in end-user devices have a negative impact on the adoption of this technology.

For this reason, the scientific and engineering community has shown interest in using cameras as receivers instead of custom photodiode arrangements. A new VLC branch is thus established, known as optical camera communication (OCC). This technology promotes the massive creation of applications using end-user devices with built-in cameras, such as smartphones, laptops, dashcams, biometric security systems, etc.

Nevertheless, OCC links have a relatively low throughput, which is inherently limited by the camera's frame rate (i.e., 60 to 120 frames per second). Specifically, global shutter cameras, i.e., those that expose all their pixels simultaneously during capture, sample the optical signal each time an image is taken. Therefore the achievable receiver's sampling rate coincides with the camera's fps, limiting the maximum achievable data rate according to the Nyquist-Shannon theorem.

On the other hand, rolling shutter (RS) cameras achieve significantly higher transmission rates than the previous ones. This is because they scan the scene sequentially row by row of pixels, sampling light variations that occur during a single capture. These variations are perceived in the image as bands of different intensities according to the illumination level present at the instant the corresponding row was exposed. Those bands correspond to the different symbols of the optical signal. In summary, signal sampling is performed row by row of pixels, so the receiver's sampling rate depends on the rate at which the rows are activated. The interval between the activation of

two consecutive rows is called the row sampling time, and it is between 100 and 10000 times shorter than the capture time. Typical row sampling times range from 8 to 10 microseconds, leading to receiver's sampling rates up to 125 kHz. In addition, super slow motion cameras can exceed sampling rates greater than 500kHz. For this reason, and the fact that most cameras on the market use the RS acquisition mechanism, these links have recently received special attention.

Although RS-OCC systems allow higher transmission rates, they require that the camera's exposure time, i.e., the time the pixels are exposed to light, be as short as possible. Otherwise, as exposure time increases, the symbol bands begin to blend together in the image because the pixels begin accumulating the irradiance from several consecutive symbols. The result of prolonged exposures is the appearance of intersymbol interference that severely impairs signal decoding. In other words, the exposure time behaves like a low-pass filter that considerably reduces the reception bandwidth. A conflict then arises with the usual operation of cameras as imaging devices. If the exposure time is reduced, the camera's sensitivity decreases, and the images it delivers are practically dark, in which the objects present in the scene are no longer clearly discernible. The exposure time sets a trade-off between the camera's sensitivity, and thus its ability to capture the scene, and the receiver bandwidth, and thus its ability to operate as an optical receiver.

Preventing cameras from losing their intrinsic functionality as imaging devices is necessary for this technology to be massively adopted. Therefore, strategies must be sought that favor increased sensitivities without compromising communications. On the other hand, developing a core solution compatible with the vast diversity of cameras on the market is required. Although most of them are RS, each one has a different internal configuration. Moreover, in some cases, their internal parameters are inaccessible, such as the clock frequency, which is related to the row sampling time (and thus to the receiver's sampling frequency), simply because it has no practical application in photography. In other cases, its parameters are dynamically adjusted, as is the case with the exposure time depending on the ambient illumination.

In conclusion, the massive adoption of OCC links depends on knowing the influence that camera parameters have on communications, with particular emphasis on the exposure time, and developing a unique solution that favors the visualization of the scene and therefore allows the effective reuse of as many cameras as possible. To achieve this, the use of artificial intelligence (AI) is proposed in this thesis.

The search for this solution is condensed in this thesis' first general objective (GO1): to develop an AI-assisted architecture that is compatible with many cameras and operates independently of the exposure time. At the same time, it has to support the use of moderate levels for the camera's sensitivity, preventing it from losing its functionality as an imaging device. In addition, two other general objectives are stated: (GO2) to develop an experimental test bench for the generation of real samples and (GO3) to evaluate the integration and feasibility of OCC in industrial and market applications.

The main core of this thesis is constituted by three contributions indexed in high-impact journals, which follow an evolution coherent with the iterative development to address this industrial technical challenge.

This thesis is based on previous work done by the author as his master's thesis. In that work, a functional prototype of an RS-OCC link was developed under laboratory conditions. The thesis began when it was decided to deploy it in an industrial environment, specifically in a microalgae cultivation plant. The objective of this project was to create an optical link using a surveillance camera for the simultaneous monitoring and surveillance of multiple photobioreactors. These photobioreactors (hereafter nodes) use an LED panel to simultaneously (i) adjust the illumination intelligently according to the light requirements of the microalgae and (ii) send optical codes with culture parameters, such as temperature, and pH level, among others. This work highlighted that selecting the exposure time is crucial in designing RS links. If it was reduced to provide faster communications, then the camera lost its function as a monitoring device. Laboratory technicians or possible intruders cannot be discerned in the image. This work also analyzed other technical requirements to be considered when deploying multiple links, such as (i) the influence of the orientation of the nodes and their irradiance profile on the link quality, (ii) the distribution of the nodes to achieve an equivalent shared capacity between them, and (iii) their placement to optimize the available space. The results and findings of this work constitute the first contribution of this thesis and provide practical technical guidance for the design and deployment of RS links in an industrial environment.

After highlighting the effects of prolonged exposures, an extensive literature review was conducted in search of solutions to this problem. A lack of research in this aspect was detected. Moreover, the growing interest in the use of AI for source detection and digital signal processing tasks suggested the idea of using AI for the equalization of the effects of long exposures. This marks the beginning of the efforts to increase the exposure time without affecting the performance of the reception routines. This work results in an AI-assisted equalization block capable of mitigating the effects of exposure in mid to severe noise conditions. This equalizer based on a convolutional autoencoder (CAE) allows increasing the exposure time up to 7 times compared to the ideal exposure time for decoding, with bit error rates lower than the forward error correction (FEC) limit. In other words, this equalizer improves the receiver bandwidth by up to 14 times compared to non-equalized links.

It is important to highlight that the results of this work have a potential impact in the field of digital signal processing and, therefore, in all communications systems, whether they are optical, radio-based, or acoustic, since they validate the ability of AI to increase the effective bandwidth in band-limited systems. In this case, the limitation is imposed by the exposure time, but it could come from any other link element in the transmission, channel, or reception.

Another contribution of this work is that the equalizer training was performed exclusively using synthetic samples. For this purpose, the RS mechanism was modeled, and an efficient algorithm for synthetic sample generation was developed. This algorithm uses only the link's temporal parameters, such as transmitter symbol rate and the camera's exposure and row sampling times. This algorithm was indirectly validated by evaluating the equalizer using real-world samples.

This synthetic training was motivated by the difficulty, complexity, and time demands of capturing real images. For this purpose, the cameras must be rigorously characterized and the link conditions adjusted accordingly. This implies that the equalizer might not generalize well and operate correctly with images from other cameras taken under different conditions. Therefore, synthetic training not only speeds up the creation of datasets and eliminates the complexity of developing a test bench but also allows abstracting the training from the experiment conditions by using many representative samples of a wide variety of transmitters, cameras, configurations, and noise levels.

Finally, when commercial smartphone cameras were proposed for practical implementation, it was discovered that many of their internal parameters are unknown, cannot be modified, or are dynamically updated. Therefore, in these cases, it was necessary to estimate these parameters directly at reception from the captured images before decoding. For this purpose, convolutional networks trained with synthetic images representative of thousands of possible configurations are proposed. The results reveal that these networks can obtain relative errors lower than 1.3% and 3% in estimating the data clock frequency and exposure time, respectively. These error values guarantee the optimal operation of the rest of the reception routines.

As a result, this estimator decouples the reception routines from the cameras used and allows to devise a reception architecture that operates exclusively with the images independently of the camera used and its configuration. In addition, this architecture includes exposure equalization stages to allow increasing the exposure time to optimal visibility conditions. In this way, the simultaneous use of the cameras as imaging devices and receivers is enabled. Ultimately, this proposal paves the way for the creation of transferable communications software for RS-OCC that can be easily integrated into developing multiple applications for smartphones, desktops, laptops, tablets, autonomous driving systems, ATMs, etc. In addition, it also enables the development of application interfaces hosted in cloud infrastructures that can process images provided by different multimedia streams in real-time. In this way, it is also possible to technologically enable devices that do not have an open ecosystem to develop their own applications.

Additionally, in a later work, the training times of the equalizers were improved up to 435 times, using only 250 training images instead of 35500. This was achieved by transferring the pre-trained estimator's knowledge to the equalizers. This improvement in the efficiency of training the equalizers allows them to be trained on demand without disrupting communications' link.

Finally, during the exercise of this thesis, other contributions have been made regarding the transmitter part. A transmitter device called Barcolit, which operates as a traditional barcode but actively using LED panels, was proposed. This system is compatible with conventional barcode readers. Thus it does not disrupt current industrial processes. Moreover, it reduces the detection and decoding complexity of traditional barcodes. The codes generated with Barcolits are not distorted in the images and appear correctly aligned horizontally regardless of the camera's relative orientation.

In conclusion, this thesis balances efforts to address (i) an industrial challenge in the particular field of rolling shutter (RS)-optical camera communication (OCC) links, which is to achieve the effective reuse of cameras as receivers without losing their primary function as imaging devices, with (ii) a scientific-technical problem extensible to any field of communications, which is the equalization of the intersymbol interference (ISI) produced by the reduction of the available receiver bandwidth and the indirect estimation of specific link parameters. Moreover, it evolves and matures in line with the most recent advances in the field of artificial intelligence (AI) and computer vision, recording results that exceed those obtained with classical algorithms and other preliminary work. Finally, it focuses on decoupling the reception routines in RS-OCC links from the cameras used, aiming to accelerate this technology's mass adoption.

Resumen en Español

La sociedad actual depende de una infraestructura de comunicaciones inalámbrica basada en radio que está alcanzando niveles de saturación constatables. Son cada vez más los dispositivos que requieren de conexiones inalámbricas de alta capacidad y baja latencia. Para solventar esta problemática, las comunicaciones ópticas inalámbricas (OWC por sus siglas en inglés, optical wireless communication) se posicionan como candidatas idóneas para ofrecer una infraestructura complementaria capaz de asumir los requerimientos de capacidad de los enlaces del futuro. Existen muchas ramas dentro de las OWC, entre las que destaca las comunicaciones ópticas de espacio libre (FSO por sus siglas en inglés, free space optical communication) que utilizan transmisores ópticos basados en láseres, que cuentan ya con implementaciones comerciales exitosas. Por ejemplo, los enlaces FSO se utilizan para las comunicaciones intrasatelitales en StarLink.

Por su parte, esta tesis se centra en las comunicaciones ópticas basadas en luz visible (VLC por sus siglas en inglés, visible light communication). Esta tecnología tiene como objetivo reutilizar los sistemas de iluminación presentes en oficinas, casas, hospitales, aeropuertos, etcétera, para establecer nuevos enlaces de comunicaciones, con diversas aplicaciones: el posicionamiento preciso en interiores, la señalización y el marketing, el soporte de entornos de realidad aumentada, la creación de puntos de acceso a Internet, las comunicaciones vehiculares, la gestión de ciudades inteligentes, la monitorización ambiental, entre otras muchas. Sin embargo, a pesar de que esta tecnología ha alcanzado el nivel de madurez técnico necesario para su despliegue, su penetración en los mercados y la industria se ha visto ralentizada por la necesidad de utilizar un hardware de recepción específico basado en fotodiodos. Los costes derivados de la integración en los dispositivos finales de usuario repercuten negativamente en la adopción de esta tecnología.

Por este motivo la comunidad científica e ingenieril ha mostrado interés en utilizar en lugar de arreglos específicos de fotodiodos, las cámaras como receptores, constituyéndose así una nueva rama conocida como las comunicaciones ópticas basadas en cámara (OCC, por sus siglas en inglés, optical camera communication). De esta forma se promueve la creación masiva de aplicaciones utilizando dispositivos finales de usuario con cámaras incorporadas como los teléfonos móviles, los laptops, las dashcams, los sistemas de seguridad biométricos, entre otros.

No obstante, los enlaces OCC tienen una tasa de transmisión relativamente baja, la cual está intrínsecamente limitada por la tasa de captura de imágenes de la cámara, con tasas típicas de 60 a 120 imágenes por segundo (fps por sus siglas en inglés, frames per second). Concretamente, las cámaras que utilizan una adquisición global shutter (GS), esto es, aquellas que exponen todos sus píxeles simultáneamente durante la captura, muestrean la señal óptica cada vez que se toma una imagen, por lo que la frecuencia de muestreo de recepción coincide directamente con los fps, lo que limita la tasa de datos

de acuerdo con el teorema de muestreo de Nyquist-Shannon.

Por otro lado, las cámaras rolling shutter (RS) alcanzan tasas de transmisión significativamente mayores respecto a las anteriores. Esto es debido a que escanean la escena de forma secuencial fila por fila de píxeles, lo que permite muestrear variaciones en la iluminación que ocurren durante la captura. Estas variaciones se perciben en la imagen como bandas de diferente intensidad de acuerdo con el nivel de iluminación presente en el instante en el que se expuso la correspondiente fila de píxeles. Cuando se utiliza un panel LED como transmisor, este genera bandas en la imagen que se corresponden con los diferentes símbolos que constituyen la señal óptica. En definitiva, el muestreo se realiza fila por fila de píxeles, por lo que la frecuencia de muestreo de recepción dependerá de la velocidad a la que se activan las filas. El intervalo entre la activación de dos filas consecutivas se denomina tiempo de muestreo de fila, y es entre 100 y 10000 veces mayor que el tiempo de captura de una imagen. A modo de ejemplo, tiempos de muestreo de fila típicos que varían entre los 8 y los 10 microsegundos, suponen frecuencias de muestreo que varían entre los 125 kHz y los 100 kHz. Además, las cámaras de súper cámara lenta pueden alcanzar y superar frecuencias de muestreo por encima de los 500 kHz. Por este motivo, y por el hecho de que la mayoría de las cámaras presentes en el mercado utilizan el mecanismo de adquisición RS es por el que estos enlaces han recibido especial atención recientemente.

A pesar de que los sistemas RS-OCC permiten alcanzar tasas de transmisión mayores, requieren, sin embargo, que el tiempo de exposición de la cámara, esto es, el tiempo que permanecen los píxeles expuestos a la luz, sea lo más corto posible. De lo contrario, a medida que el tiempo de exposición aumenta las bandas de símbolos comienzan a mezclarse en la imagen, ya que los píxeles acumulan la irradiancia de varios símbolos consecutivos. El resultado de exposiciones prolongadas es la aparición de una interferencia intersimbólica (ISI, del ingles intersymbol interference) que perjudica gravemente la decodificación de la señal. En otros términos, el tiempo de exposición se comporta como un filtro paso bajo que reduce considerablemente el ancho de banda de recepción. Surge entonces un conflicto con la operación habitual de las cámaras como dispositivos de visualización. Si se reduce el tiempo de exposición, la sensibilidad de la cámara disminuye y las imágenes que esta entrega son prácticamente oscuras, en las que los objetos presentes dejan de percibirse claramente. El tiempo de exposición juega entonces un papel crucial en la mejora de la sensilibidad de la cámara, y por ende su capacidad para visualizar la escena, y en la reducción del ancho de banda del recepción, y por ende su capacidad para operar como receptor óptico.

Evitar que las cámaras pierdan su funcionalidad intrínseca es necesario si se desea que esta tecnología se adopte masivamente. Por tanto, hay que buscar estrategias que favorezcan el aumento de la sensibilidad sin que esto repercuta en un daño para las comunicaciones. Por otro lado, es requisito desarrollar una solución única que sea compatible con la gran diversidad de cámaras que existen en el mercado. A pesar de que la mayoría de ellas son RS, cada una tiene una configuración interna diferente. Además, en algunos casos sus parámetros internos son inaccesibles, como la frecuencia de reloj, que está relacionada con el tiempo de muestreo de fila simplemente porque carece de aplicación práctica en la fotografía. En otros casos, se ajustan dinámicamente su parámetros, como es el caso del tiempo de exposición, que se selecciona automáticamente en función de la iluminación ambiente.

En conclusión, la adopción masiva de los enlaces OCC pasa por conocer la influencia que los parámetros de las cámaras tienen en las comunicaciones, haciendo especial hincapié en el tiempo de exposición, y desarrollar una solución única que favorezca la visualización de la escena, y que, por tanto, permita el reuso efectivo del mayor número de cámaras posible. Para conseguir esto en esta tesis se propone el uso de la inteligencia artificial (AI por sus siglas en inglés).

La búsqueda de esta solución se condensa en el primer objetivo general (GO1) de esta tesis: desarrollar una arquitectura asistida con AI que sea compatible con un gran número de cámaras, y que opere de forma independiente al tiempo de exposición. Al mismo tiempo ha de favorecer el uso de sensibilidades altas para la cámara, evitando así, que esta pierda su funcionalidad como dispositivo de visualización. Además, se enuncian otros dos objetivos generales: (GO2) desarrollar un banco de pruebas experimental para la generación de muestras reales y (GO3) la evaluación de la integración y viabilidad de OCC en las aplicaciones industriales y de mercado.

El núcleo principal de esta tesis es un compendio de tres publicaciones en revistas de alto impacto las cuales siguen una evolución coherente con el desarrollo iterativo llevado a cabo para abordar este desafío de carácter técnico industrial.

Esta tesis parte de un trabajo previo realizado por el autor como trabajo de fin de máster. En este trabajo se desarrolló un prototipo funcional de enlace RS-OCC y se evaluó en condiciones de laboratorio. La tesis comienza en el momento en el que se decide desplegarlo en un entorno industrial, concretamente en una planta de cultivo de microalgas. El objetivo de este proyecto era crear un enlace óptico utilizando una cámara de vigilancia para la monitorización y supervisión simultáneas de múltiples fotobiorreactores. Estos fotobiorreactores (en adelante nodos) utilizan un panel LED para simultáneamente (i) ajustar la iluminación inteligentemente acorde los requerimientos de luz de las microalgas, y (ii) enviar códigos ópticos con parámetros del cultivo, como la temperatura, el nivel de pH, entre otros. Este trabajo trajo a la luz el hecho de que el ajuste del tiempo de exposición es crucial en el diseño de este tipo de enlaces. Si se reducía para favorecer una comunicación más rápida, entonces, la cámara perdía su función como dispositivo de vigilancia. Por consiguiente, los técnicos del laboratorio, o posibles intrusos, no se pueden discernir en la imagen. Este trabajo, también analizó otros requerimientos técnicos a considerar a la hora de desplegar múltiples enlaces, como: (i) la influencia de la orientación de los nodos y su perfil de irradiancia en la calidad del enlace, (ii) su distribución para conseguir tasas de transmisión equivalentes entre ellos, y (iii) su colocación para optimizar el espacio disponible. Los resultados y descubrimientos de este trabajo se recogen el primer artículo de compendio de esta tesis, y ofrecen una guía técnica útil para el diseño y despliegue de enlaces RS en un entorno industrial.

Tras resaltar los efectos de las exposiciones prolongadas se realizó una extensa revisión bibliográfica en la búsqueda de soluciones para este problema. Se detectó una carencia respecto a la investigación en este aspecto, y se observó un creciente interés por el uso de la AI para de detección de fuentes en la imagen, y en otras ramas en el campo del procesamiento digital de la señal. Esto sugirió la idea de utilizar AI para la ecualización de los efectos de las exposiciones prolongadas. El fruto del trabajo realizado con este objetivo se publicó en el segundo artículo de compendio y consiste en un bloque de ecualización asistido con AI, capaz de mitigar de los efectos de la exposición en condiciones de ruido moderado/alto. Este ecualizador basado en un autoencoder convolucional (CAE por sus siglas en inglés, convolutional autoencoder) permite aumentar el tiempo de exposición hasta 7 veces más en comparación con el tiempo de exposición ideal para la decodificación, con tasas de error de bit inferiores al límite impuesto por las técnicas de corrección de errores hacia adelante (FEC por sus siglas en inglés forward error correction). En otras palabras, este ecualizador mejora el ancho de banda de recepción hasta 14 veces comparado con enlaces no ecualizados.

Es importante destacar que los resultados de este trabajo tienen un impacto potencial en el campo del procesado digital de las señales, y, por tanto, en todos los sistemas de comunicaciones sean estos ópticos, basados en radio o acústicos, puesto que demuestran y validan la capacidad de la AI para aumentar el ancho de banda en sistemas fuertemente limitados en banda. En este caso la limitación viene impuesta por el tiempo de exposición, pero podría derivarse de cualquier otro elemento del enlace.

Otra de las contribuciones de este trabajo es que el entrenamiento del ecualizador se realizó exclusivamente con muestras sintéticas. Para ello se modeló el mecanismo de RS con el tiempo de exposición y se desarrolló un algoritmo eficiente para la generación sintética de muestras. Este algoritmo utiliza únicamente parámetros temporales del enlace como son la tasa de símbolos del transmisor, el tiempo de exposición y el tiempo de muestreo de fila de la cámara. Esta generación sintética se validó indirectamente con la evaluación del ecualizador utilizando muestras reales. Este entrenamiento sintético fue motivado por la dificultad, la complejidad y el gasto de tiempo que implicaba la captura de imágenes reales. Para este propósito habría que caracterizar rigurosamente las cámaras utilizadas, y ajustar convenientemente las condiciones del enlace. Esto provocaría, además que el ecualizador no fuese capaz de generalizar bien y operar correctamente con imágenes provenientes de otras cámaras y tomadas en otras condiciones. Por lo tanto, el entrenamiento sintético, no sólo agiliza la creación de datasets, y elimina la complejidad de desarrollar un banco de pruebas, sino que además permite abstraer el entrenamiento del receptor utilizado, empleando para ello multitud de muestras representativas de una amplia variedad de transmisores, cámaras, configuraciones, niveles de ruido, entre otros parámetros.

Finalmente, cuando se empezaron a utilizar cámaras de móviles convencionales, en lugar de cámaras de laboratorio, para una implementación práctica, se descubrió que muchos de sus parámetros internos son desconocidos, no se pueden modificar, o se actualizan dinámicamente. Por tanto, en estos casos resulta necesario estimar estos parámetros directamente en la recepción, a partir de las imágenes capturadas, antes de proceder a la decodificación. Para este propósito se propone en el tercer artículo de compendio, el uso de redes convolucionales entrenadas con imágenes sintéticas generadas con miles de configuraciones posibles. Los resultados demuestran que estas redes son capaces de obtener errores inferiores al 1.3% y el 3% en la estimación la frecuencia de reloj de los datos y el tiempo de exposición de la cámara respectivamente. Estos errores garantizan la operación óptima del resto de las rutinas de recepción. Como resultado, este estimador desacopla las rutinas de recepción de las cámaras utilizadas, y permite idear una arquitectura que opere directamente con las imágenes con independencia de la cámara utilizada y su configuración. Asimismo, esta arquitectura incluye etapas de ecualización de la exposición para permitir aumentar el tiempo de exposición hasta condiciones de visibilidad óptimas. De este modo se favorece el uso de las cámaras de manera simultánea para la visualización y la recepción de datos. En definitiva, esta propuesta permite la creación de un software de comunicaciones transferible que puede integrarse fácilmente en aplicaciones para teléfonos inteligentes, ordenadores de sobremesa y portátiles, tabletas, sistemas de conducción autónoma, cajeros automáticos, etc. Además, también permite el desarrollo de aplicaciones alojadas en la nube que pueden procesar imágenes provenientes de diferentes flujos multimedia en tiempo real. De esta forma, se habilitan tecnológicamente aquellos dispositivos que no disponen de un ecosistema abierto para el desarrollo de sus aplicaciones.

Adicionalmente, en un trabajo posterior se mejoraron, los tiempos de entrenamiento de los ecualizadores hasta 435 veces, usando sólo 250 imágenes de entrenamiento en lugar de 35500. Esto se consiguió utilizando técnicas para transferencia del conocimiento acumulado por el estimador a los ecualizadores. Esta mejora en la eficiencia del entrenamiento de los ecualizadores permite que estos se puedan entrenar bajo demanda sin llegar a interrumpir la comunicación.

Por último, durante el ejercicio de esta tesis también se ha trabajado en la parte del transmisor, creando un dispositivo llamado Barcolit que opera como un tradicional código de barras, pero de manera activa usando paneles LED. Este sistema, además, es compatible con los convencionales lectores de barras, por lo que no disrumpe los procesos industriales actuales. Además, ofrece ventajas significativas respecto a los códigos de barra impresos, ya que los códigos generados no se distorsionan en la imagen y siempre permanecen correctamente alineados horizontalmente independientemente de la orientación de la cámara.

En conclusión, esta tesis equilibra los esfuerzos por abordar (i) un reto industrial en el campo particular de los enlaces RS-OCC, que es conseguir la reutilización efectiva de las cámaras como receptores sin perder su función primaria como dispositivos de imagen, con (ii) un problema científico-técnico extensible a cualquier campo de las comunicaciones, que es la ecualización de la ISI producida por la reducción del ancho de banda de recepción disponible y la estimación indirecta de ciertos parámetros del enlace. Además, evoluciona y madura en línea con los avances más recientes en el campo de la AI y la visión por computador, registrando resultados que superan los obtenidos con algoritmos clásicos y otros trabajos preliminares. Por último, se centra en desvincular la recepción en enlaces OCC de las cámaras utilizadas, con el objetivo de acelerar la adopción masiva de esta tecnología.

Dedicado a mi madre, mi padre, mi hermana y mi queridísima sobrina.

Acknowledgements

Funding

This work has received funding from:

European Cooperation in Science and Technology, NEWFOCUS COST action (Ref.: CA19111).

European Union's Horizon 2020 Marie Skłodowska-Curie grant agreement (Ref.: 764461).

Spanish Research Administration, Project OCCAM (Ref.: PID2020-114561RB-I00).

Spanish Research Administration, Project: OSCAR (Ref.: TEC 2017-84065-C3-1-R).

Canary Island Regional Government, Project: ATICCuA (Ref.: 2017010053)

Team members and collaborators

Main supervisor Jose Alberto Rabadán

Second supervisor Víctor Guerra Yánez

Academic staff

- Prof. Rafael Pérez Jiménez (IDeTIC)
- Prof. Stanislav Zvánovec (CTU)
- Prof. Zabih Ghassemlooy (NU)
- Prof. Jose Luis Pinchetti (BEA)
- Dr. Veronique Moevaert (MONS)

Technical support

- Dr. Patricia Chávez Burbano (IDeTIC)
- Dr. Behnaz Majlesein (IDeTIC)
- Dr. Vicente Matus Icaza (IDeTIC)
- Ing. Carlos Guerra Yánez (CTU)

- Ing. Veronique Georlette (MONS)
- Dr. Zun Htay (ECRN)
- Dr. Othman Younus (ECRN)
- Dr. Xicong Li (NU)
- Dr. Shivani Teli (CTU)
- Dr. Neha Chaudhary (UA)
- Dr. Carlos Almeida (BEA)
- Ing. Victor Alonso Eugenio (IDeTIC)
- Ing. Jaime Roberto Ticai Rivas (IDeTIC)
- Ing. Alberto Clavijo Rodriguez (IDeTIC)
- Ing. Klára Eöllős Jarošíková (CTU)

External Reviewers

- Dr. Mónica Jorge Carvalho de Figueiredo (PL)
- Dr. Matěj Komanec (CTU)

Institutions and groups

IDeTIC Institute for Technological Development and Innovation in Communications, IDeTIC, Universidad de Las Palmas de Gran Canaria, ULPGC, Las Palmas de Gran Canaria, Spain

CTU Dept. of Electromagnetic Field, Czech Technical University in Prague, Prague, Czech Republic

NU Optical Comm. Research Group, Northumbria University, Newcastle-upon-Tyne, UK

UA Instituto de Telecomunicações, University of Aveiro, Aveiro, Portugal

BEA Spanish Bank of Algae, Universidad de Las Palmas de Gran Canaria, ULPGC, Las Palmas de Gran Canaria, Spain

MONS Electromagnetism and Telecom. Dept, University of Mons, Mons, Belgium

PL Electrical Engineering Department, Polytechnic of Leiria, Leiria, Portugal

ECRN European Collaboration Research Group in Optical Wireless Communications.

CRUE Conferencia de Rectores de las Universidades Españolas, Madrid, España.

Contents

List of Figures

List of Tables

Acronyms

1D one-dimensional. 23, 24, 52, 54, 138, 139

2D two-dimensional. 24, 56, 57, 97

Adam adaptative moment estimation. 40

ADC analog-to-digital converter. 28, 34

AE autoencoder. 38, 41, 47, 48, 138

AI artificial intelligence. vi, xi, 5–9, 11, 19–24, 27, 38, 39, 41, 51, 67, 96, 97, 117, 136, 137, 140–142

AoV angle of view. 32, 61

AWG arbitrary waveform generator. 61, 62

AWGN additive white Gaussian noise. 23, 57, 142

BER bit error rate. 22–24, 52, 64, 97, 118, 138

CAE convolutional autoencoder. 9, 24, 48, 49, 57, 64, 97, 137–139

CCD charge-coupled device. 27

CMOS complementary metal-oxide semiconductor. 27

CNN convolutional neural network. 23–26, 38, 41, 43–48, 64, 117, 140

COTS commercially available off-the-shelf. 2

CSK color shift keying. 25

DOS denial-of-service. 2

ESR exposure-to-symbol ratio. 35, 52, 62, 65, 97, 117, 118, 136, 138

FDA floating diffusion amplifier. 28

FEC forward error correction limit. 22, 97, 118, 138

FoV field of view. 22, 32

FPS frames per second. 3

FSO free space optics. 1

GS global shutter. 3, 14, 15, 20–27

HCL human centric lighting. 2

IS image sensor. 2–4, 13–17, 23, 31, 34, 38, 54, 61, 96, 118, 137, 142

ISI intersymbol interference. vi, xi, 3, 17, 24, 34, 35, 51, 69, 96, 97, 137, 141

JPEG joint photographic experts group. 53, 59, 138, 142

LED light emitting diode. 1, 2, 68

LoS line of sight. 14

ML machine learning. 39, 40, 43, 47, 48

MLP multi-layer perceptron. 23, 38, 41–43, 45, 48

MSE mean squared error. 47, 52, 64

NN neural network. 20–26, 39, 41–44, 47, 52, 63, 97, 138, 139

NPPS number of pixels per symbol. 30, 52, 54, 62, 64, 65, 117, 118, 136

NRZ non-return-to-zero. 54, 57, 62

OCC optical camera communication. vi, xi, 2–11, 13–15, 19, 20, 22, 30, 34, 51, 53, 67, 68, 95–97, 118, 137, 140–142

OOK on-off keying. 4, 23, 24, 26, 29–31, 62

OWC optical wireless comunication. 1, 142

PAM pulse amplitude modulation. 24, 25, 138, 143

PWM pulse width modulation. 22, 25

QAM quadrature amplitude modulation. 20, 143

RE relative error. 65

ReLU rectified linear unit. 42, 43

RF radiofrequency. 2

RMSE root MSE. 65

RMSProp root mean square propagation. 40

RNN recurrent neural network. 25, 143

ROI region of interest. 15, 16, 21, 33, 139

RS rolling shutter. vi, xi, 3–11, 13–15, 17, 18, 20–30, 33–35, 51–53, 60, 61, 96, 97, 137, 140–143

RX receiver. 54–56

SFA source follower amplifier. 28

SGD stochastic gradient descent. 39, 40

SNR signal-to-noise ratio. 3, 15, 22, 24, 26, 31, 33, 34, 53, 57, 59, 60, 96, 97, 137–139

SSH secure shell. 62

SUR space utilization ratio. 68, 69

TX transmitter. 54, 55

UUID universally unique identifier. 62

V2X vehicle to everything. 14

VLC visible light communication. 1–3, 13

YOLO you only look once. 21

Chapter 1

Introduction

Today's society is moving inexorably toward real-time communications of millions of interconnected devices. Hence, the future communications infrastructure must provide high-capacity and low-latency links. Next-generation mobile networks must face the challenge their predecessors have overcome with increasing intricacies: the radio spectrum's saturation. The 6G strategy of expanding the spectrum towards sub-thz frequency bands [1] faces the technical handicap of manufacturing and integrating increasingly smaller antennas (with waveguides in the micrometer range) in user devices. To which is added the high sensitivity demanded by current receivers to support increasingly complex modulations. Therefore, this strategy's success is strongly linked to the research efforts invested in pushing the limits reached in semiconductor technology.

Alternatively, optical wireless comunication (OWC) are positioned as a promising candidate to offer a complementary infrastructure capable of assuming a part of the network capacity requirements of tomorrow's society. Free space optics (FSO) is one of the most successful OWC technology considering its acceptance in markets and industry. This technology provides long-range laser-driven links extensively used as backbone support in hybrid RF/FSO networks and satellite communications [2]. One of this technology's most commercially successful applications is the intra-communication links of Starlink satellites. [3].

On the other hand, visible light communication (VLC) technology, based on light emitting diode (LED) light sources, focuses on medium to short-range links. It aims to reuse diverse lighting elements that are widely available, such as office light panels or downlights, to provide high-speed connection access points. The advantages of this technology compared to radio communications are the following:

1. **Broad unlicensed electromagnetic spectrum availability**. The visible light spectrum ranges from infrared to near-ultraviolet wavelengths, potentially increasing the capacity of communications links. However, it must be taken into account that the efficient exploitation of this spectrum depends on utilizing narrow-spectrum LEDs and passive optical filters.

2. **Interoperability**. Visible light wavelengths cannot traverse opaque objects and do not interfere with radio communication transceivers. Hence, light enables high-speed communications in rooms sensitive to radio interference. Still, it should be mentioned that, as validated in recent experiments [4], light can interact with the exposed silicon of some electromagnetic components (such as silicon microphones).

3. **Simplified coverage customization**. The luminous radiation pattern of LED sources can be easily adapted through passive optical elements. This allows for delimiting network microcells in which the medium is not shared between nodes, eliminating the need for medium access mechanisms that highly impact conventional WiFi networks. Remark that this microcell coverage can be attained using radiofrequency (RF) technology at the cost of deploying dense arrays of antennas that have no other practical function rather than communications as opposed to LEDs.

4. **Privacy/Security enhancement**. The possibility of confining light within a space prevents eavesdropping and denial-of-service (DOS) attacks. This physical property of light propagation does not imply perfect security but adds an extra layer of protection against intrusions, illegitimate network access, and communication interferences.

5. **Human awareness**. Humans are adapted to the perception of light. This can be exploited to provide unique interactions between humans and networks, in other words, to make humans sensitive to networks. For instance, VLC transmitters can adapt their light parameters (color, intensity, flickering) to notify users of network events (inoperability, congestion, high latency). In addition, the design of transmitting light sources can be aligned to the recent concept of human centric lighting (HCL), which seeks to equate artificial light to natural light as much as possible to improve the quality of life and the well-being of people.

Among its drawbacks is the need for a continuous light-up source (even during the day), which can irritate the human eye in particular scenarios. Moreover, the transmitted light must not be dangerous to humans or harms the environment. For instance, flickering perceived in the light source must be alleviated as it can cause migraines and persistent headaches. Furthermore, the ambient light can significantly interfere with VLC links, which detriments its performance.

Despite these drawbacks, this technology is trying to find diverse application niches to offer added value. The application with the most support from the engineering community is indoor localization. Optical beaconing based on VLC provides millimetric precision that exceeds the performance of previous radio-based methods [5, 6, 7]. Another market-ready application is a general purpose VLC access points (for internet access) developed by pureLifi [8], Oledcomm [9] and Philips (Signify) [10]. Other applications of interest are vehicle-to-infrastructure, and vehicle-to-vehicle communications (V2X) [11], underwater optical wireless communications [12, 13, 14] and energy harvesting communications [15, 16].

Despite these efforts, VLC technology has yet to be massively adopted by the market. Photodiode-based links need aggressive investment in integrated circuit design and manufacturing and their further embedding in commercially available off-the-shelf (COTS) devices. Alternatively, OCC, a branch of VLC, which replaces the photodiodes at the receiver side with image sensors (ISs) has generated a strong interest in the community, being finally included in the 2018-revised version of the VLC IEEE 802.15.7 [17] standard, and in the ISO 22738:2020 [18]. This technology aims to reuse conventional cameras embedded in an increasing number of end-user devices (such as mobile phones, laptops, and vehicle dashcams), ultimately paving the way for VLC to break the market's entry barriers.

This field is divided into two subfields with different application niches based on the camera's image acquisition mechanism: global or rolling shutter. On the one hand, global shutter (GS) cameras simultaneously expose all the pixels of IS at the time of capturing one frame. Therefore, these cameras sample the light source with a frequency equal to the camera's frame rate (which conventionally varies from 30 to 120 frames per second (FPS)). Accordingly, the receiver bandwidth of these cameras is extremely low. Alternatively, RS cameras expose the IS sequentially row-by-row[1] of pixels, sampling the light at different instants while the camera's shutter remains open. As a result, a compound of different illuminated stripes appears in the image depending on the light source's illumination state at the exposition of each row. Hence, the light is sampled periodically at each row's sampling instants. The interval between the activation of two consecutive rows is the row sampling time. It is directly related to the receiver's sampling time, whose inverse corresponds to the receiver's sampling rate. Conventionally, this interval extends from tens to hundreds of microseconds, resulting in sampling rates of up to 100kHz. Consequently, the attainable sampling rate in RS cameras is several times greater than in GS cameras.

This thesis focuses on RS cameras since they are the predominant technology in the market due to their low manufacturing cost. Furthermore, the techniques used with GS cameras can be easily transferred to RS cameras in those cases where the source's light intensity varies more slowly than the frame capture time.

Despite RS-based OCC significantly speeding up the time to market of VLC products, it must face six technical challenges, compiled in this thesis, before its successful assimilation in market applications.

CH1 **Bandwidth constrained by the camera's exposure**. In RS cameras, the exposure time, which is the time a pixel remains exposed to light, behaves as a low pass filter (Chapter 3). In long exposures, pixels may integrate the irradiance levels of several consecutive symbols, producing an ISI that is more harmful depending on the ratio between the exposure time and the symbol transmission time. Ultimately, the exposure time further restricts the attainable receiver bandwidth.

CH2 **Windowed reception**. The light rays from the transmitting source reach different parts of the image sensor (after traversing the camera's optical system). The result is a two-dimensional map (the image) with different intensity values for each pixel. Therefore, each pixel is affected by a different signal-to-noise ratio (SNR) based on its incident irradiance. Consequently, It is crucial to propose mechanisms for identifying those image regions with the highest SNR to proceed with data decoding. These regions commonly belong to the light source's projection within the image, although they can also come from its reflections in walls or mirrors. Thus, the number of valid data samples depends on the source's projection size. The greater it is, the higher number of pixels that successfully sample the light source. Consequently, the geometrical arrangement, physical properties (physical size of the light source), and specific optical parameters affect the achievable data throughput (Chapter 3). In addition, the relative movement of the light sources with respect to the camera causes the irradiance distribution

---

[1] Depending on the scanning dimension of the IS, vertical or horizontal, the camera exposes row-by-row or column-by-column.

map over the IS to vary over time, causing dynamic windowing, which implies the use of channel coding and data detection techniques in each frame.

CH3 **Non-linear and time-variant system**. On the one hand, cameras apply a gamma correction (a non-linear transformation) to encode the pixel luminance values, considering how humans perceive light and color. This gamma correction boosts low-intensity values, producing a more evident non-linear effect at those levels. On the other hand, the sequential acquisition of RS cameras makes the link time-variant. To justify this statement, let's consider a static transmitter located in a scene that repeatedly sends an on-off keying (OOK) optical signal, encoding bits using two illumination states ("ON" for ones and "OFF" for zeroes). In that case, if the irradiance distribution map over the IS is non-uniformly distributed (when the source is "ON"), then each pixel's value will depend on its position within the IS. On the other hand, the sequential acquisition of RS cameras relates both the pixel position (space) and the sampling instant (time). Consequently, the intensity value of equivalent "ON" state samples will be determined by the time the sample was taken. The justification behind this statement is further discussed in Chapter 3. Therefore, RS-based OCC requires spatially uniforming the received energy across the IS by either using diffusers at the transmitter side or by using equalization techniques at reception.

CH4 **Lack of synchronization**. During acquisition, cameras wait, between frames, for the allocation of memory resources. During those times, the camera remains blind to the variations light is undergoing. This contributes to the reception's windowed behavior and introduces a variable shift in the sampling instants after each frame capture. Hence, data synchronization techniques are demanded at each frame.

CH5 **Simultaneous visualization and data detection**. The industry requires that the functionality of cameras as visualization devices be preserved when they are used as optical receivers. This is challenging when working with RS cameras, as to increase the receiver bandwidth, the camera's exposure time must be as short as possible (Challenge I). However, this penalizes the camera's sensitivity, as pixels accumulate less irradiance from the scene. Consequently, the camera delivers dark images in which objects in the scene are barely perceptible. In other words, the resulting images provide poor lighting conditions for human or machine-supervised applications. The exposure time ultimately establishes a trade-off between the camera's sensitivity and the receiver's bandwidth (Chapter 3).

CH6 **Massive adoption**. The wide variety of cameras that flood the market differ significantly in their hardware and configuration parameters. In addition, the vast majority of cameras do not reveal specific manufacturing parameters important in communications, such as their source clock frequency, from which the row sampling frequency is derived (the receiver's sampling rate in RS cameras). Moreover, some cameras do not allow access to their internal settings, such as the exposure time. Other cameras use automatic closed-loop control mechanisms to regulate their internal parameters. Therefore, the mass adoption of OCC requires characterizing all those camera parameters that influence communications and developing architectures that allow decoupling the reception routines from the cameras.

1.1 Motivation

This thesis departs from a previous research work conducted by the author for his master's thesis, and it is aligned with the research lines of the *Photonic Technology Division* of the *Institute for Technological Development and Innovation in Communications (IDeTIC)* at the *Universidad de Las Palmas de Gran Canaria (ULPGC).*

The previous work entitled "Design, characterization, and implementation of an optical communications system based on cameras" [19] proposed an ad-hoc OCC academic prototype that aimed to address challenges CH3 (equalization) and CH4(synch.) raised in the introduction. The plan to integrate this prototype into an industrial environment marks the starting point of this thesis. The preliminary analysis of state-of-the-art and the practical study of the technical link requirements for the proposed scenario led to the compilation of a preliminary version of the challenges presented in this introduction. As the thesis evolved, those challenges became more solid and consistent.

This thesis focuses on challenges CH1 (to increase the attainable bandwidth), CH3 (to mitigate the non-linearity effects of the gamma), CH5 (to enable the simultaneous operation of cameras as visualization and receiver devices), and CH6 (to decouple cameras from the receiver routines). For this purpose, it suggests using AI as an alternative to conventional signal processing strategies. The following section presents the raised, reviewed, and assessed hypothesis along with the objectives proposed in the research plan.

1.2 Hypotheses

The four hypotheses raised in this thesis follow a hierarchical dependency, in which the validation of any of them has repercussions on the immediate superior, taking a step forward in achieving its validation (although it does not entirely validate them). These hypotheses are presented in a hierarchical order, from the top (higher dependency) to the bottom.

H1 **It is possible to simultaneously use a RS camera for visualization from a human-centric viewpoint and data detection.**

In the design RS-based OCC links, there is an underlying trend of decreasing the exposure time to a minimum without a proper consensus. This strategy is widely assumed throughout the literature as the only solution to perfectly visualize the symbol bands of the transmitted optical signal in the image. Conversely, increasing the exposure time produces the blend of several consecutive symbols, causing the signal to be incorrectly decoded. This literature trend of reducing the exposure time, which lacks a theoretical framework that supports it, causes, on the other hand, the delivered images to be practically dark. Therefore, objects in the image cannot be perceived. Thus, the camera loses its functionality as a visualization device. This hypothesis can then be reformulated as *"Is there a method that allows increasing the exposure time of a RS camera without harming the received signal"*. This search raised the hypothesis H2.

H2 **It is possible to reconstruct bandwidth-limited signals using AI-assisted equalizers.**

The search for methods to increase the exposure entails modeling the sequential scanning of RS cameras and analyzing the effects of the exposure time. In this model, the sequential exposition behaves as a low-pass filter acting on the transmitted signal. Therefore, it limits the attainable receiver bandwidth, significantly attenuating the signal's frequency components that fall outside the specified filter's cutoff frequency. For this reason, a possible solution to increase the exposure time is to exploit those attenuated frequency components to reconstruct the transmitted signal as if it were received with a receiver with a broader bandwidth (i.e., lower exposure time). For this purpose, the use of AI is proposed as a promising candidate capable of dealing with the complexity associated with (i) the great diversity of exposure times considered, (ii) the limitation of traditional digital signal processing algorithms, and (iii) the non-linearity of the cameras (which apply a gamma transformation on the luminance values). This hypothesis can be further reformulated with the following statement: "A neural network model can exploit the attenuated frequency components of a bandlimited received signal that falls beyond the cutoff frequency to reconstruct the original signal as if it were acquired with a broader bandwidth.".

The validation of this hypothesis is one of the main contributions of this work. Furthermore, as can be witnessed from the reformulated hypothesis, the results of this work are transferable to any bandlimited communications system. In this case, the band limitation is imposed by the camera behavior related to the exposure time; however, it could also be imposed by the transmitter (e.g., the LED driver's response), the channel, or any element in the communications chain.

H3 **It is possible to train an RS OCC-based AI system using synthetic samples.**

The unprecedented challenge of obtaining real OCC samples for the training and validation of neural networks leads to the formulation of this hypothesis. Furthermore, its validation will be an important contribution to this thesis since it indirectly validates the model used for the synthetic generation of samples.

H4 **It is possible to recover both the signal clock and the camera's exposure time from severely distorted images.**

The search for an effective strategy to decouple the camera's exposure time from the reception routines, taking a step forward in the search for the total decoupling of cameras, leads to envisioning an element capable of delivering this parameter even when it cannot be selected or obtained from the camera.

The solution is to estimate it (along with other link parameters) directly from the images using AI techniques.

1.3 Objectives

This thesis has three general objectives (**GO**). They are stated below, along with a brief description, and broken down into technical objectives (**TO**).

GO1 **To develop an RS-based OCC receiver assisted with AI with support for simultaneous visualization and data detection.**

This objective aims to validate the hypotheses: H1, H2, and H3.

As mentioned in the introduction, one of the challenges addressed in this thesis (Challenge CH5) is enabling the simultaneous operation of RS cameras. Preliminary work indicates that the camera's exposure time is crucial, as it is involved in the camera's sensitivity (visualization) and the receiver's bandwidth (communications). Therefore, evaluating the impact of this parameter on the transmitted signal is a central part of this thesis. Moreover, mitigating its harmful effects on the signal will contribute to efficiently exploiting the achievable bandwidth offered by cameras, overcoming the Challenge CH1. The use of AI techniques is considered for this purpose. Furthermore, as discussed in Chapter 5, the complexity faced in capturing real-world training samples forces the adaptation of specific objectives to perform training based on synthetic images.

TO1.1 To analyze in depth the camera's exposure influence in visualization (sensitivity) and communications (bandwidth).

TO1.2 To evaluate the impact of the camera's exposure effects on the signal.

TO1.3 To elaborate a comprehensive mathematical model of the RS acquisition mode.

TO1.4 To develop a procedural method for generating synthetic RS-based OCC samples.

TO1.5 To design and train AI-assisted models for exposure equalization and regression using synthetic datasets and conduct their validation using real images.

GO2 **To develop an experimental testbed for OCC and generate research datasets.**

This objective aims to validate experimentally the hypotheses: H1, H2, and H3.

The development and rigorous characterization of a connected experimental platform for capturing real-world images will allow the generation of valuable datasets. These datasets released to the scientific community during the exercise of this thesis will initially serve as a tool for (i) analyzing the impact of the exposure time on the received signal, (ii) validating novel AI-based approaches for exposure equalization, and (iii) validating novel estimators of the intrinsic parameters of the received signal.

TO2.1 To design an experimental setup for OCC image acquisition.

TO2.2 To assess an in-depth characterization of both transmitter and receiver devices.

TO2.3 To produce a real-world OCC dataset for system validation.

GO3 **To evaluate the integration and viability of OCC in industrial and mass-market applications.**

This objective aims to validate the hypothesis H4.

As presented in the introduction, the potential integration of OCC in the industry and the market comes from bringing solutions that consider the wide variety of cameras available (Challenge CH6). To do this, it is necessary to (i) study the cameras as receiving devices, (ii) analyze how their internal parameters influence the optical signals embedded in the images, and, most importantly, (iii) establish mechanisms to decouple cameras from the reception routines. Consequently, the focus in designing new OCC links shifts from working with cameras to working with images, regardless of their origin. In addition, the distribution of transmitters within a room or over a wall depends on the camera's position and orientation to establish a line of sight while reserving a space free of obstructions between them. Therefore, analyzing how the deployment of transmitters impacts communications must be carried out, considering space limits or restrictions. Finally, emphasize again that the scene monitoring provided by cameras is essential in many scenarios and should be kept from being interrupted for other purposes. For example, this is the case with closed-circuit television surveillance cameras.

Ultimately, this objective paves the way to the practical integration of OCC in the industry and the market.

TO3.1 To analyze the potential integration of OCC into mass-market applications.

TO3.2 To propose preliminary architectures to decouple cameras from receiver routines.

TO3.3 To analyze the impact of multi-transmitter deployments in industrial scenarios.

1.4 Thesis structure

The chapters of this thesis are structured as follows. Chapter 2 of this book provides a literature review on the most current and important contributions related to OCC and AI. First, It follows a top-down approach, starting with a general overview of the current challenges, trends, and applications of OCC and then focusing on contributions related to RS cameras. Next, this chapter examines topics such as analyzing the influence of camera parameters on communications, the simultaneous use of cameras as both imaging and communication devices, and the widespread adoption of OCC. Finally, it dissects the contributions related to AI and OCC.

Chapter 3 delves into the technical and theoretical considerations for using RS cameras as optical receivers. It begins by discussing the RS acquisition mechanism and how it relates to data acquisition, focusing on the effects of the camera's exposure time on the signal. This chapter also introduces the fundamental concepts of AI necessary to understand the contributions of this thesis.

Chapter 4 discusses the methods and experiments carried out in this thesis. It includes the latest version of the synthetic generation procedure used to generate and compile the training datasets. It also examines the experimental setup used for capturing real-world validation samples.

Chapter 5 presents the research outcomes supporting this thesis's objectives. It includes three original journal manuscripts sorted by publication date. The original manuscripts are embedded in this chapter, including a brief introduction that describes the context of the work, the challenges addressed, the results obtained, the contributions, and highlighted findings. This chapter also summarizes related works published by the author of this thesis that are not included in the compendium.

Finally, Chapter 6 summarizes the findings and the potential impact of this thesis's results in the scientific community, industry, markets, and society and suggests derived future research lines.

1.5 List of publications

This thesis has generated three publications in indexed journals, included in the compendium, and three in conference proceedings. In addition, the author has collaborated with other researchers from international institutions in 10 publications. The following sections list the mentioned articles, including a brief description.

1.5.1 Compendium

This section lists the publications included in this thesis's compendium, which are detailed in Chapter 5.

[20] C. Jurado-Verdu, V. Guerra, V. Matus, C. Almeida, and J. Rabadan, "Optical camera communication as an enabling technology for microalgae cultivation," *Sensors*, vol. 21, no. 5, 2021

This work proposes a novel photobioreactor that uses artificial lighting to feed a microalgae culture and simultaneously send optical codes to a surveillance camera. In addition, it analyzes the optimal deployment of several transmitter nodes (photobioreactors) in a room in terms of efficient space utilization and quality of communications. It addresses the challenge of the massive adoption of OCC (Challenge CH6).

[21] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, and R. Perez-Jimenez, "Convolutional autoencoder for exposure effects equalization and noise mitigation in optical camera communication," *Opt. Express*, vol. 29, pp. 22973–22991, Jul 2021

This work proposes a novel AI-assisted equalizer based on convolutional autoencoder (CAE) to alleviate the harmful exposure-related effects on the received signals in RS-based OCC links, in which the camera's exposure time is greater than the symbol time. It addresses the challenges of bandwidth constrained by the camera's configuration and the simultaneous operation of cameras as visualization and receiver devices (Challenges

CH1 and CH5, respectively).

[22] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Deep learning for signal clock and exposure estimation in rolling shutter optical camera communication,” *Opt. Express*, vol. 30, pp. 20261–20277, Jun 2022

This work proposes a novel regression convolutional network to estimate important OCC link parameters directly from the received images: the camera’s exposure time and the signal clock frequency. Ultimately, this network allows the decoupling of the camera’s configuration and hardware from the reception routines. This work addresses the challenges of the simultaneous operation of cameras as visualization and receivers devices and the decoupling of cameras from the reception routines (challenges CH5 and CH6, respectively).

1.5.2 Master thesis publications

This section lists the author’s related journal work published before the doctorate program started that paved the way for the formulation of the research plan of this thesis. This publication is briefly discussed in Chapter 5 and appended in the Annex A.

[23] C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, “Correlation based receiver for optical camera communications,” *Opt. Express*, vol. 27, pp. 19150–19155, Jul 2019

This work proposes a fully operating RS-based OCC in which source detection and tracking, spatial equalization, and data decoding are implemented using correlation-based techniques. It addresses the challenges of windowed communications and lack of synchronization (challenges CH2 and CH5, respectively).

1.5.3 Conference proceedings

This section lists the camera-ready conference proceedings briefly described in Chapter 5 as related contributions to the compendium. In addition, they are appended in the Annex B).

[24] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, R. Perez-Jimenez, J. Luis Gomez-Pinchetti, and C. Almeida, “Application of optical camera communication to microalgae production plants,” in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2020

[25] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Barcolits: Barcodes using led tags and optical camera communications,” in *2022 IEEE 18th International Conference on Factory Communication Systems (WFCS)*, pp. 1–8, 2022

[26] C. Jurado-Verdu, V. Guerra, C. Guerra, J. Rabadan, S. Zvánovec, and R. Perez-Jimenez, “On-demand training of deep learning equalizers for rolling shutter optical camera communications,” in *2022 13th International Symposium on Communication*

Systems, Networks and Digital Signal Processing (CSNDSP), pp. 1–5, 2022

1.5.4 Datasets

This section describes the open datasets generated as supplemental material for [22], which was released to the scientific community.

[27] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Effects of the camera’s exposure time on rolling shutter based optical camera communication links.,” *Figshare*, 3 2022

This dataset contains 7281 example images of a RS-based OCC link, in which the camera is configured with increasing exposure times and signal clock frequencies. The original purpose of this dataset is to provide a tool for (i) analyzing how the exposure time affects the received signal, (ii) validating novel AI-based approaches for the exposure equalization of the received signal, and (iii) validating novel estimator models of the intrinsic parameters of the received signal.

1.5.5 Collaborations

This sections presentes the journals and conference proceedings in which the author has collaborated.

[28] A. Mederos-Barrera, C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Discovering and tracking-based detection system for optical camera communication,” in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2020

[29] V. Matus, S. R. Teli, V. Guerra, C. Jurado-Verdu, S. Zvanovec, and R. Perez-Jimenez, “Evaluation of fog effects on optical camera communications link,” in *2020 3rd West Asian Symposium on Optical and Millimeter-wave Wireless Communication (WASOWC)*, pp. 1–5, 2020

[30] V. Matus, V. Guerra, C. Jurado-Verdu, S. R. Teli, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, “Experimental evaluation of an analog gain optimization algorithm in optical camera communications,” in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–5, 2020

[31] A. Mederos-Barrera, C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Design and experimental characterization of a discovery and tracking system for optical camera communications,” *Sensors*, vol. 21, no. 9, 2021

[32] V. Georlette, F. Piras, C. Jurado–Verdu, S. Bette, N. Point, and V. Moevaert, “Content triggering system using tricolor led strips and optical camera communication in rolling shutter mode,” in *2021 Third South American Colloquium on Visible Light Communications (SACVLC)*, pp. 01–06, 2021

[33] V. Matus, V. Guerra, C. Jurado-Verdu, S. Zvanovec, and R. Perez-Jimenez, “Wireless sensor networks using sub-pixel optical camera communications: Advances in experimental channel evaluation,” *Sensors*, vol. 21, no. 8, 2021

[34] V. Matus, V. Guerra, C. Jurado-Verdu, J. Rabadan, and R. Perez-Jimenez, “Demonstration of a sub-pixel outdoor optical camera communication link,” *IEEE Latin America Transactions*, vol. 19, no. 10, pp. 1798–1805, 2021

[35] V. Matus, V. Guerra, C. Jurado-Verdu, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, “Design and implementation of an optical camera communication system for wireless sensor networking in farming fields,” in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2021

[36] C. Guerra-Yánez, V. Guerra, C. Jurado-Verdú, J. Rabadán, R. Pérez-Jiménez, Z. Ghassemlooy, and S. Zvánovec, “General framework for calculating irradiance distributions of symmetric surface sources,” *Opt. Express*, vol. 30, pp. 43910–43924, Nov 2022

[37] B. Majlesein, V. Matus, C. Jurado-Verdu, V. Guerra, J. Rabadan, and J. Rufo, “Experimental characterization of sub-pixel underwater optical camera communications,” in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 150–155, 2022

Chapter 2

Literature review: a state-of-the-art analysis

This chapter provides a literature review of the most important contributions to the exercise of this thesis and is updated at the time of writing the thesis. Following a top-down approach, this chapter lists the general findings and insights on the current challenges, trends, and applications of OCC. Then, it examines those contributions related to RS cameras, emphasizing those that justify the challenges raised in this thesis. The selected topics are (i) the analysis of the influence of camera parameters on communications (such as exposure time, the analog gain, among others), (ii) the simultaneous operation of cameras as imaging and communication devices, and (iii) the massive adoption of OCC. Finally, it details general and specific research lines related to the application of artificial intelligence to the field of OCC.

OCC technology is considered a branch of VLC technology, which uses cameras (i.e, ISs) as optical receivers instead of custom photodiode arrangements. This technology has recently been included in the latest version of the IEEE 802.15.7-2018 standard [17], which reveals the interest of the engineering and scientific communities in this type of communication system. This interest was also reflected in a survey conducted by Zaman et al. in 2018 [38], in which the authors compared wireless optical communication technologies. This comparison positioned OCC as an emerging technology that offers potential technical advantages over traditional VLC, including (i) easiness at discerning light sources due to the camera's spatial diversity capabilities, (ii) stability of the received power over distance thanks to the camera's optical lenses[1], and (iii) robustness to interferences and partial blocking of the sources. It also listed as disadvantages: (i) the limited bandwidth, (ii) the increased risk of flickering due to the required slow switching speed of transmitting sources, and (iii) the presence of unavoidable signal distortions produced by image transformations (including image compression) included in most cameras' acquisition pipelines. To these limitations, other authors, such as Le et al. [40], added the challenges arising from the lack of synchronization due to the variability in the cameras' frame rate.

Despite its disadvantages, this technology has found potential niche applications as listed by Mohsan et al. [41] in their most recent survey conducted in 2021, including

---

[1] This statement is correct insofar as the projection of the emitter onto the surface of the image sensor is larger than one pixel. The use of image-forming optics compensates for the power loss due to spherical propagation with the projected size of the optical source on the image sensor [39]

vehicle to everything (V2X), indoor positioning, digital signage (for advertising and marketing), drone-to-drone communications, and augmented reality. In V2X communications rear and front cameras are used to discern a multitude of transmitting light sources, such as car headlights and traffic lights, enabling directional links with minimal interference (2022 survey by Hasan et al. [11]). Indoor positioning systems also benefit from the spatial diversity of cameras. OCC systems record location errors in the order of tens of millimeters, considerably lower than those achieved with traditional radio antennas.

Therefore, much effort and resources are invested in developing novel coding and modulation techniques adapted to this technology's constraints, applications, and use-case scenarios. The obtained results are reported and compared in the survey of 2019 carried out by Saed et al. [42]. This report classified and compared these systems based on the transmitter: (i) stand-alone LEDs or wide are LED panels, (ii) LED arrays, (iii) and screens or displays, focusing on the standardization, channel characterization, modulation, coding, synchronization, and signal processing techniques. This report concludes by enumerating several open research lines to address technical challenges in OCC, including (i) the dependency (i.e., coupling) between the receiver routines with the camera, which affects the fair comparison between OCC systems and hinders their massive adoption, (ii) the ambient light interference, (iii) the defocusing of transmitter sources, which might lead to increase the data rates, (iv) the modeling of the receiver's noise, (v) the synchronization considering different camera's frame rates, (vi) the unexplored MAC and network layers for this systems, and (vii) the blockage of the line of sight (LoS). The non-LoS links established from light reflections off walls, and objects are examined by Weijie et al. in 2020 [43].

On the other hand, OCC systems can be classified based on the camera used, or more specifically, based on the image acquisition mechanism they use (Le et al. [44], and Willy et al. [45]: (i) global shutter and (ii) rolling shutter. The survey by Willy et al. [45] analyzed in depth the differences in techniques, modulations, and main constraints of these two variants.

In the former case, the GS-based IS exposes all its pixels simultaneously during capture time. As a result, the sampling rate coincides with the frame rate. However, this rate is considerably low (compared to traditional radio samplers), around 60 to 120 Hz. This slow sampling not only limits the available receiver bandwidth but can also lead to the slow switching of the transmitting sources, increasing the risk of perceived flickering. To overcome this issue, undersampled-based modulation schemes (Luo et al., [46]) are used at the cost of further reducing the available bandwidth. In addition, the variability in the capture times of consecutive frames results in an irregular sampling (i.e., jitter).

In the latter case, in RS cameras, the exposure of the IS is sequential row by row of pixels, sampling the light at each row interval, resulting in the possibility of intra-frame communication. In this case, stripes or bands of different intensities are perceived in the image, corresponding to the scene illumination during each row exposure. Therefore, communication is faster, and the risk of flickering is reduced. However, this technology also has several limitations. The following section highlights some of these limitations concerning the camera parameters, with particular emphasis on the camera's exposure time.

Finally, for ease of reading, systems, transmitters, or techniques are labeled with the terms RS or GS for convenience from now on. This terminology must be clarified since

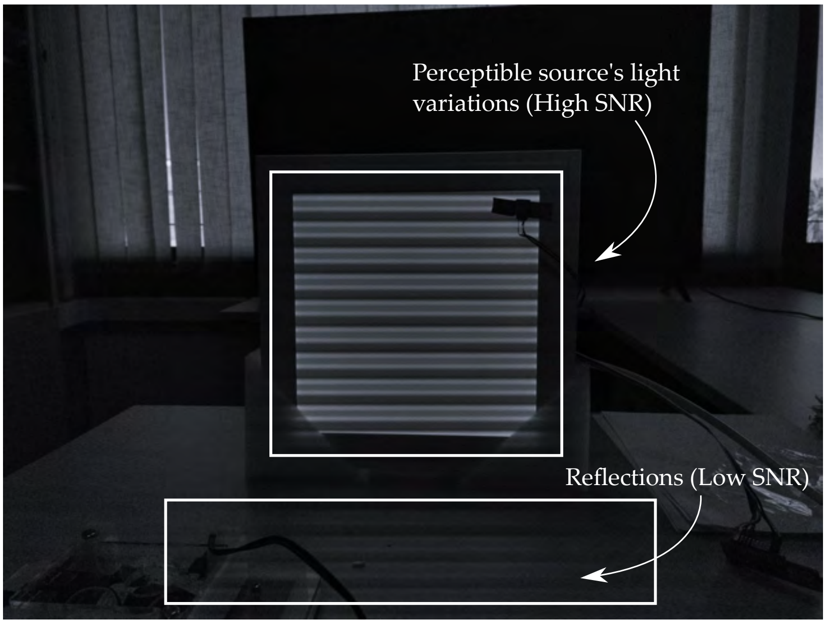


Figure 2.1: Example of a rolling shutter link using a LED flat panel as a transmitter.

it refers exclusively to camera acquisition. In the rest of this thesis, this terminology is extended to the rest of the elements involved in an OCC link to help differentiate GS and RS systems. Hence, an RS transmitter is the one that works with baud rates higher than the frame rate of the camera, and GS in the opposite case. Furthermore, a link classified as GS might not use a GS camera. In fact, the vast majority of cameras are RS. In other words, GS techniques can be combined with RS cameras. Conversely, systems classified as RS can only work with RS cameras.

This thesis and the remainder of this chapter focus on RS techniques and links.

2.1 Rolling shutter systems

The main limitation of OCC systems is the achievable receiver bandwidth. As aforementioned RS systems can achieve higher sampling rates than their GS counterparts. This is because they sample the light at each row interval. However, the light source's illumination changes might not be perceptible in some rows, as the irradiance levels over those areas of the IS need to be stronger. As a result, the effective sampling rate is lower than the attainable sampling rate. The effective sampling is determined by the number of rows in which the optical signal variations are discernible; in other words, the number of rows in which the SNR exceeds a suitable threshold for decoding. In general terms, the projection of the light source in the image, as shown in Figure 2.1, is where the signal-to-interference-to-noise ratio (SINR) is the highest (Chapter 3)), and it is considered the region of interest (ROI). Therefore, the larger its projection, the higher the attainable transmission rate, as more symbol bands can fit within it. In conclusion, the cameras' spatial diversity capability sharpens the contours of the transmitting sources, reducing the effective sampling rate.

Two consequences can be extracted from this phenomenon. First, variables such as the distance of the light source and its physical size come into play, as well as many other considerations regarding its position relative to the camera. This issue was high-

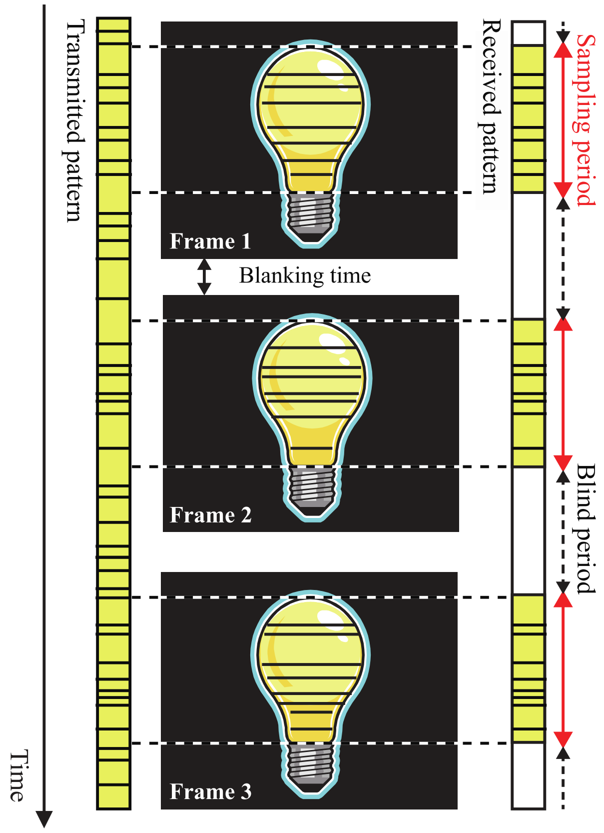


Figure 2.2: Receiving and blind periods at reception by Aoyama et al. (June 2015) [50].

lighted in previous work by the author of this thesis, [47, 48], and, later that year (2019), reiterated by Nguyen et al. [49]. The camera's optical system also affects. Its magnification, angle of vision, and aberrations, among many other parameters, influence the light projection in the IS. Second, the reception is not continuous but windowed, as symbol bands are lost in those image regions where the SNR is unsuitable for demodulation. To complicate matters further, this windowing is dynamic. It changes over time depending on multiple factors, including (i) the relative movement of the sources, (ii) the presence of partial or total occlusions, and (iii) the attenuation caused by foggy, rainy, or snowy conditions, among others. Furthermore, between two consecutive captures, the IS remains blind to the transmission while waiting for memory resources to be allocated. These waiting times are variable and ultimately result in an unstable frame capture rate, which may induce a jitter in the received signal. This windowed reception is carefully examined by Aoyama et al. [50] and illustrated in Fig. 2.2. Therefore, precisely identifying and locating the light sources in each frame is one of the community's main challenges.

On the other hand, the signal strength is not uniform within the ROI as shown in Fig.2.3. Accordingly, intra-frame spatial equalization is required to adapt the image's dynamic evolution of the light source. Several authors utilize different thresholding schemes at the demodulation stage to address this challenge, which are compared by Liu et al. in [51]. On the other hand, Liang et al. [52] used an entropy-based approach for this task. Also, Zhang et al. [53] proposed to demodulate those columns where the extinction ratio is low, i.e., where the difference between the intensity of the highest levels (transmitted 1) and the lowest levels (transmitted 0) is minimum. This generally

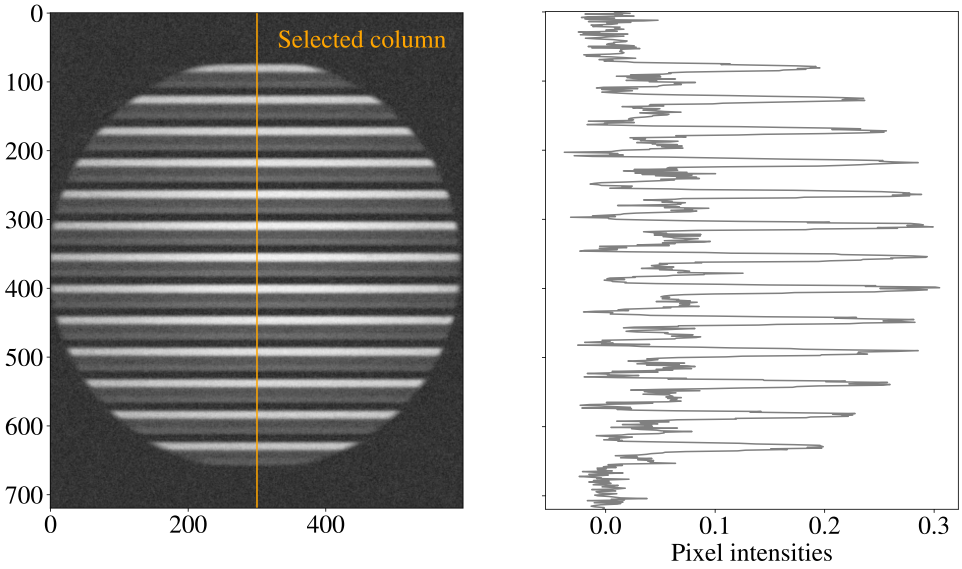


Figure 2.3: Pixel intensities for the green channel (in grayscale) [23]. The right graph shows the pixel intensities values for the selected column highlighted in orange.

happens at the boundaries of the light source. Finally, this challenge has also been addressed in a previous work of the author of this thesis. In this case, RGB lamps [23] are decoded using third-order polynomial thresholding. In conclusion, it is necessary to segment the source in the image and evaluate the distribution of its intensity values for effective decoding.

However, this equalization is only sometimes possible. For example, if the source saturates the sensor due to its intensity, any equalization would produce distortion due to signal clipping and information loss. In those cases, it is necessary to adapt the receiver's sensitivity. This sensitivity is related to two camera parameters: the overall sensor gain and the exposure time (Chapter 3). The overall gain, in turn, is divided into analog and digital gains, but for practical purposes, they will be considered together. Increasing or decreasing the camera's gain directly affects the receiver's sensitivity. A later work [30, 54] reveals the improvement that adaptive analog gain adjustment brings to data reception in a low visibility environment.

On the other hand, the exposure time is related to the time the pixels of the IS are exposed to light (after activation row by row). The longer they are exposed, the more charge they accumulate, thus improving sensitivity. This is advantageous in static environments, where the scene's illumination hardly changes during capture. Otherwise, for example, if the camera or objects are moving, the image is blurred. The same effect occurs when an optical signal is transmitted. When the camera is set with high exposure times, the pixels accumulate the irradiance of several consecutive symbols. The effect on the image is mixing adjacent bands as shown in Fig. 2.4. Ultimately, prolonged exposure times produce the appearance of ISI, limiting the attainable receiver bandwidth. This phenomenon has been noticed in the community. Most of the contributions for RS systems agree that the camera's exposure time has to be set to minimum values and counteract the loss in sensitivity by increasing the analog gain. However, these works do not state or justify the reason behind this decision. Oth-

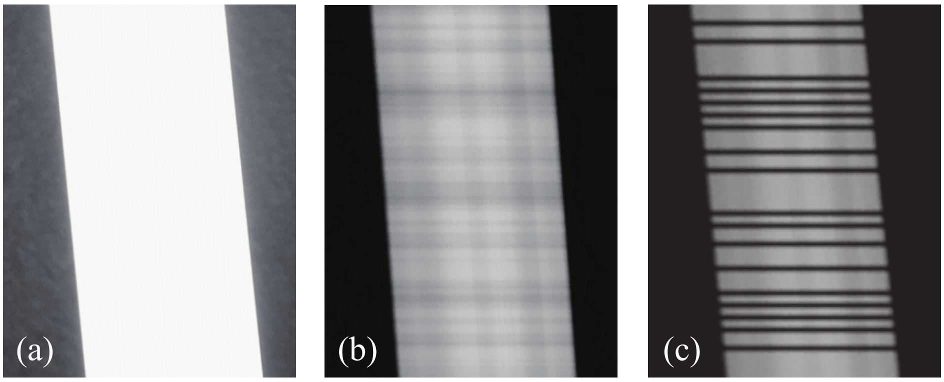


Figure 2.4: Examples of the effects of decreasing the exposure time by Aoyama et al. [50]. From long to short: (a) 1/100, (b) 1/1000, (c) 1/10000 seconds.

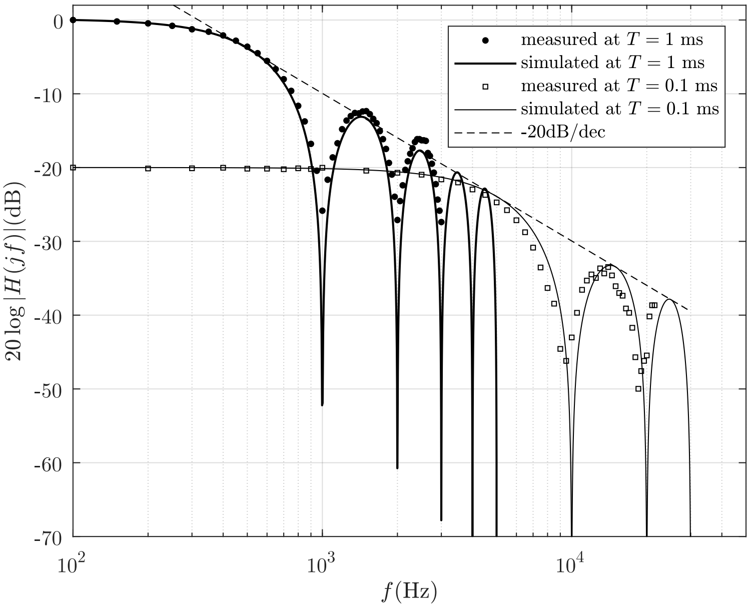


Figure 2.5: Frequency response of the camera for different exposure times ($t_{\text{exp}} = 1\,\text{ms}$ and $t_{\text{exp}} = 0.1\,\text{ms}$) by Li et al. [55].

ers, on the other hand, evaluate the detriment that increased exposure time brings to the bit error rate without characterizing its effect on the signal. Hence, a theoretical framework is required to model the exposure of RS cameras. To date, invested efforts within these lines are limited. In previous work, in collaboration with Li et al. [55], a preliminary approach is made, modeling the effect of exposure on the signal as a low-pass filter. Figure 2.5 shows the camera's measured and simulated frequency response for two different exposure times. In this model, the first null frequency corresponds to the inverse of the exposure time. However, this model is greatly simplified and needs to consider, for example, dynamics in the received intensity. Two years later (2021), Younus et al. [56] recovered this preliminary modeling to use for equalizing its effects. In conclusion, the exposure time is a crucial parameter to consider in the design of RS links. The adjustment of this parameter raises a trade-off between sensitivity and receiver bandwidth.

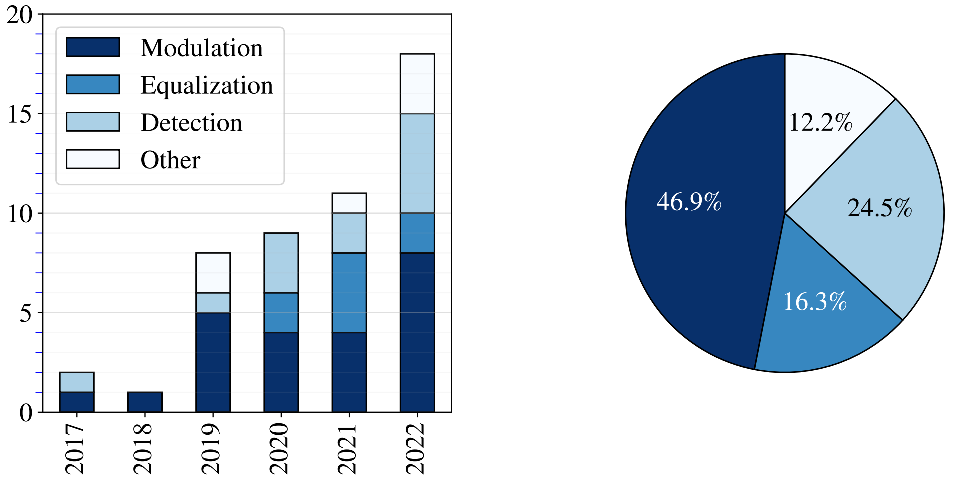


Figure 2.6: Distribution of contributions related to AI and OCC.

Finally, the mass adoption of the OCC technology requires the development of solutions that cover as many cameras as possible (Saed et al. [42]). Hence, given the incredible diversity of cameras flooding the market, it is crucial to analyze the parameters that influence communications, as mentioned above. In this way, they can be favorably configured for the best performance. However, it is also necessary to consider cases where this is unattainable. For example, in many cameras, the exposure time is automatically adjusted based on the ambient lighting. Other parameters are inaccessible, such as their internal clock, or are protected by firmware, which happens with most Android and iOS smartphones (Nguyen et al. [57]). Therefore, it is necessary to search for strategies to enable communications using those cameras, such as the blind estimation of link parameters and channel conditions and the equalization of the effects derived from selecting the non-optimal settings. Furthermore, accelerating the technology adoption in the industry requires these systems to be compatible with the actual methods and processes and to operate to a certain extent with the current systems.

The remainder of this chapter reviews the literature on the AI applied to OCC systems.

2.2 Artificial intelligence and optical camera communication

In recent years, the interest in using artificial intelligence to address complex challenges in OCC has grown, as shown in Figure 2.6. The first contributions in this field date back five years ago, up to the author's knowledge. Despite being a recent line of research, it shows great solidity and maturity since it uses tools developed and validated in other branches of science and engineering, especially in computer vision and digital signal processing. To have a better understanding of the current research interests, the contributions have been classified into four different categories: (i) source detection, (ii) modulation, (iii) equalization, and (iv) others. Figure 2.6 shows the distribution of the research according to the above classification. It should be noted

Table 2.1: Contributions related to AI and OCC.

Category	Link	Reference
Detection	GS	[59, 60, 61, 62, 63, 64, 65, 66, 67, 68]
	RS	[69, 70, 71]
Modulation	GS	[59, 72, 73, 74, 75, 76, 61, 64, 77, 66, 67, 71, 78, 68]
	RS	[79, 80, 81, 82, 83, 84, 70, 85, 86, 87]
Equalization	GS	[64]
	RS	[88, 89, 56, 90, 91, 92, 93]
Other	GS	[94, 95, 96, 97, 58]

that this distribution considers the fact that the same article belongs to more than one category.

As can be observed, the main interest lies in modulation techniques and source detection. Many of these works combine two independent models: one for source detection and one for source demodulation/decoding.

Table 2.1 shows the contributions classified by category and type of link (either GS or RS). All these works focus mainly on reception. Only Islam et al. [58] shift their attention to the transmitter. Their work aims to optimize the spectral efficiency of OCC systems in a vehicular environment. For this purpose, reinforcement learning techniques are used to learn the best decision policies based on adapting the modulation order of a quadrature amplitude modulation (QAM) signal and the relative speed between vehicles.

On the other hand, the "Other" section includes those works that use artificial intelligence for tasks that indirectly support data reception. For example, Li et al. (2019) [94] propose to use neural networks (NNs) to improve the estimation of a user's position in an indoor scenario. In their work, they identify LED beacons with known positions using GS techniques and then feed these coordinates to the network to accurately estimate the user's position. Teli et al. (2019) [95] use AI to predict the movement of a user's finger that occludes a transmitting LED array. In other terms, it focuses on predicting partial blockages in communications. Jeong et al. (2021) [96] use NNs marginally for position estimation using information extracted from accelerometers. They use optical beacons that update the accumulated error in this estimation to improve the accuracy. Zhao et al. (2022) [97] use NNs to estimate the human pose to indirectly ease the detection of transmitters located on their head.

The remainder of this section chronologically dissects the rest of the contributions collecting them in two subsections: (i) detection and (ii) modulation and equalization. It has been decided to gather the work related to modulation and equalization in a single subsection to improve understanding of the context since, in many works, the equalization is implicitly performed during demodulation. However, the distinction between demodulation procedures, which deliver data, and equalization, which deliver signal samples, is emphasized.

It should be added that since many papers combine two or more categories, they are either repeated as appropriate or, if they do not contribute significantly to a particular category, they will not be mentioned.

Detection

In the field of source detection, Nguyen et al. [59] proposed in 2017 the use of an intelligent receiving stage based on two independent NNs used for: (i) tracking and (ii) decoding GS transmitters in vehicular environments. Although they do not specify the network topology used for source detection, or its input parameters, they point out that the use of AI could improve, compared to classical computer vision techniques, the position estimation of light sources, even when they appear switched off, due to their cyclic blinking. Using several consecutive frames would allow the network to deliver accurate estimates. Moreover, this author proposed the design of a custom network that differs from the current trend in this research line, which generally consists of reusing models already validated in object detection tasks. These models return the ROI via bounding boxes, in which there is a high probability of containing the object of interest.

For example, Choi et al. (August 2019) [60] proposed to use an adaptation of the you only look once (YOLO) [98] version 3 [99] model for the detection of traffic lights operating as GS transmitters. The use of this network reduces the computation times while enabling multiple source detection. However, this system did not have a method for tracking sources, especially when they were switched off. Instead, it used the bounding boxes detected in previous frames.

This problem might be solved, according to Pham et al. (June 2020) [61], by using a hierarchical detection, where the cars' rear is first detected using YOLOv2 [100], and then the headlights. This way, even if either or both of their headlights are off, they can still be detected correctly. In addition, they proposed to track the source using an algorithm based on the Euclidean distance between the ROI's position detected in the current and the previous n frames. However, this technique might produce errors when cars are crossing. To solve this, cars use active signaling of the ROI. This signaling uses a low-rate link that operates simultaneously with a high-speed data link (hybrid scheme).

On the other hand, Sun et al. (May 2021) [64] developed a new model that combines the functionality of two preexisting models: detection and image enhancement. These authors stated that in vehicular environments, the relative motion between cars causes GS transmitters to appear blurred in the image. Hence, the proper data decoding requires to deblur the image previously. To this end, they proposed a feasible approach using a YOLOv5 [101] network for detection and a single blind motion deblurring network, DeblurGAN-v2 [102, 103], for image deblurring. However, when both models are cascaded, it leads to high computations and long delays. Therefore, they proposed a novel end-to-end network, D2Net, inspired by both models, which can achieve LED detection and motion deblurring simultaneously.

These previous works rely on GS transmitters. Rahman et al. (July 2021) [69] shifted their focus to RS links. This must face the consequences of reducing the camera's exposure, which reduces the receiver's sensitivity, producing dark images in which only the most intense light sources are perceived. Thus, the context of the image is lost, and the emitters are difficult to classify. To overcome this challenge, Rahman et al. proposed a two-stage detection in which: (i) all light sources present in the image are detected and (ii) then classified into transmitters using Support Vector Machine (SVM) models. This classification benefits from the fact that transmitters, contrary to non-flickering lighting systems, will appear composed of symbol bands due to the RS scanning.

Nevertheless, RS systems are strongly limited in the distance, so it is not surprising that for long distances, GS systems are preferred. For example, Takano et al. (May 2022) [65] achieved the detection of three independent RGB LEDs mounted on a drone flying at a distance of up to 300 meters, using YOLOv3. Note that at these distances, the LEDs are hardly visible in the image, so high-magnification lenses are needed (x20.6 used in this work). This drastically reduces the field of view (FoV) of the camera, which makes the system more sensitive to movement.

Nguyen et al. (August 2022) [71] adapted YOLOv5 for RS transmitter detection. Compared to Rahman et al., the novelty is that they train the network using images captured with different exposure times. In addition, they performed a small preliminary analysis of how exposure time affects the received signal in terms of SNR and bandwidth.

On the other hand, Cheng et al. (August 2022) [66], from Sun's team at Jilin University (Changchun, China), introduced a new architecture based on the GhostNet [104] model that significantly improves inference times, efficiency, and accuracy over the previous reported D2Net, and recent YOLO versions. However, this new network loses the deblurring functionality offered by D2Net. The authors stated bit error rates (BERs) below the forward error correction limit (FEC) limit for distances up to 4 meters in real traffic scenarios.

Finally, Kim et al. (November 2022) [68] used the YOLOv4 [105] model to detect color bars on displays.

Modulation and equalization

As mentioned above, works related to modulation, coding, or equalization in OCC links focus actually on the AI-assisted demodulation, decoding, or equalization of the received signal. The demodulation pipeline for translating signal samples into data is generally divided into three parts: (i) symbol sampling, (ii) feature extraction, and (iii) classification using AI techniques. The equalization pipeline, on the other hand, will ingest and deliver signal samples.

In 2017, Nguyen et al. [59] proposed the use of NNs for symbol classification, using GS transmitters and a novel modulation called spatial-2-phase-shift-keying (S2-PSK). This encoding maps each bit into the phase difference between two synchronized LEDs. When both LEDs are in phase, a bit 1 is transmitted, and 0 otherwise. During decoding, the network classifies the intensity values of both LEDs into their corresponding bits. In this way, non-linearities in the received intensities are exploited by the NN, providing a more accurate and robust (against noise) classification. Although the authors do not mention their causes, these non-linearities might be related to the camera's gamma, the LED non-linear behavior (emitted power vs. current), and the channel response, among other factors.

In December 2018, Li et al. [79] published the use of NNs for demodulating colored symbols (6 RGB symbols), using RS transmitters. Each symbol comprises three independent pulse width modulation (PWM) signals, one for each color, with different frequencies, duty cycles, and phases. In the demodulation stage, features related to the color stripes perceived in the image, such as their number and perimeter, are utilized.

Pham et al. (June 2019) [72] presented a dimmable S8-PSK (DS8-PSK) AI-assited decoding. This modulation uses eight reference LEDs and eight data LEDs to transmit in the phase difference eight symbols. In this case, the extracted intensities of each LED are used as the input parameters of the NN. This work aims to evaluate the performance of NNs at decoding GS transmitters in vehicular environments when the image is affected by variable blurring. This blurring is associated with multiple phenomena: vehicle vibration and movement, road defects, and atmospheric conditions (rain, snow, haze, ...), among others. Although this work needed an in-depth characterization of this blurring (its sources and distribution), they proposed a primary method for generating synthetic samples based on different levels of blur and additive white Gaussian noise (AWGN) noise. These samples were used for training and validating the network. However, there was no actual validation using authentic images. In [74], a related work, they proposed an AI error correction scheme. These associated works were condensed in [61] (June 2020).

Zhong et al. (August 2019) [73] proposed to detach the camera optical lenses to eliminate the need to detect GS sources in the image, as the LED light will flood the IS entirely. Consequently, the camera loses its spatial diversity capability and functionality as an imaging device. In this work, they relied on NNs to decode the ON and OFF states of OOK signals by using the luma, Y, and chrominance, Cb/Cr, values.

Liu et al. (November 2019) [80] proposed the use of convolutional neural network (CNN) for the assisted decoding of RS transmitters. These networks utilized the convolution operation for extracting features from input samples. Each convolutional layer uses different kernels whose coefficients are adjusted during training. In this work, the input corresponds to a single column of the image of MxN rows of pixels. Where M is the number of adjacent symbols, and N (referred to as oversampling ratio) is the number of pixels (i.e., samples) per symbol. Hence, the input has dimensions of (M-N)x3, where 3 corresponds to the RGB channels. The features extracted by the CNN are delivered to a multi-layer perceptron (MLP) that classifies them into one of 8 possible symbols. This network exploits M-1 adjacent symbols to classify one. Therefore, the complete demodulation requires iterating the operation of the network using M-N input segments until it traverses the entire image. The authors stated that this system consistently improves BER compared to the use of one-dimensional (1D)-NNs or other classical algorithms, such as the double-equalization proposed by the same authors [106].

Alfarozi et al. (November 2019) [75] evaluated and compared several different CNNs models (ResNet, DenseNet, VGG, and a proprietary adaptation that the authors called MiniDenseNet) to perform the decoding of square wave quadrature amplitude modulation (SW-QAM) signals, sent with GS transmitters. This assisted decoding is intended to be robust against interference between nearby sources (given that two LEDs are used in the transmission), noise in the communication, and other parameters that may deteriorate the signal. This can be achieved with the training with thousands of representative samples. In addition, they proposed the use of Spatial Transformer Networks [107] to deal with the movement of LEDs in the image. These networks learn to perform spatial transformations on the input images that improve the invariant geometry of the model. As a result, this network copes with relative camera movements and avoids the need to use techniques for detecting the position of LEDs in the image area. The authors demonstrated that using STN and MiniDenseNet improves source decoding in moderate motion environments without needing prior source detection or tracking algorithms.

In February 2020, Younus et al. [88] proposed the use of NNs for exposure effects equalization in RS links. The neural network performance was validated against

Manchester-encoded OOK signals. It proved to recover the transmitted information with a bandwidth higher than the cut-off frequency imposed by the exposure time. As a result, a bandwidth improvement of up to 9 times is registered (compared to non-equalized reception chains. In terms of temporal parameters, it can equalize signals whose symbol time is approximately four times shorter than the exposure time, equivalent to the mixing of 4 consecutive symbols. However, this procedure is evaluated under significantly high SNR conditions. In addition, training samples must be acquired under the same conditions and using the same instruments (i.e., camera and transmitters) as expected during the link's deployment and operation. In short, this ad-hoc solution can hardly be applicable in a real environment. Subsequently, in April 2021, the authors used the same procedure with constant 4-pulse amplitude modulation (PAM) signals [56]. On this occasion, the improvement is less significant, given the complexity introduced by the multilevel modulation. The bandwidth improvement is lower than three times. The exposure time must be up to 1.3 times the symbol time.

Pepe et al. (April 2020) [76] used CNNs to decode 16-CSK signals transmitted with screens acting as GS transmitters. The purpose of these networks is to overcome distortions and disparities in the camera's color representation with the screen set at low brightness.

Yu et al. (August 2020) [82] used CNNs at decoding RS transmitters to overcome the distortion effects produced on the generated bands due to motion. These authors evaluate lateral movements, which produce skewness in the LED projection, and horizontal movements, which expand or shorten the height of the bands, depending on whether the action follows the scanning direction of the sensor or backward, respectively.

Ahmed et al. (October 2020) [89] used NNs to correctly align the bands generated in RS links. This misalignment occurs when the transmitter moves laterally in the image. The authors reported improvements in the BER using this AI-assisted warping procedure. In addition, they found that the camera's exposure time impacts the BER, concluding that it is related to ISI. To avoid it, they set the exposure time to a relatively short duration (1.2 times the symbol time).

Sun et al. (May 2021) [64] cascades their customized D2Net network used to detect sources (mentioned in the previous subsection) with a decoding stage based on CNNs. The decoding is conducted independently on equal patches of the LED matrix. The same procedure will be used later by Cheng et al. (August 2022) [66]. In this case, the differences between these two papers are related to detection, as discussed in the previous subsection.

Lin et al. (June 2021) [83] and later Chow et al. (July 2021) [84] used conventional NNs to facilitate the decoding of RGB panels whose backlight is OOK modulated. Together with preliminary background removal algorithms, these networks aim to reduce the interference caused by the projected image in front of the panel.

To continue with the chronological narrative, it is at this point, in July 2021, that one of the contributions of the author of this thesis ([21]) detailed in Chapter 5 is published. This contribution was published three months after Younus et al. [56] publication related to exposure equalization in RS systems. The discussion of the differences is reviewed in the conclusions (Chapter 6) and include: (i) synthetic training combined with real validation versus ad-hoc real training, (ii) a two-dimensional (2D)-CAE versus a 1D-NN that exploits the redundancy in nearby columns, (iii) added denoising capabilities, and (iv) higher efficiency and fewer network parameters. It should be remarked that more work will subsequently be developed regarding exposure equalization, albeit done implicitly during demodulation, such as the work of Lin et al. [90], He et al. [85], and Ling et al. [86] This demonstrates the community's interest in this line of research. Despite being later works, the differences will be detailed in the conclusions (Chapter 6).

Park et al. (August 2021) [77] proposed to use a CNN that ingests the entire image containing a GS LED-matrix transmitter to decode it. The fact that it decodes the whole image avoids the use of detection algorithms. However, this system lacked a mechanism to determine whether there were none or several sources in the picture.

Lin et al. (September 2021) [90] proposed the demodulation of 4-PAM signals in RS links. This work was similar in content and results to the previous work by Younus et al. [56]. In technical terms, they propose the classification of a group of PPS pixels, where PPS is the number of pixels per symbol, with labels corresponding to each symbol. These labels are actually numerical values corresponding to the signal levels. This work registered up to 4 times bandwidth improvement, with exposure times reaching up to 2.5 times the symbol time. It should be highlighted that the input size in this network is significantly smaller. Instead of using an input with dimensions close to the packet size in pixels, it uses inputs with a size equal to the PPS, which improves the efficiency of the network. This work will be presented again a month later at a WOCC 2021 conference [91]. Finally, in March and May 2022, they used the same procedure but with transmitters based on light-diffusing fiber [92, 93]. In this work, they also evaluated the effect that the relative angle between the transmitter and the camera has on reception.

He et al. (March 2022) [85] also proposed to use an NN for demodulating 4-PAM signals, using RS transmitters. However, in this case, each 4-PAM symbol is generated by repeating several cycles of the same PWM signal whose period is significantly shorter than the exposure time. In this way, at reception, each pixel accumulates the irradiance of several PWM cycles accumulating the corresponding charge of a PAM symbol. On this occasion, the network input has dimensions of 2N0x1x1, where N0 (referred to as the symbol resolution) is the number of pixels per symbol (i.e., rows of pixels), and 1x1 corresponds to a single column in grayscale (one channel). The network's output, M0, has one neuron per each possible combination of two slots (two transmitted input symbols) and four possible symbol choices. It should be noted that the results obtained are within the records of the previous works in terms of longer exposure times. The exposure times were set to 40 microseconds and the symbol time to 93 microseconds, equivalent to exposure times of about half the symbol time.

Ling et al. (April 2022) [86] published a paper in which they combined CNNs and recurrent neural network (RNN) for decoding 8-color shift keying (CSK) signals in RS systems. In this case, RNNs manages to capture the dynamics of the sequences (i.e., the transmitted symbols) through recurrent connections, which can be thought of as cycles in which the internal state of the network is fed back. In this way, they are capable of acquiring short and long-term memory. Long-short-term memory (LSTM) networks achieve the latter case. This work also focused on mitigating exposure time's effects on the signal, among other factors. However, despite identifying that exposure is detrimental, they still failed to analyze or quantify its impact and did not report the exposure time setting.

In June 2022, another contribution of this thesis (Chapter 5) was published [22]. This work proposes estimating fundamental parameters in RS links: the exposure time and the signal clock. The contributions and implications of this work will be detailed in Chapter 5.

Nguyen et al. (August 2022) [87] used, on the other hand, NNs for estimating the preamble position in constant-OOK signals in RS systems and compared them with matched filters. The purpose was to address the distortions introduced by the mobility of the sources. In addition, they experimentally demonstrated improvements in the SNR by increasing the camera's exposure. At the same time, although without experimental or theoretical validation, they pointed out that increasing the exposure time reduces the bandwidth. Moreover, they did not analyze the behavior of the network versus exposure time but adjusted and fixed it so it does not affect the receiver bandwidth. In August 2022, they published the same procedure in [71] for preamble estimation in orthogonal frequency division modulation (OFDM) schemes.

Again, Sun et al. (October 2022) [64] proposed a system based on CNNs for decoding an LED matrix operating in GS mode. In this case, they analyzed how increasing the camera sensitivity can lead to saturation of the source and increased interference between nearby sources. The use of CNNs could alleviate these interferences.

Finally, Kin et al. (November 2022) [68] used CNNs for color barcode decoding, addressing disparities in the color representation by the camera.

Chapter 3

Theoretical framework

This chapter explores the technical and theoretical considerations for using RS cameras as optical receivers. The chapter starts by discussing the RS acquisition mechanism and how it relates to data acquisition, focusing on the effects of the camera's exposure time on the signal. Other parameters that may affect signal reception, including frame rate, analog gain, and physical world parameters, such as transmitter's size, camera optics, and distance, are also discussed in terms of their influence on the communications link. On the other hand, it also introduces the AI fundamentals required to understand the rest of this thesis.

3.1 Rolling shutter cameras

RS cameras scan the scene sequentially, row by row of pixels. The interval between the beginning of the scan of one row (the activation) and the subsequent row has taken several names in the literature, including row shift time, row interval, and row sampling time. The latter terminology will be used for convenience due to its equivalent relationship with the sampling time terminology used in communication systems.

This acquisition mechanism is representative of complementary metal-oxide semiconductor (CMOS)-based cameras. On the other hand, charge-coupled device (CCD) cameras generally use a GS acquisition, which simultaneously exposes all the sensor pixels. Despite this strong relationship between the fabrication technology and the acquisition mechanism, GS-based CMOS cameras can be found in the market, although it is less frequent.

Figure 3.1 shows the timing diagram of the RS acquisition. In this diagram, the frame time corresponds to the total invested time in capturing a frame, while the acquisition time is the interval between the start to exposure of the first row and the readout of the last one. The rest of the frame time is used internally: (i) to accommodate a variable (sometimes adjustable by the user) exposure and (ii) to guarantee the availability of hardware resources for subsequent captures.

The RS acquisition starts when the camera opens its shutter. Suddenly, all pixels are exposed to the incoming light and start accumulating charge. This seems to contradict the sequential scanning mentioned before, but the key that solves this contradiction is the following. Immediately after the shutter is opened, the first row of pixels is activated by an electronic signal. This signal discharges the charge previously accumulated in the pixel capacitors, thus initiating the actual exposition for that row. Hence, the activation time, or reset time, t_{reset}, is the time necessary to discharge the

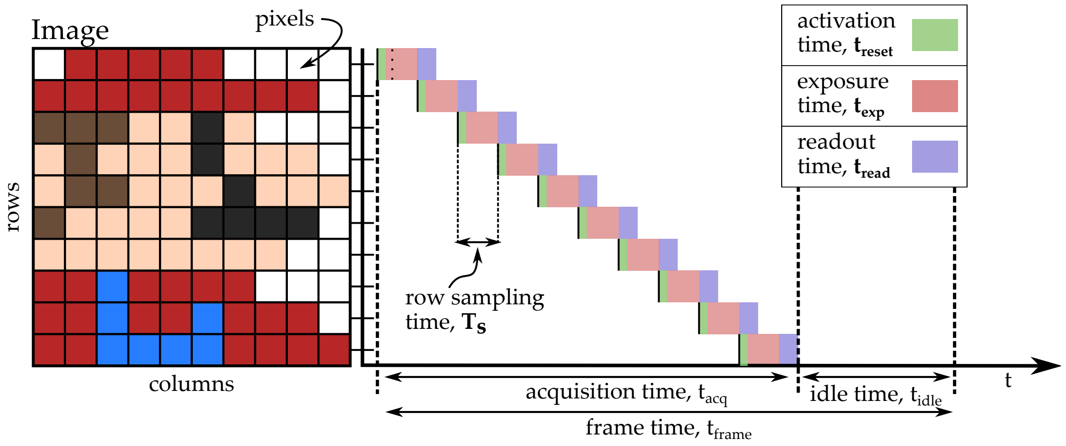


Figure 3.1: Timing diagram of the RS acquisition mechanism.

previously accumulated charge on the photodiode (by biasing it with reverse voltage). Afterward, the rest of the sensor rows are activated sequentially. As introduced at the beginning of this chapter, the interval between activations is known as the row sampling time, t_{row}. After the activation, each row is exposed to light during a preselected time, the exposure time, t_{exp}. For convenience, the exposure time in this graph is considered to be the same as the row sampling time, which corresponds to the ideal sampling conditions. Otherwise, exposure effects start to be noticeable. Those effects will be discussed at the end of this section. Later, once the exposure has been finished, the sequential readout of the pixel values proceeds pixel by pixel for each row. The readout time, $t$read, corresponds to the readout time of just one row. This readout process varies depending on whether passive or active pixels are involved. In passive pixels, the charge is transferred to a floating diffusion amplifier (FDA) common to the entire column, which converts the charge into a voltage with a charge-voltage conversion gain. Finally, this voltage is amplified at the general output amplifier and quantified by the analog-to-digital converter (ADC). In active pixels, the charge-voltage conversion is performed at the pixel level, and the voltage is transferred to the general output amplifier using source follower amplifiers (SFAs). This procedure is repeated until all the pixels in a row and all the sensor rows have been read.

As it was discussed in the introduction (Chapter 1), the RS mechanism can lead to several potential issues in images, including:

- Distortion and blurring (Fig. 3.2): When an object is moving quickly through the scene, the camera may capture it at different positions at different times, leading to distortion or ”skew” of the object. This can be especially noticeable in images of fast-moving objects or when the camera moves quickly.
- Banding (Fig: 3.3[1].): If there is a flickering light source in the scene, such as office lights, a television or a computer monitor, the RS can cause bands of different light intensity to appear in the image. This is because the camera captures the light at different times, leading to uneven exposure.

The last phenomenon suggests using RS cameras for sampling optical signals, which produces a recognizable pattern of different illuminated bands or stripes in the image as

---

[1] Images obtained from https://www.dpreview.com/forums/post/57566427 and https://www.dpreview.com/forums/post/60552191

Figure 3.2: A de Havilland Canada Dash 8 propeller, with severe RS distortion from a Pixel 3 camera by Dicklyon (Wikipedia).

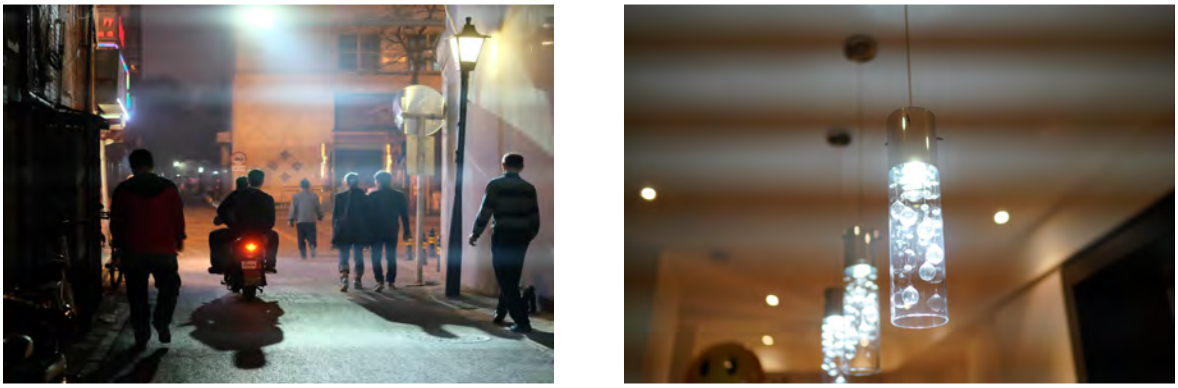


Figure 3.3: Examples of RS banding (DPreview).

shown in Fig. 3.4. Each band represents a potentially transmitted symbol. On top of the right diagram are displayed the temporal variations of a OOK modulated squared signal, with "ON" and "OFF" illumination states. The image on the left represents the pixel values obtained after the capture.

This graph illustrates direct relationships between the camera's parameters and conventional communications receivers. First, it shows that the temporal signal variations are captured along the horizontal domain of the image. Therefore, the RS mechanism establishes a direct relationship between the pixel position (space) and the sampling instant (time). This space-time relationship has profound consequences on the link, as will be discussed in the following subsections. Second, it also shows that the columns contain highly correlated signal samples. This is because the pixels for each row are exposed simultaneously. Hence, each column represents redundancy samples of the transmitted signal. In summary, the rows sample the signal, and the columns introduce redundancy[2].

Third, the camera's row sampling time, t_{row}, corresponds directly to the receivers receiver's sampling period (or period), T_{S}. Conventional user-ready cameras typically sample the optical signal every 8 to 10 microseconds. Therefore, the receiver's sampling rate, F_{S} ($F_{\text{S}} = 1/T_{\text{S}}$) can reach up to 125 kHz. In super slow motion cameras, even

---

[2] It should be remarked that the sampling dimension changes if the scanning is performed in columns instead of rows.

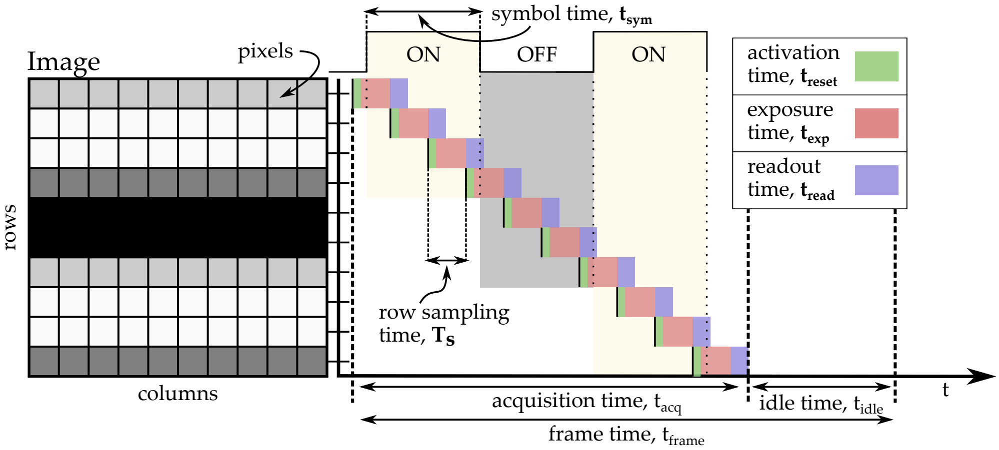


Figure 3.4: Timing diagram of the RS acquisition of an OOK signal.

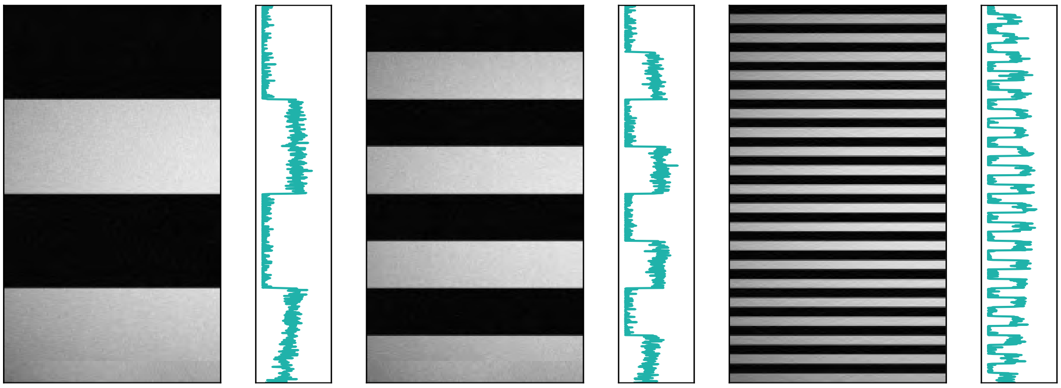


Figure 3.5: Image examples for increasing transmitter's baud rates.

higher sampling rates can be achieved (up to 5 times higher than conventional cameras, i.e., 500kHz). As in any generic communication system, the transmitter's baud rate in OOK signal, which corresponds to the inverse of the symbol time, t_{sym}, must be at most half of the sampling frequency to comply with Nyquist-Shannon sampling theorem.

Figure 3.5 provides some examples of increasing transmitter baud rates.

As seen, increasing the transmitter's frequency decreases the band size (in pixels). Ultimately, each band's height (in pixels) is related to the number of symbol samples. The latter is the ratio between the symbol time and the sampling period. In this thesis, as the focus in RS OCC are images, this ratio is called the number of pixels per symbol (NPPS). The reason behind not calling it just samples per symbol is that generic communication systems consider one receiver sampler to provide all the signal samples. In RS systems, each signal sample comes from independent pixels, which can be understood as independent samplers, only related by the time they are activated. This will be further detailed in the following subsection. Finally, it should be mentioned that, theoretically, despite working with discrete systems, the NPPS may be a non-integer value.

The sampling time is upper limited by the capture time. In other words, the sampling frequency must be considerably higher than the frame rate (or frames per second) fps. Generally, the higher the capture rate, the greater the capacity to sample the signal faster. Another factor influencing the sampling rate is the number of rows, i.e., the vertical resolution (in pixels), h_{pixels}. The higher the vertical resolution, the shorter the sampling time should be. Many literature texts established a direct relationship between the sampling rate, the frames per second, fps, the exposure time, t_{exp}, and the vertical resolution, h_{pixels}, as expressed in Eq. 3.1.

$$F_{\mathrm{S}} = \mathrm{fps} \cdot h_{\mathrm{pixels}} - \frac{1}{t_{\mathrm{exp}}} \tag{3.1}$$

Although the computed sampling rates using this equation approximates well as the camera's fps increases, the reality is that they are generally significantly higher. This is because this equation implicitly considers that the acquisition time equals the frame time, which is different from what is shown in Fig. ef fig:diagram. This is because the cameras use idle times to (i) accommodate variable exposure times and (ii) wait for resources.

Consequently, the IS will remain blind to the transmission during these idle times resulting in an interrupted reception. To address this challenge, channel coding techniques must be applied to ensure reliable communications. This will be discussed later in this section, along with analyzing the camera's exposure time. But before entering this discussion, it is mandatory to analyze how the light projects into the IS to generate the intensity distribution profile perceived in the images.

3.1.1 The image intensity distribution profile

The images shown in Fig.3.5 result from captures of uniform illuminated surfaces (such as LED flat panels). Hence, the light source's irradiance uniformly floods the entire IS, generating an intensity distribution profile that is practically uniform for the "ON" and "OFF" states.

However, this condition of equally distributed illumination is rarely met. The image's intensity distribution profile hereafter referred to as the intensity profile, depends on a multitude of link parameters, including the size and shape of the light source, its irradiation pattern, its relative distance and orientation to the camera, the presence of obstacles or other objects that blocks or reflect the light, among many others.

Figure 3.6 shows two examples (for increasing distance) of the simulated reception of an OOK modulated signal using a synthetically generated 2D Gaussian distributed intensity profile.

As can be seen in this figure, the intensity profile varies both in the vertical (rows) and horizontal (columns) dimensions (data and redundancy domains, respectively). The graphs on the right show the row intensity profile of three columns. They manifest that the signal strength is significantly higher in the source's projection and drops sharply outside. With maintaining rigor, it can be concluded that communication quality depends considerably on the source's image size (in pixels). This statement is conveniently used to make introducing complex concepts in subsequent paragraphs easier. It remains valid to the extent that the luminous profile drops sharply beyond the source's borders, which happens in most cases (especially when using flat LED panels). However, it should be noted that the signal can be recovered anywhere in the image, even from reflections in walls and mirrors, ultimately where SNR is higher than the threshold established by the system's modulation scheme.

As it was aforementioned, the size of an object in pixels, as captured by a camera, is affected by several parameters, including:

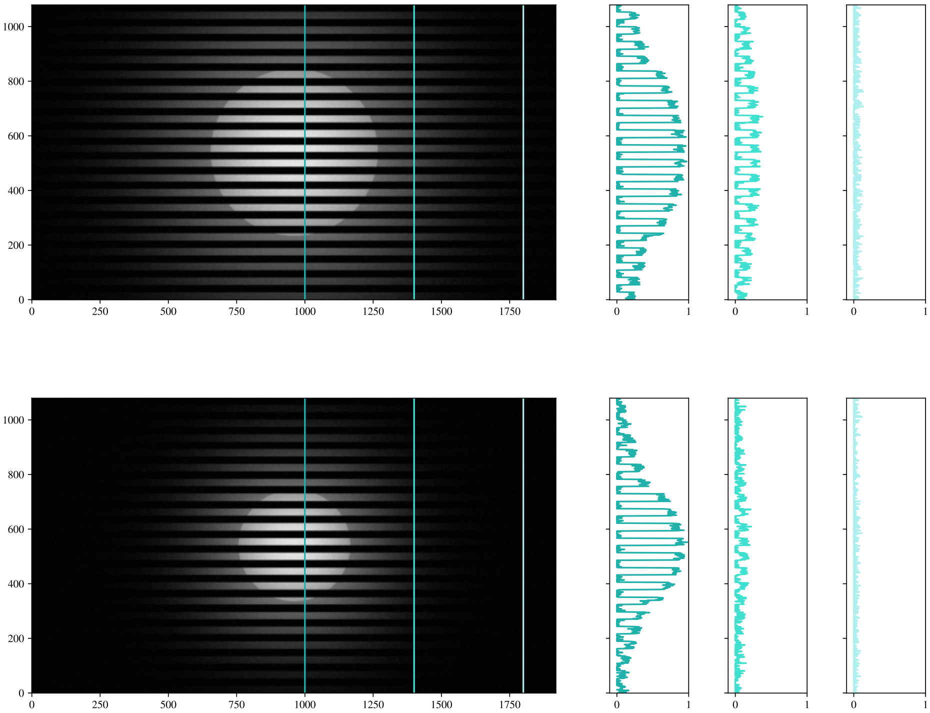


Figure 3.6: Examples of the simulated reception of square signals using a synthetic 2D Gaussian distributed intensity profile. In the top image, the source is closer to the camera.

- The physical dimensions (in centimeters) of the object itself (H, the height, and W, the width, and, D).
- The distance between the object and the camera, d, and its orientation.
- The focal length of the camera lens, f, and the sensor size, s, (in micrometers). The focal length of a lens is a measure of its ability to focus (redirect) light on the image sensor. It is typically expressed in millimeters and is a key factor along with the sensor size in determining the angle of view (AoV) of a lens, which corresponds to the angular extent (in degrees) of a given scene that the camera can capture. A lens with a short focal length, such as a wide-angle lens, has a wide AoV and can capture a larger portion of the scene. A lens with a long focal length, such as a telephoto lens, has a narrow AoV and can magnify distant objects. The angle of view can be expressed by Eq 3.2.

$$AoV = 2 \cdot tan^{-1}\left(\frac{s}{2 \cdot f}\right) \tag{3.2}$$

  The AoV should be distinct from the FoV. Although these related terms have been used interchangeably in technical writing, they do not refer to the same concept. The FoV measures the size (typically in meters) of the scanning area in the object plane for a certain distance and AoV. Equation 3.3 relates both quantities. Finally, it should be highlighted that both values can be referred to as either the vertical or horizontal dimension of the camera.

$$FoV = 2 \cdot d \cdot \tan\left(\frac{AoV}{2}\right) \tag{3.3}$$

- The sensor resolution (the number of pixels on the image sensor).
- Defocus. Defocusing a lens, causing the focus point to fall in front of or behind the object, might lead to making objects in the image appear larger. The defocusing effect on the object's size will depend on the amount of defocus and the distance between the camera and the thing. Objects closer to the camera will be more affected by defocus than objects further away.

Ultimately those parameters condition the irradiance map in a discrete sampling instant. The final image results from the time evolution of consecutive irradiance maps during the frame capture, which may vary over time due to other circumstances rather than the actual transmission. For example, because of: (i) the dynamics of the scene, such as the relative movement between the transmitter and the receiver, (ii) the presence of external active interference agents, or (iii) the environmental conditions, such as rain, snow, fog, among others. In short, the dynamic nature of the link further complicates the task of accurately demodulating the transmitted data using a RS camera.

This non-uniformly distributed intensity profile greatly impacts communications and confers the following characteristics on the link:

- Time-variant. The irradiance reaching each pixel for each symbol varies depending on its location within the image sensor. For example, as shown in Fig. 3.6, the top pixels accumulate less charge than the middle pixels, although they sample equivalent “ON” and “OFF” illumination states. On the other hand, the sequential acquisition of RS cameras relates both the pixel position (space) with the signal sampling instant (time). As a consequence, the signal's sample strength is influenced by the time the sample was taken. Overall, the system behaves as a time-variant system, increasing the complexity at reception. Solutions adopted to address this challenge includes: (i) applying row equalization stages to adjust the gain for each pixel, using various techniques, including linear equalization, decision feedback equalization, and maximum likelihood sequence estimation, (ii) applying global equalization techniques that exploit the redundancy of nearby columns, and (iii) identifying those regions where the signal strength is more uniform, e.g., outside the source's projection (Fig. 3.6 middle graphs).
- Windowed reception. To proceed with decoding the transmitted data, it is necessary to ensure that the SNR is above a threshold imposed by the modulation scheme. For this reason, it may be required to discard those samples that are too weak to be accurately demodulated. For example, in Fig. 3.6, the further top and bottom pixel rows might be discarded. Recalling the space-time relationship in RS systems, this means that the receiver operates in time windows of varying length, resulting in a windowed reception. This implies that it is necessary to identify those ROIs within the image where the data can be decoded and to devise appropriate channel coding techniques to protect information from losses. To address this challenge, the solutions that the community has widely adopted include: (i) adapting the packet length and sending them repeatedly or (ii) adding a significant amount of redundancy to the transmitted data. By sending variable-length packets repeatedly, it is possible to increase the probability that the transmitted data will fall entirely within an active time window. Similarly, by adding redundancy, it is possible to improve the reliability of the

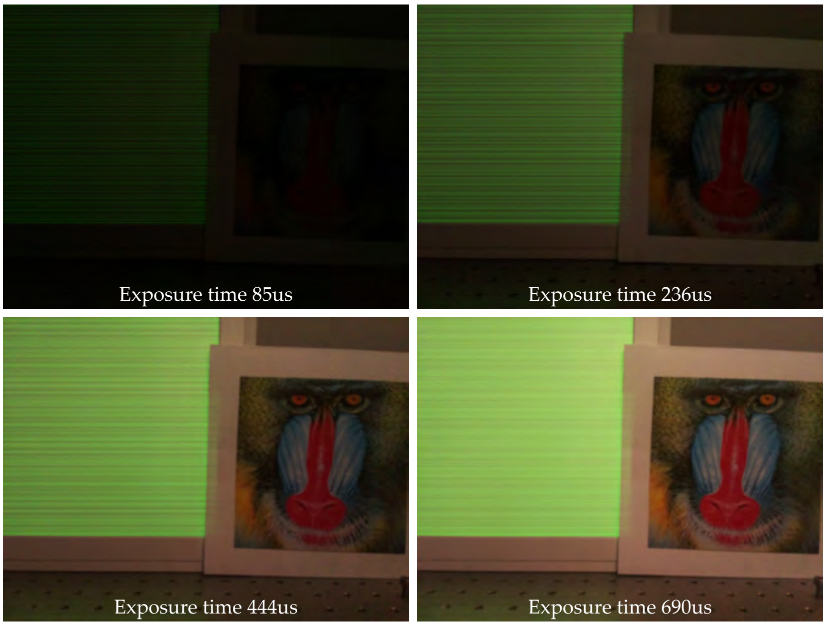


Figure 3.7: Examples of images for increasing exposure times.

communication link at the cost of increasing the system's complexity. Ultimately, this windowed reception further restricts the attainable data rate.

In conclusion, the RS systems are time-variant, and the active reception happens in variable-length time windows. The link reliability increases as the irradiance reaching the IS become more uniform. Finally, data decoding takes place in regions where the SNR is above the modulation's imposed threshold. Moreover, as discussed in the following section, the SNR can be improved by adjusting the receiver's sensitivity.

3.1.2 Receiver's sensitivity and exposure time

The receiver's sensitivity is directly related to the camera's sensitivity, measured with the ISO 12232:2019 standard value. The higher the ISO value, the more sensitive the camera is to light. The parameters influencing the receiver's sensitivity are (i) the analog and digital gains and (ii) the exposure time.

Cameras adjust their sensitivity to light by applying analog and digital gains to the analog signal coming from the pixel before and after it is digitized by the ADC, respectively. Increasing both gains can improve the image quality in low light conditions but also introduce noise into the image, particularly the digital gain. In addition, two additional digital gains are used to adjust the relative sensitivity for each independent channel in RGB cameras. Although being out of the scope of this thesis, It should be mentioned that the presence of a Bayer optic filter in cameras to split the incoming irradiance in different wavelengths have an impact on OCC systems based on colored modulations. The misalignment between the transmitter's spectrum and the Bayer filter response might lead to potential ISI between channels. This ISI might be equalized prior decoding as proposed in a previous work [47].

On the other hand, the exposure time refers to the time each pixel in the image sensor remains exposed to light. The longer the exposure, the higher the sensitivity, as the pixels accumulate more radiant energy from the scene. Figure 3.7 shows images captured for increasing exposure time values.

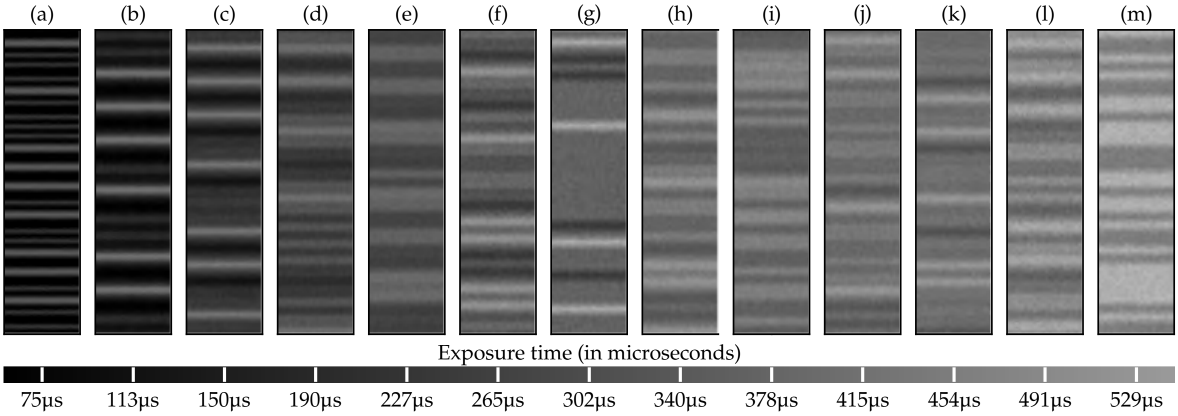


Figure 3.8: Examples of the same transmission for increasing exposure times.

As can be seen, the mandrill in the bottom left picture is discernible, as the exposure time is the highest in this setup. However, increasing the exposure time has a detrimental effect on the signal. To improve the visualization, Fig. 3.8 shows some samples cropped from the source projection. These samples show the effects on the received signal as the exposure increases. It should be emphasized that the same symbol sequence is transmitted in all the samples. Therefore, the differences between the images derive mainly from the exposure time and, to a lesser extent, from a slight shift in the symbol location due to the lack of synchronization between the transmitter and the camera.

As the exposure increases, it can be seen that consecutive symbol bands appear to blend together. This is because each pixel remains exposed while transmitting several consecutive symbols, accumulating a variable charge depending on their corresponding irradiance. Consequently, increasing the exposure time produces an ISI in the signal. Figure 3.9 shows the complete acquisition diagram derived from Fig.3.4, including the exposure time. To simplify the visualization, the activation and readout times have been removed. It can be seen that as the exposure time increases, more symbols are mixed.

The degree of impact of the exposure can be expressed in terms of the ratio between the exposure and the symbol times. In this thesis, this ratio is called exposure-to-symbol ratio (ESR). As an example, in Fig. 3.9(b), the exposure time is twice as long as the symbol time, and hence the ESR is 2. The higher the ESR value, the more detrimental impact on the signal.

Consequently, the sequential exposure of RS cameras acts as a low-pass filter, reducing the receiver's bandwidth considerably. Figure 3.10 shows the normalized frequency response experimentally measured for increasing exposure times. As discussed in Chapter 5, the receiver's 3 dB cutoff frequency, f_{cutoff} imposed by the exposure time can only be estimated numerically. However, to have reference values for the bandwidth reduction, the first null frequency, f_0, corresponds to the inverse of the exposure time, and the 2-dB bandwidth, to the first null frequency divided by 2. Therefore, the impact of the exposure time is neglectable for ESR equals 0.5. For this reason, many researchers choose to reduce the exposure time to a minimum, which is only sometimes possible. The consequence of this strategy is that only the active light sources and those regions with a higher intensity will be discernible in the images.

This sequential acquisition can be modeled as a two-stage system with (i) a filtering and (ii) a sampling stage (Fig. 3.11(c)). The filtering stage acts directly on the

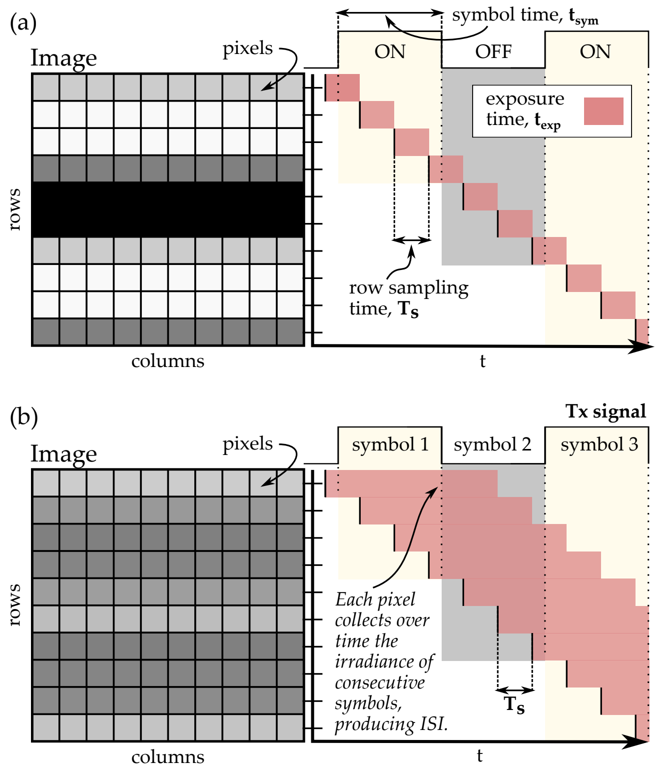


Figure 3.9: Timing diagram of the RS acquisition with variable exposure time.

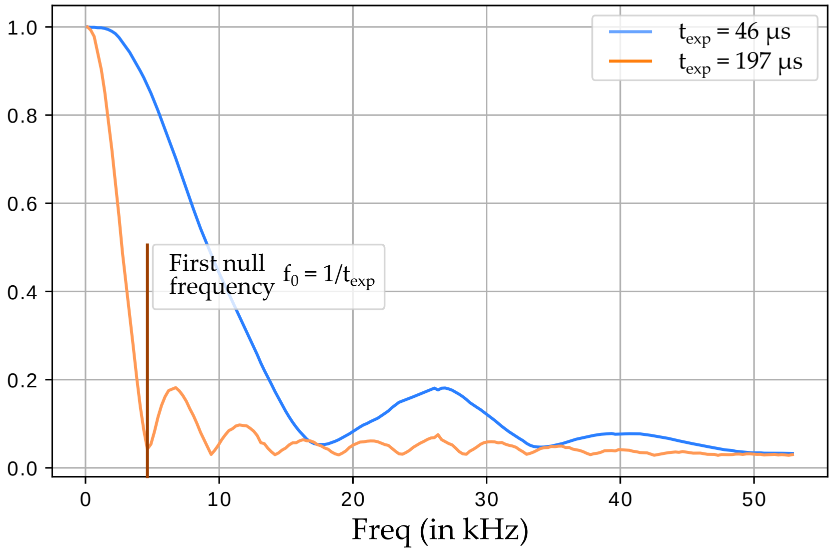


Figure 3.10: Normalized frequency response for different exposure times.

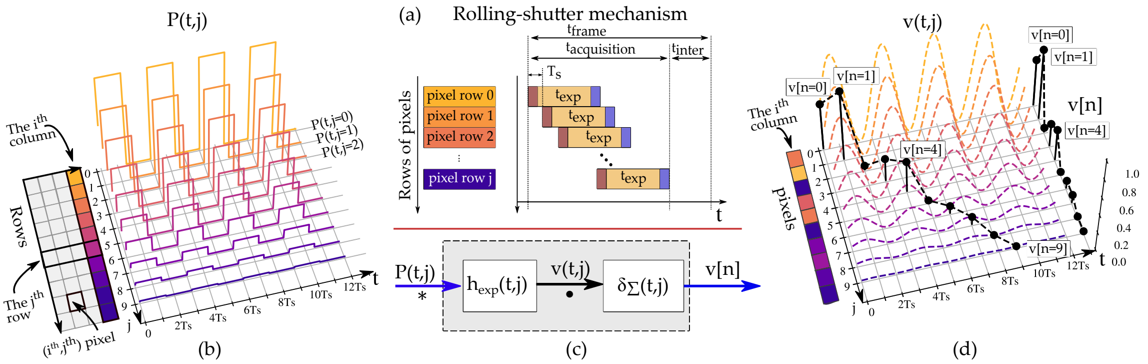


Figure 3.11: RS acquisition mechanism. (a) RS timing diagram. (b) Normalized optical power $P(t,j)$ reaching each pixel of the i-th IS column. (c) RS system modeling: impulse response, h_{exp}, and sampling function, δ_{Σ}. (d) Filtered curves, $v(t,j)$, and the discrete signal, $v[n]$ after sampling.

incident irradiance samples, $P(t,j)$ for each pixel (Fig. 3.11(b)). Note that the pre-filter sampling rate is considerably higher than the row sampling rate of the camera. Generally, an oversampling factor of 100:1 or 1000:1 is used, which means that 100 or 1000 temporal irradiance samples are used to compute each pixel intensity values, $v[n]$ (Fig.3.11(d)). As seen in Chapter 4, in the synthetic generation, the filtering step corresponds to a sliding uniform filter with a unit shift that keeps the original dimension (number of samples) by adding zero-padding if necessary. The size of this uniform filter is related to the exposure time. The result of this stage is time-averaged irradiance levels for each pixel, $v(t,j)$ (Fig. 3.11(d)).

Then, all filtered signals (one per pixel) are sampled, one at a time, depending on the oversampling factor and the camera's row sampling rate. Therefore, this procedure associates the k-th pixel irradiance value with the sample of the k-th time-averaged irradiance signal, $v(t, j = k)$, based on the sampling instant, T_s (i.e., $v(t = k{\cdot}T_s, j = k)$). Finally, the pixel's irradiance value is multiplied by a scaling factor based on the analog and digital gains, the loading capacity of the photodiodes, and the exposure time, to obtain the intensity values for each pixel.

This procedure can be used for the synthetic generation of image samples, but it is rather complex. First, computing the irradiance samples for each pixel is an unmanageable task. Considering only one row of pixels, 1080 filtered signals are required for an HD camera (with a resolution of 1920x1080). The length of these signals depends on the simulation time (i.e., the frame capture time) and the sampling rate, which must be 100 or 1000 times higher than the camera's sampling rate. More than 1M samples are required to simulate the reception of just one column. Moreover, this computation gets worse when there is high dynamic content in the scene, e.g., sources are moving, or there are interferers that contribute detrimentally to the final irradiance. To speed up the simulation, several conditions can be established to reduce the generation times considerably.

The first condition is that the irradiance only varies due to the usual operation of the transmitter (i.e., flickering) and not due to relative movements between sources or the action of interference agents.

The second condition assumes that each pixel's behavior is linear with respect to the incident irradiance and that the irradiance reaching each pixel can be modeled as a linear function of the total irradiance reaching the lens surface. If so, a significant computational reduction is achieved since only a single irradiance signal is required, which will be weighted accordingly after the filtering stage depending on the location of each pixel and the image intensity distribution profile.

The third condition ensures that the irradiance is uniformly distributed in the IS. In this way, the irradiance samples for each pixel are interchangeable, i.e., all pixels can be considered replicas of a single pixel sampling a single temporal signal.

In conclusion, the exposure time introduces a trade-off between the receiver's sensitivity and the attainable bandwidth. Therefore, efforts should be invested to increase the performance of these systems, for example, by applying AI-assisted equalization stages in alleviating the effects of prolonged exposures, which is one of the objectives of this thesis (Chapter 1), and which is addressed in Chapter 5.

3.2 Artificial Intelligence for optical camera communication

This section introduces the generic methodology for solving tasks using AI. This methodology involves several steps, including (i) data preparation, (ii) model selection, (iii) training, validation, and testing, (iv) hyperparameter tuning, and (v) deployment. Moreover, it details the models that are used in this work, including MLPs, CNNs, and autoencoders (AEs).

It should be highlighted that this section is not focused on providing a detailed analysis of AI methods or explaining the mathematical background behind them. For a more comprehensive understanding of these topics, readers can refer to the book by Goodfellow et al. in 2016 [108] or the practical online book by Zhang et al. and updated in 2022 [109]. This chapter is intended to provide only an introduction to the fundamental concepts that will be used and revisited throughout the rest of the chapters.

3.2.1 Solving tasks using AI

The generic methodology for solving a specific task using AI involves several steps. This section briefly details those steps and includes a few considerations to achieve better results.

Identification of the problem

The first step is to define the problem that the network must solve clearly. This involves identifying the task the network will be trained to perform and any constraints or requirements the network must satisfy (e.g., computational complexity, time or space limitations).

Data collection

The next step is to collect a suitable dataset for training the network. This typically involves identifying a relevant source of data (e.g., a database or a collection of text or images) and pre-processing the data to extract the features used as inputs to the network. The quality and diversity of the data will significantly affect the performance of the trained network, so it is important to ensure that the dataset is representative of the problem at hand and contains enough examples to enable the network to learn effectively.

Exploratory data analysis

Before training a NN, it is important first to understand the characteristics of the data that will be used to train the network. This involves exploring the dataset and performing visual and statistical analyses to identify patterns, trends, or anomalies. This can help inform the design of the network and the choice of architectures to use.

One potential use of EDA is to assist in generating synthetic data, which involves creating artificially generated samples. Synthetic samples can be helpful in cases where it is difficult or time-consuming to acquire real-world data or where there are privacy or security concerns. However, to use synthetic samples for training (which might speed up the acquisition and training times), it is necessary to have sufficiently good models to generate high-quality representative samples of real-world data.

Dataset preparation

The next step is to prepare a dataset for training the network. This typically involves dividing the available data into at least two sets: a training set, which will be used to train the network, and a test set, which will be used to evaluate the performance of the trained network. The training set should be sufficiently large and diverse to enable the network to learn effectively, while the test set should be representative of the data the network will be applied to in practice.

In fact, data is split into three sets: training, validation, and test. Testing and validation are two closely related concepts in the context of training a machine learning (ML) model. Both involve using a dataset to evaluate the model's performance, but they have different goals and are typically used at different stages of the model development process. Testing refers to assessing a trained model's performance on a dataset it has not seen before to estimate its generalization performance. This is done to ensure that the model can make accurate predictions on new, unseen data, not just the training data. Testing is typically done after the model has been trained. On the other hand, validation refers to using a dataset to tune the hyperparameters of a model. The model's hyperparameters are parameters that are not learned during the training process but are set before training and control the model's behavior. Examples of hyperparameters include the learning rate, the number of layers in the model, and the regularization coefficient. This hyperparameter tuning is typically done during training to optimize the model's performance on a specific task or dataset. Validation helps ensure that the model can generalize well to new, unseen data.

Finally, in addition to dividing the data into training and validation/test sets, this step may involve other pre-processing steps, such as normalizing the data, removing outliers, or generating synthetic data to augment the training set.

Model selection

Once the dataset has been generated, the next step is to select the suitable model for the task at hand. This involves choosing the AI algorithm to be applied, and specifically, if a NN model is selected, the network architecture (e.g., the number of layers and the type of each layer) and the learning algorithm (e.g., stochastic gradient descent (SGD) or adaptative moment estimation (Adam)) that will be used to train the network. These learning optimization algorithms are algorithms used to train ML models, typically by minimizing a loss function that measures the difference between the predicted outputs and the ground truth. Learning optimization algorithms aims to find the appropriate parameters (e.g., weights and biases) that minimize the loss function and produce the most accurate predictions.

One common learning optimization algorithm is minibatch SGD, a variant of the SGD. In SGD, the model parameters are updated using the gradient of the loss function with respect to the current set of parameters, calculated using a single example from the training dataset. In minibatch SGD, the gradient of the loss function is computed using a small batch (or "minibatch") of examples from the training dataset rather than a single example. This allows the algorithm to make more efficient use of the available data and typically leads to faster convergence and better performance than SGD. The key advantage of minibatch SGD over SGD is that it improves the training process's stability and leads to more accurate parameters. Additionally, minibatch SGD allows the use of parallelization techniques, which can further improve its computational efficiency. However, one potential disadvantage of minibatch SGD is that it may require more careful tuning of the hyperparameters, e.g., the batch size, to achieve optimal performance.

Other optimization algorithms include: (i) Adagrad, which adapts the learning rate of the model during training based on the sum of the squares of the past gradients, (ii) root mean square propagation (RMSProp), which uses an exponentially decaying average of the squares of the past gradients to scale the learning rate for each parameter of the model, and (iii) Adam, which is an extension of the RMSProp algorithm, among many others.

Due to the complexity and diversity of models, architectures, optimization algorithms, and hyperparameters, it may be necessary to try out multiple different models and compare their performance on the validation or test set to select the best one.

Training

Once the model has been selected, the next step is to train the network on the training dataset. This involves feeding the data through the network and using the learning algorithm to adjust the network's parameters. The training process typically involves multiple epochs over the training dataset. An epoch refers to one full pass through the entire training dataset. It consists of one or more iterations over the training dataset. Each iteration involves the forward propagation of the model, the computation of the loss and gradients, and the backward propagation of the gradients to update the model's parameters. Finally, the network's performance is evaluated on the validation set at regular intervals to ensure that it is learning effectively.

Testing

After the network has been trained, the next step is to evaluate its performance on the test set. The network's performance is typically measured with metrics such as accuracy, precision, and recall, among many others. The evaluation results can be used to identify areas where the network is not performing well and may need further improvement.

Hyperparameter tuning

Another step in training a model is fine-tuning its hyperparameters. This involves adjusting the values of the hyperparameters (e.g., the learning rate or the regularization strength) to improve the network's performance on the test set. This step can be time-consuming and may require experimentation and trial and error, but it is important for achieving the best possible performance. Several algorithms can be used for hyperparameter tuning, including grid search, random search, and Bayesian optimization. Grid search involves exhaustively searching over a specified grid of hyperparameter values, while random search involves random sampling values for the hyperparameters from a specified distribution. Bayesian optimization uses Bayesian statistics to model the function that maps from hyperparameters to the model's performance and then uses an optimization algorithm to find the set of hyperparameters that maximizes this performance.

Deployment

The final step is deploying the trained AI model in a production environment, making it available to end users. This typically involves integrating the model into an application or system and configuring it to work with the desired input and output formats. It may also involve deploying the model on a specific hardware platform, such as a server or a mobile device. During the deployment process, it is important to consider scalability, reliability, and security issues. The deployed model should be able to handle a large volume of requests and should be able to function consistently and accurately over time. Additionally, measures should be taken to ensure the security and privacy of the data being processed by the model.

It should be remarked that deployment is a crucial step in the AI methodology. However, this research does not focus on deployment.

3.3 Artificial intelligence models

This section introduces and details the models used in this thesis. These models include MLPs, CNNs, and AE, which are commonly used in a wide range of applications. Each section examines the key features of a model, introduces some examples and potential use cases, and concludes by linking it with the related tasks to be solved in this thesis.

3.3.1 Multi-layer perceptron

A MLP is a type of NN that consists of multiple layers of interconnected "neurons". A neuron is a basic unit of computation that receives input data in the form of a matrix, processes it using a non-linear mathematical function, and produces an output. Neurons are typically organized into layers. The input data is passed through the input layer, which is then processed by one or more hidden layers and then transformed into the output data by the output layer. The number of layers and the number of neurons in each layer can vary, but generally, the more layers and neurons an MLP has, the more complex patterns it can learn to recognize.

The computations performed by each MLP layer can be represented by Eq.3.4:

$$\mathbf{H} = \Phi(\mathbf{X}\mathbf{W} + \mathbf{b}) \tag{3.4}$$

where $\mathbf{X}$, $\mathbf{H}$ are the input and output matrices, repectively, $\mathbf{W}$, the weights for the connections between neurons, $\mathbf{b}$, the biases vector, and Φ, the activation function.
The activation function serves a crucial purpose which is introducing non-linearities into the network, allowing to model complex relationships between the input and output data. Without these non-linearities, the MLP would be limited to learn linear relationships between inputs and outputs, which would significantly limit its ability to perform many tasks. In conclusion, using activation functions allows NNs to learn and model more complex and non-linear relationships between the input and output data. Moreover, choosing the appropriate activation function can have a significant impact on the performance of the network. Common activation functions include the sigmoid, the hyperbolic tangent, and the rectified linear unit (ReLU) functions. The choice of activation function depends on the specific task and other factors, such as the network architecture and the used training algorithm.

The sigmoid function is defined in Eq.3.5

$$\Phi(x) = \frac{1}{1 + e^{-x}} \tag{3.5}$$

This function maps the input values to the (0, 1) range, which makes it useful for modeling binary or probabilistic outputs, such as in classification tasks where the network must output the probability that a given input belongs to a particular class. However, the sigmoid function has the disadvantage of being computationally expensive to evaluate and differentiate, which can slow down the training. It also can saturate for large positive or negative values of x, leading to the vanishing gradient problem. This means that as the input becomes very large in magnitude, the output of the sigmoid function approaches either 0 or 1, and the gradient becomes very small. In simple terms, the gradient of the activation function is used to compute the weight updates during the training process, so if the gradient is very small, then the weight updates will also be small, which can slow down the learning process and make it difficult for the network to converge.

The tanh function is defined in Eq.3.6:

$$\Phi(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.6}$$

The tanh function maps the input values to the (-1, 1) range, which makes it well-suited for modeling outputs that can have both positive and negative values and where the range of possible output values is symmetric around 0. For example, the tanh function might be used in networks trained to predict a company's stock price based on various input factors. In this case, the output could be a continuous value representing the predicted stock price, which could be positive or negative depending on the input data. The tanh function presents the same issues as the sigmoid function.

The ReLU function is defined in Eq.3.7:

$$\Phi(x) = \max(0, x) \tag{3.7}$$

The ReLU function outputs 0 for negative values of x and x for positive values. This activation function is useful for modeling either 0 or positive outputs. One advantage of the ReLU function is that it is computationally efficient, as it does not require any complex mathematical operation such as exponentiation, which can be computationally demanding. This can make it useful when the NN needs to be trained quickly or deployed on hardware with limited computational resources. Another advantage of the ReLU function is that it has been shown to improve the convergence of networks during the training process. However, it also has some limitations, such as the potential rise of "dying" neurons, i.e., neurons that have become inactive and no longer contribute to the computation of the outputs. This might happens if the parameters of the network are not correctly initialized.

Some specific tasks that MLP models can be used for include:

- Regression: to predict a continuous output value based on a set of input features. For example, to predict the price of a stock based on historical data such as the stock's previous prices and trading volumes.
- Classification: to predict a discrete class label based on a set of input features. For example, to classify an image as belonging to one of several different object categories, such as "dog" or "cat".
- Clustering: to group data points into different clusters based on their similarity. For example, to group customers into different segments based on their purchasing behavior.
- Dimensionality reduction: to reduce the number of dimensions (encoding) in the data while preserving as much of the original information as possible to make it easier to visualize or speed up the training of other ML algorithms.

Other tasks that can be performed include image classification and recognition. However, there are several limitations to using MLP for working with images. One of the main limitations is that MLP does not inherently capture the spatial structure of images. The neurons in an MLP model do not have any notion of spatial location or orientation of the input values. As a result, this model is not translation invariant. Translation invariance (or translation equivariance) refers to the property of a model where its output is the same regardless of the position of the input data in the input space. For example, a translation invariant model should recognize an object in an image regardless of where the object is located in the image. Furthermore, an MLP cannot prioritize the input data values based on their location and orientation and instead treats all input values equally. In other words, it cannot generate local representations that consider only nearby pixels without directly relating them to other content in the image. This property is known as the locality principle. This principle refers to the property of a model where the output is only influenced by a small, local region of the input data. For example, a model that follows the locality principle should be able to recognize an object in an image based on the features of that object rather than being influenced by the features of other objects in the image. In addition, MLP is not well-suited for working with high-dimensional data, as it requires considerable computational resources, which can be a significant disadvantage when working with large and complex images.

To address this issue, other types of NN architectures have been proposed that are specifically designed to capture the spatial structure of the data, such as CNNs. These architectures use convolutional layers and other techniques to explicitly incorporate the concepts of translation invariance and the locality principle, making them more effective for tasks involving images.

3.3.2 Convolutional neural networs

A CNN is a type of NN specifically designed to work with multi-dimensional arrays such as images. CNNs are composed of several layers, including (i) convolutional layers and (ii) pooling layers.

Convolutional layers are the core building blocks of CNNs. They perform the discrete convolution operation on the input data for each filter in a filter set to extract specific features. The outputs of the filters are then stacked together to produce a feature map.

Considering an input matrix, $\mathbf{I}$, with dimensions, $H \times W \times D$, where H is the height, W the width, and D the number of channels, and a three-dimensional filter's kernel, K, with dimensions, $k_h \times k_w \times k_d$, the output, $\mathbf{Z}$, after applying the discrete convolution is given by Eq. 3.8.

$$\mathbf{Z}[i,j] = \sum_{m=0}^{k_h-1} \sum_{n=0}^{k_w-1} \sum_{l=0}^{k_d-1} \mathbf{I}[m,n,l] \cdot \mathbf{K}(i-m, j-n, l) \tag{3.8}$$

In this formula, $\mathbf{Z}(i,j,k)$ is the value of the output at position (i,j), $\mathbf{I}(m,n,l)$, the value of the input at position (m,n) for the l-th channel, and $\mathbf{K}(i-m, j-n, l)$, the value of the filter at position $(i-m, j-n)$ for the l-th input channel. This formula shows that the convolution is performed by sliding the kernel over the input data and computing the dot product between the input and kernel values at each position. The result of the convolution, $\mathbf{Z}[i,j]$, is then biased ($\mathbf{b}[i,j]$) and transformed using a non-linear activation function, ϕ, generating the corresponding feature map, $\mathbf{F}[i,j]$ (Eq. 3.9).

$$\mathbf{F}[i,j] = \Phi(\mathbf{Z}[i,j] + \mathbf{b}[i,j]) \tag{3.9}$$

Finally, the output of the convolutional layer consists of a set of feature maps resulting after applying each filter separately.

On the other hand, pooling layers are used to reduce the dimensionality of the input data by sub-sampling it, which can help to make the network more efficient and reduce overfitting. There are several types of pooling strategies, but one of the most common is max pooling, which takes the maximum value from each pooling window. For example, given an input feature map F with dimensions $H \times W \times D$, maximum pooling is defined by Eq. 3.10.

$$\text{maxpool}(F) = \max_{i=1}^{p} \max_{j=1}^{q} F_{i,j} \tag{3.10}$$

where p and q are the vertical and horizontal sizes of the local pooling neighborhood, respectively. Another strategy is average pooling, defined by Eq. 3.11.

$$\text{avgpool}(F) = \frac{1}{p \cdot q} \sum_{i=1}^{p} \sum_{j=1}^{q} F_{i,j} \tag{3.11}$$

Pooling layers help reduce the network's computational complexity and improve its overall performance by making the learned features more robust and invariant to small translations in the input data.

Regarding network architectures, LeNet [110] is the classic example of a CNN. It was one of the first successful implementations of a CNN, and it is often used as a starting point for teaching their basics. LeNet mainly consists of two blocks: (i) a block for extracting features, which is composed of several convolutional and pooling layers, followed by (ii) a fully connected network (i.e., a MLP) for making predictions.

However, architectures have evolved over time, achieving excellent performance in classification, segmentation, regression, and other tasks discussed below. The following are just some examples that have marked significant milestones in the advancement of CNNs according to Zhang et al. [109].

AlexNet, developed by Krizhevsky et al. in 2012 [111], was the first large-scale network to outperform conventional computer vision methods on a large-scale vision challenge. It achieved state-of-the-art performance on the ImageNet dataset and paved the way for developing deeper and more complex CNN architectures.

The VGG network, developed by Simonyan et al. in 2014 [112], uses several repeating blocks of elements. Each block consists of several convolutional and pooling layers, allowing the network to learn hierarchical features at different scales.

GoogLeNet, developed by Szegedy et al. in 2015 [113], uses inception blocks that perform multi-branch convolutions. This allows the network to learn multiple features simultaneously, which can improve its performance. GoogLeNet achieved state-of-the-art performance on the ImageNet dataset using this approach.

The residual network (ResNet), which was developed by He et al. in 2016 [114], is a CNN architecture that uses skip connections to allow the network to learn residual functions. In a traditional CNN architecture, each layer receives input from the previous layer and transforms it using a series of convolutional, non-linear, and pooling operations. In a ResNet, each layer receives input not only from the previous layer but also from skip connections that bypass one or more layers. This allows the network to learn residual functions, which are the difference between a layer's input and output. To implement this, ResNet uses residual blocks. One of the key advantages of ResNet is that it can learn complex functions using fewer layers than a traditional CNN. This is because the residual blocks allow the network to learn the residual functions, which are typically easier to learn than the underlying functions. Additionally, the skip connections help reduce the vanishing gradient problem, which can improve the network's performance. This network remains one of the most popular off-the-shelf architectures in computer vision.

ResNeXt blocks, developed by Xie et al. in 2017 [115], are a type of building block based on the idea of splitting the features learned by a layer into multiple branches and processing each branch separately. This allows the network to learn more complex features and reduces the number of parameters in the network.

DenseNet, developed by Huang et al. in 2017 [116], uses dense connections between layers. In a dense connection, each layer receives input from all previous layers rather than only from a few adjacent layers as in traditional CNN architectures. This allows the network to use features learned by other layers, which can improve its overall performance. In addition, to further improve the efficiency of the network, DenseNet uses transition layers, which reduce the number of feature maps and the size of the feature maps at each layer. This reduces the computational complexity of the network and allows it to learn more complex features without significantly increasing the number of parameters.

Over time, many special optimizations for efficient CNNs have been developed, such as coordinate shifts (ShiftNet) by Wu et al. in 2018 [117] and the automatic search for efficient architectures (MobileNet v3) by Howard et al. in 2019 [118].

In addition, semi-automatic design exploration methods, such as the work of Radosavovic et al. in 2020 [119], have led to the development of architectures such as RegNetX/Y, which have demonstrated state-of-the-art performance on various tasks. These methods allow for the efficient exploration of the design space of CNN architectures, which can help to identify optimal architectures for specific tasks.

CNNs are widely used for several tasks, including:

- Image classification. It involves assigning an image to one of several predefined classes (or labels), such as ”dog” or ”cat”. This task is typically performed by training a CNN on a large dataset of labeled images, where each image is associated with a specific class.

- Object recognition. It goes one step further by not only assigning a class label to an image but also identifying the location of each object in the image. For example, a CNN trained for object recognition can identify an image containing a dog and where the dog is. This is often achieved using object detection, which involves sliding a small window across the image and using the CNN to predict whether each window contains an object of interest.

- Object segmentation. It is a related task that involves not only identifying the presence and location of objects in an image but also assigning a unique label to each pixel in the image. This allows the CNN to create a detailed map of the objects in the image, segmenting them from the background and each other. This can be useful for tasks such as medical image analysis, where it is important to identify and differentiate between different structures in an image accurately.

- Image generation. It involves training a CNN on a large dataset of images to learn the underlying patterns and structures present in the data to generate new images later. Typically, image generation using CNNs involves generative models. There are several different generative models, but one of the most commonly used for image generation is the generative adversarial network (GAN). A GAN consists of two main components: a generator and a discriminator. The generator generates new images using noise as input, while the discriminator is trained to classify images as real or fake. The generator and discriminator are trained simultaneously, with the generator trying to produce images that are indistinguishable from real ones and the discriminator trying to classify the generated images accurately.

Moreover, CNNs can also be used for a wide range of other tasks, including:

- Natural language processing: to perform tasks such as sentiment analysis, text classification, and language translation.

- Music generation: to learn the underlying patterns and structures present in music, and then used to generate new, unique music.

- Speech recognition: to process audio data and recognize spoken words and phrases.

- Anomaly detection: to identify unusual or anomalous patterns in data, such as fraud in financial transactions or malfunctioning machinery in a manufacturing plant.

In conclusion, CNNs are a versatile tool that can be used for many tasks beyond image recognition and generation.

In this work, the proposed CNN must perform a regression task accurately to retrieve estimates of the link parameters directly from the images captured with a camera (Chapter 5).

3.3.3 Autoencoders y Convolutional Autoencoders

An AE is a type of NN that is used to learn a latent (compressed) representation of a dataset. This is done by training the network to map the input data, x, into a lower-dimensional representation, z, and then to reconstruct the original input data from this representation. Hence, AEs are composed of two main parts: the encoder, which maps the input data to the encoding, and the decoder, which maps the encoding back to the reconstructed input. The encoder and decoder are typically implemented as NNs with a series of interconnected layers.

This underlying process is mathematically described in Eq. 3.12.

$$\bar{\mathbf{x}} = \mathcal{D}\left(\mathcal{E}\left(\mathbf{x}\right)\right) \tag{3.12}$$

Where $\mathbf{x}$ is the input matrix, $\tilde{\mathbf{x}}$ is the reconstructed version of $\mathbf{x}$, $\mathcal{E}(\cdot)$ is the encoding operation, and finally $\mathcal{D}(\cdot)$ is the decoding procedure.

In general, AEs are unsupervised learning algorithms since they do not require labeled training data. Instead, the network learns to encode and decode the input data by minimizing a loss function that measures the difference between the input and the reconstructed output as defined in Eq. 3.13. This allows the AE to learn useful representations of the data that can then be used for various downstream tasks, such as classification or clustering.

$$L\left(\mathbf{x}, \mathcal{D}\left(\mathcal{E}\left(\mathbf{x}\right)\right)\right) \tag{3.13}$$

As an example, a possible loss training function used extensively in this thesis, $L(\mathbf{x}, \tilde{\mathbf{x}})$ corresponds to the L_2-norm (i.e., mean squared error (MSE)), which includes a regularization term that has been added to prevent over-fitting, is defined in Eq. 3.14.

$$L(\mathbf{x}, \bar{\mathbf{x}}) = E\left[\left(\mathbf{x} - \bar{\mathbf{x}}\right)^2\right] + \lambda R\left(\mathcal{E}, \mathcal{D}\right) \tag{3.14}$$

$E\left[\cdot\right]$ denotes expected value, λ is the regularization coefficient, and $R(\cdot)$ is the regularization function, which in this thesis combines $L1$ and $L2$ regularization penalties.

AEs have many applications, including data compression, denoising, and feature learning. By learning a compact representation of the input data, AEs can help to reduce the storage and computational requirements of working with large datasets and can be used as a pre-processing step for other ML algorithms. Additionally, by learning to remove noise from the input data, AEs can be used to improve the quality of data that has been corrupted or degraded.

The latter task is performed by denoising AEs. Those networks are trained to minimize the difference between the input and its reconstructed version whilst previously corrupting the input with different noise sources. Equation 3.15 exemplifies this case.

$$L\left(\mathbf{x}, \mathcal{D}\left(\mathcal{E}\left(\tilde{\mathbf{x}}\right)\right)\right) \tag{3.15}$$

where $\tilde{\mathbf{x}}$ is a corrupted version of $\mathbf{x}$.

On the other hand, when working with images, CAEs are preferred for the same reasons given in the discussion between MLPs and CNNs. These networks use convolutional layers for the encoding and introduce a new layer, the transpose convolutional layer, for the decoding part.

Transpose convolutional layers perform the inverse operation of a convolutional layer. However, although it is also (wrongfully) known as a deconvolutional layer, it does not perform the deconvolution operation. Instead, it performs a regular convolution on an upsampled input version. This is the opposite of the downsampling performed by a convolutional layer, allowing the network to reconstruct the original input data from the encoding.

In general, a CAE operates in the same way as a traditional AE, but using convolutional and transpose convolutional layers to learn efficient encodings of the input data, improving the performance on images and other data that has a spatial structure.

CAEs can be used for a variety of applications, including:

- Image denoising: to remove noise from images (Xiao et al. in 2016 [120]). This is especially useful for cleaning up images corrupted by low-quality cameras or taken under low-light conditions.

- Image super-resolution: to increase the resolution of images, effectively making them appear more detailed and sharper (Dong et al. in 2015 [121]. This can be useful for improving the quality of images for applications such as image recognition and object detection.

- Image generation: to generate new images that are similar to a given dataset of images (Radford et al. in 2015 [122]. This can be useful for creating artificial images for use in training ML models or for generating new images based on user-specified parameters.

- Image compression: to reduce the dimensionality of the inputs, effectively reducing the size of the images (Cheng et al. in 2018 [123]).

- Object detection: to perform a pre-processing step for object detection tasks (Williams et al. in 2017 [124]. By denoising and enhancing the resolution of input images, a CAE can improve the performance of a subsequent object detection model.

- Anomaly detection: to detect anomalies, i.e., to identify unusual or abnormal patterns in a dataset. By training a CAE on a dataset of normal instances, it can learn to identify instances that deviate from the norm, potentially indicating anomalies (Siddalingappa et al. in 2021 [125].

- Semantic segmentation: to perform a pre-processing step for semantic segmentation tasks, which involve partitioning an image into multiple segments or regions and labeling each segment with a semantic class (Badza et al. in 2021 [126]. By denoising and enhancing the resolution of input images, a CAE can improve the performance of a subsequent semantic segmentation model.

In conclusion, CAEs perform well as denoising tools and for image generation. The results obtained in the works mentioned above suggest considering the use of CAEs for the equalization of images affected by different exposure times, which is one of the objectives of this thesis (Chapter 5).

Chapter 4

Methodology

The general objective GO1 of this thesis involves the design of an RS-based OCC link in which the camera does not lose its functionality as an imaging device. To do this, as seen in Chapter 3, the camera's exposure time must be adapted to the ambient light, which might lead to the appearance of ISI depending on the transmitter baud rate, reducing the receiver bandwidth. As a solution, this thesis proposes using AI to reconstruct the original signal as if it were acquired with a receiver with a significantly broader bandwidth. Furthermore, AI is also used to estimate link parameters required by the reception algorithms directly from images.

To achieve this goal, it is necessary to rigorously generate/compile datasets for (i) the analysis, modeling, and validation of the sequential acquisition of the RS cameras and (ii) the training and evaluation of the proposed neural network models (General objective GO2).

Section 4.1 details the methods and experiments carried out for the procedural generation of synthetic samples and capturing real-world images. Section 4.2 describes the methods used in sample preprocessing, network training and validation, and network hyperparameter optimization. Moreover, it introduces the proposed models and justifies their use.

4.1 Data collection

The samples used for training the NN models depend on the addressed task:

Task A equalization of the exposure effects. In this case, the training samples consist of a set of two images corresponding to the same transmitted signal: (i) one affected by a particular exposure time (the network input), and (ii) another one corresponding to the ground truth, i.e., the signal expected to be reconstructed by the network (network output). The latter corresponds to a version of the original signal in which the camera's exposure time is set to the minimum attainable time.

Task B estimation of link parameters. In this case, the input is the same as the previous case, and the output is replaced by a 1D tensor with the expected link parameters.

Regarding the data collection methods for the training images, it should be highlighted that obtaining real-world samples is exceptionally challenging (especially for Task A). To demonstrate this challenge, let's consider the simplest generation pipeline. It consists of the following steps: (i) configuring the camera with the expected exposure time, (ii) taking the first image, which will be related to the network's input, (iii) configuring the camera again with the lowest exposure time, and (iv) taking another image, related to the network's output. There are two issues with using this pipeline. On the one hand, the impossibility of accurately controlling the camera's internal trigger to start the acquisition may cause a lack of synchronization between the transmitter and the receiver. For this reason, as the input and output images are captured at different instants, the received signal samples are not perfectly aligned between images. In other words, the expected signal samples are not located in the same place within the image. They are shifted based on a variable offset at the start of the acquisition. On the other hand, in most cameras, the shortest available exposure time is not low enough to consider an ideal transmission. These challenges reveal that the generation of training samples based on real-world images is not the most suitable option in terms of efficiency or complexity.

Alternatively, synthetic generation of samples based on a precise model of the RS acquisition can be advantageous, as: (i) it enables the rigorous control of the sample parameters, adjusting them to different cameras and transmitters (if the acquisition model is accurate enough), (ii) it allows inserting effects on the signal in a controlled way, such as noise and image compression artifacts, and (iii) it is efficient in terms of time and resources, only requiring a parallelizable computing unit instead of an experimentation testbed.

Regarding the validation samples, they differ slightly from the training samples. Each sample consists of an image used as the network input and the sample descriptors required for computing the validation metric. For instance, Task A (i.e., the equalizer) requires the transmitted bits to compute the BER, and Task B (i.e., the regressor) requires the original NPPS and ESR to compute the MSE. Therefore, the validation samples can be extracted from a generic dataset containing all the required descriptors. The procedure to generate the validation samples requires a rigorous experimental setup that is fully automated to speed up the time invested in taking samples.

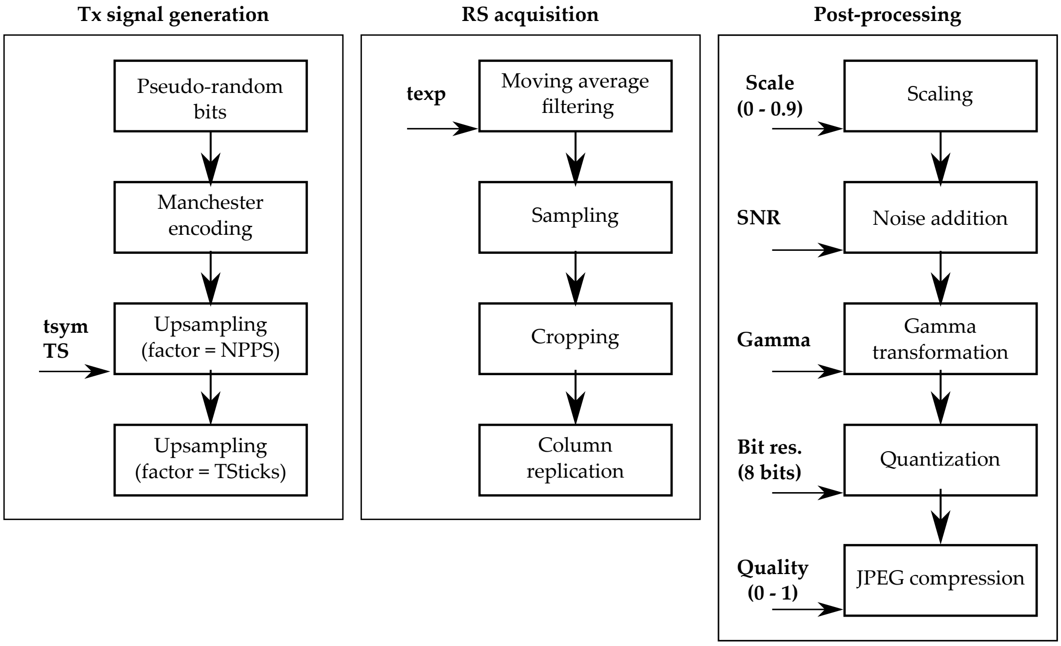


Figure 4.1: Diagram of the synthetic generation.

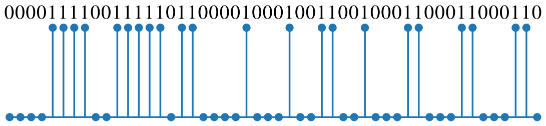
The following subsections present the methods for generating the synthetic training samples and datasets and the experimental setup for obtaining the real-world validation dataset.

4.1.1 Synthetic generation

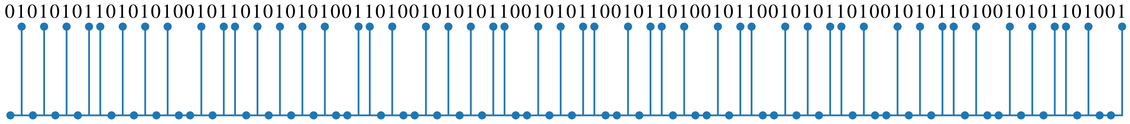
One of this thesis' main contributions is the procedural method for generating synthetic RS-based OCC images. This process was first proposed in [21] and improved in [22]. These two articles are included in the compendium (Section 5.2 and Section 5.3 respectively).

Figure 4.1 shows the processing steps of the latest version of this method. It uses three input variables, which are temporal parameters of the link: the transmitter's symbol time, t_{sym}, the camera's exposure time t_{exp}, and its sampling period, $T_{\text{S}} = t_{\text{row}}$. These parameters are expressed in terms of the number of clock ticks considering a reference clock frequency. For instance, using a reference clock with $1\mu s$ period, the number of ticks for a symbol time, $t_{\text{sym}}^{\text{ticks}}$, t_{sym}, of $300\mu s$ is 300. Ultimately, the reference clock refers to the simulation sampler with a higher sampling rate than the camera's sampling rate. The ratio between the simulation sampler and the camera's sampling rates must always be greater than 1 (typically 100 or 1000).

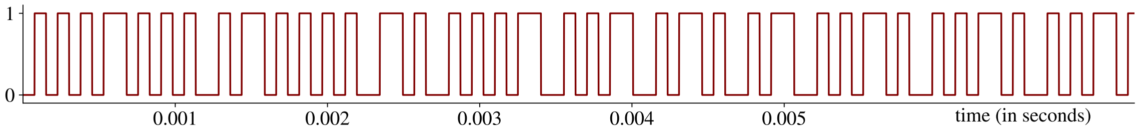
Other parameters used throughout the method are the scaling factor, the SNR, the camera's gamma, and the joint photographic experts group (JPEG) quality. The remainder of this section describes each step in the processing algorithm, which is further divided into (i) signal generation, (ii) RS acquisition, and (iii) post-processing.



(a) Pseudo-random bits generated.



(b) NRZ Manchester-encoded bits.



(c) TX signal samples.

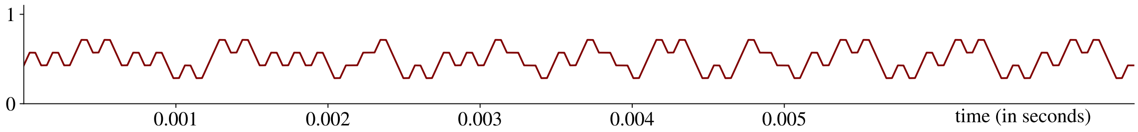
Figure 4.2: TX signal generation.

(i) TX signal generation

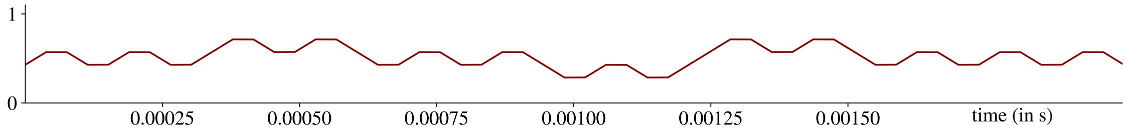
The process starts generating the transmitter (TX) samples. Figure 4.2 shows the output at each step in this stage. First, a 1D tensor containing pseudo-random bits is generated (Fig.4.2a). Next, those bits are encoded following the non-return-to-zero (NRZ) Manchester encoding convention, mapping the logic bits '0' and '1' to the bit chains '01' and '10' respectively (Fig.4.2b). Finally, the resulting vector is upsampled by two factors to generate the TX samples: (i) the NPPS, and (ii) the number of sampling ticks, $T_{\mathrm{S}}^{\mathrm{ticks}}$ (i.e., the ratio between the reference clock and the camera's sampling frequencies).

(ii) RS acquisition

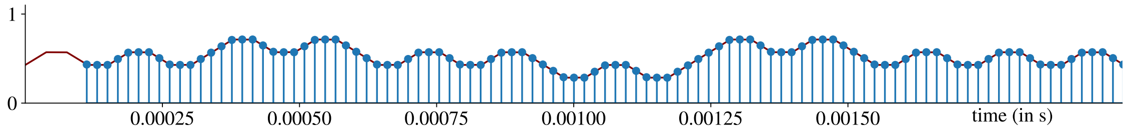
Next, the output tensor is filtered and sampled to obtain the receiver (RX) samples (corresponding to one IS's pixel column). Figure 4.3 shows the output at each step in this stage. The filtering step applies a moving average function over the TX samples. It convolves the signal with a uniform 1D-kernel with size, S, equals to the number of exposure ticks, $t_{\mathrm{exp}}^{\mathrm{ticks}}$. The kernel's coefficients are equal to $1/S$. It should be highlighted that to generate these samples it is assumed that the incident irradiance is uniform across the IS and that the behavior of the pixel voltage-charge conversion is linear (Chapter 3).



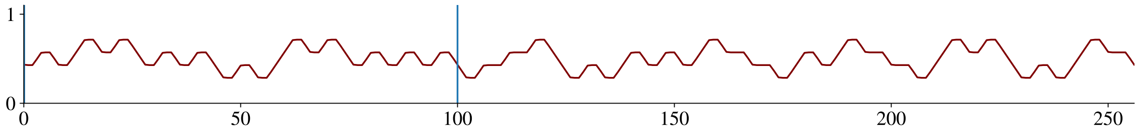
(a) Tx signal filtered.



(b) Tx signal filtered zoomed (20000 first samples).



(c) Sampling TX signal. The sampling offset corresponds to the time of the first sample.

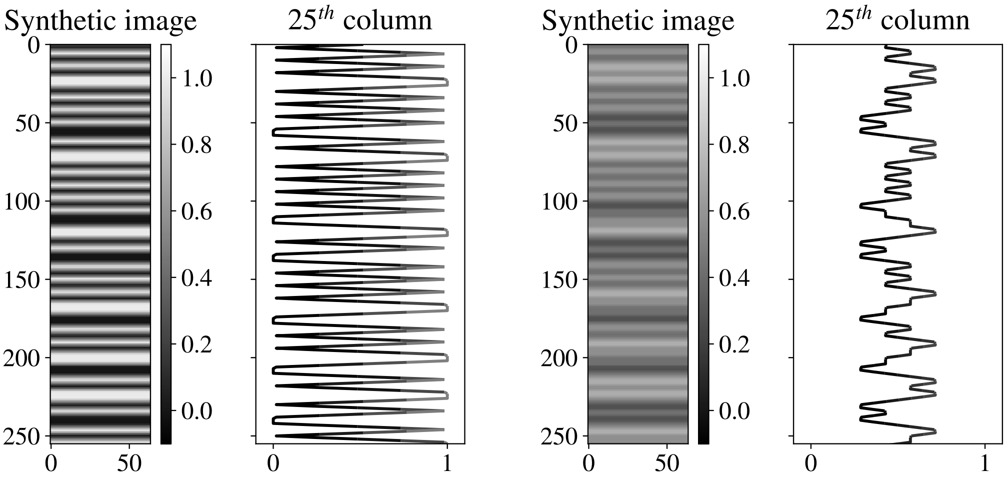


(d) Rx samples.

Figure 4.3: Rolling shutter filtering and sampling.

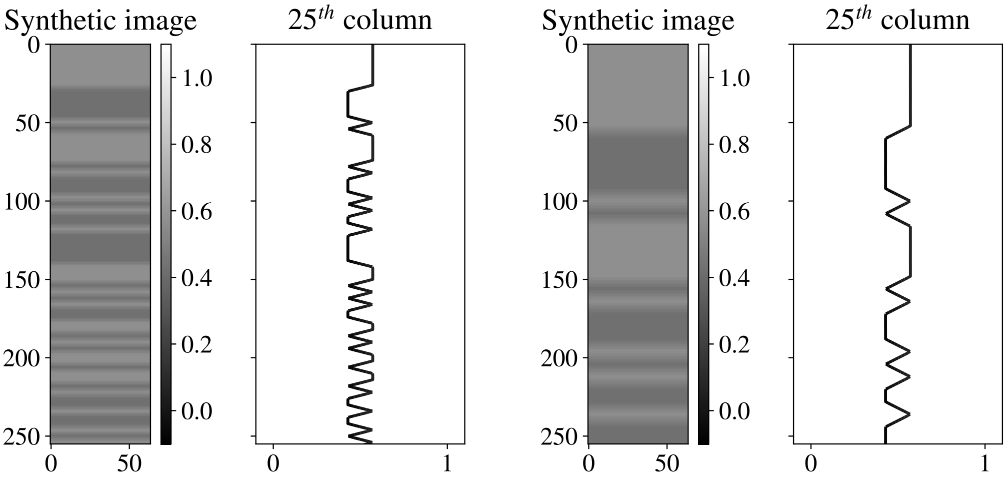
Figure 4.3a shows the filtered TX samples. Then, the signal is sampled at fixed instants depending on the sampling ticks, $T_{\mathrm{S}}^{\mathrm{ticks}}$. In addition, the sampling starts at a random offset to account for misalignments between the transmitter and the receiver. This offset ranges from 0 to $t_{\mathrm{sym}}^{\mathrm{ticks}}$. To illustrate this operation, Fig. 4.3b zooms a segment of the filtered TX samples (showing the 20000 first samples), Fig. 4.3c shows the sampling instants over the filtered TX signal, and Fig. 4.3d shows the final RX samples (corresponding to 256 samples).

Once the RX samples tensor is obtained, the image is generated by replicating those values in the column dimension, resulting in a 2D image. Figure 4.4 shows four examples of different symbol and exposure times.



(a) $t_{\text{sym}} = 75.614\mu s$ (NPPS=4)
$t_{\text{exp}} = 75.614\mu s$ (ESR=1)

(b) $t_{\text{sym}} = 75.614\mu s$ (NPPS=4)
$t_{\text{exp}} = 264.649\mu s$ (ESR=3.5)



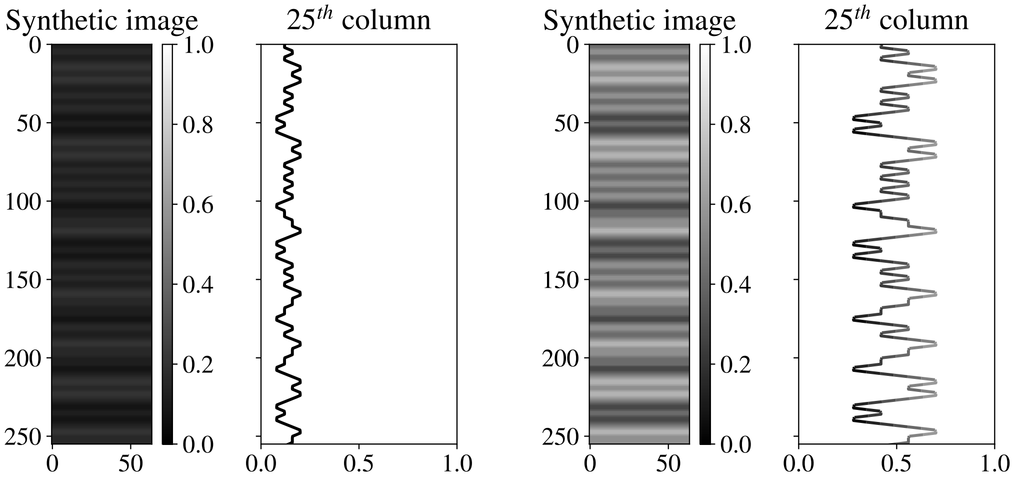
(c) $t_{\text{sym}} = 75.614\mu s$ (NPPS=4)
$t_{\text{exp}} = 529.298\mu s$ (ESR=7)

(d) $t_{\text{sym}} = 151.228\mu s$ (NPPS=8)
$t_{\text{exp}} = 1058.596\mu s$ (ESR=7)

Figure 4.4: Examples of different synthetic images.

(iii) Post-processing

Scaling. The second part of the method starts by dividing the 2D tensor by its maximum value. Then, the resulting tensor is multiplied by a scaling factor which takes values between 0 and 0.9. This scaling factor accounts for different received signal powers and is only relevant when considering the camera's gamma and image compression transformations. This is because these steps apply non-linear transformations to the input values. In any other case, the scaling step can be bypassed (and hence the received power neglected), as the standardization applied to the network inputs (in the preprocessing part) will output equivalent signals independently of the received power. In addition, recall that the scaling factor is lower than one. Otherwise, values will be saturated after adding noise to the signal.



(a) Scaling factor equals 0.2.

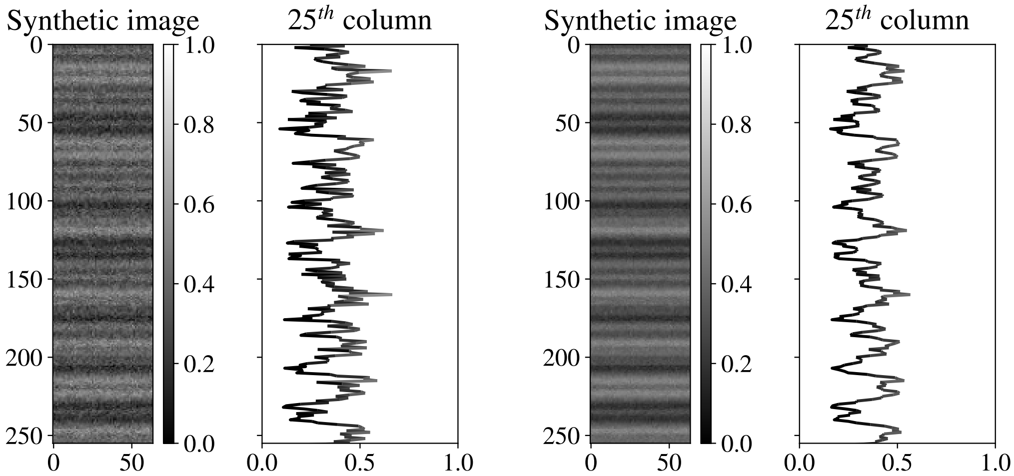
(b) Scaling factor equals 0.7.

Figure 4.5: Examples after the scaling step.

Noise addition. The following step adds AWGN to the signal based on the desired SNR. This adds robustness to the training and allows the CAE network model to improve its denoising capabilities. The standard deviation, σ_{noise} of the AWGN samples can be computed from the SNR using Eq. 4.1.

$$\sigma_{\text{noise}} = \sqrt{\frac{P_{\text{signal}}}{10^{\text{SNR}/10}}} = \frac{\sigma_{\text{signal}}}{10^{\text{SNR}/20}} \tag{4.1}$$

where P_{signal} is the power of the signal. Since the NRZ Manchester encoded samples correspond to an ergodic signal, its power can be computed using the variance of the samples, σ^2_{signal}. Figure 4.6 displays two output examples for SNR of 2 and 10dBs.

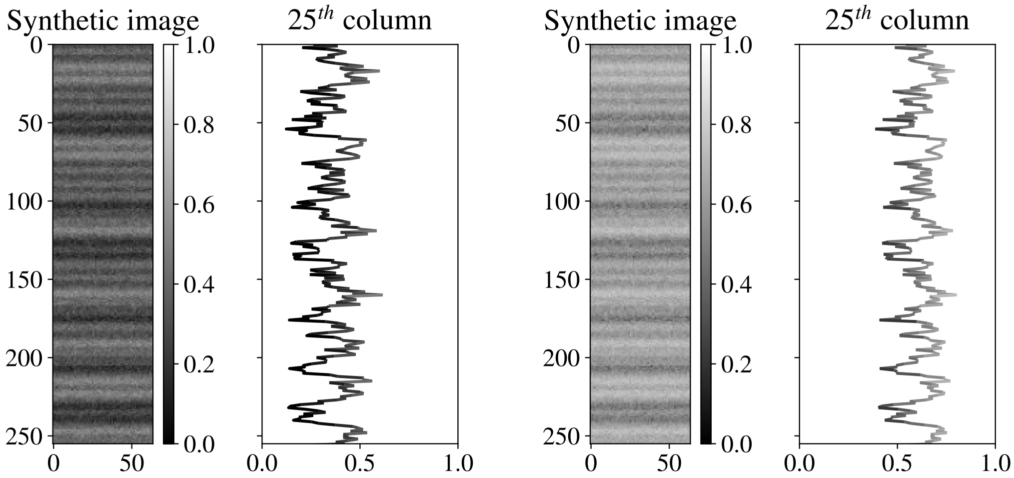


(a) SNR equals 2dBs.

(b) SNR equals 10dBs.

Figure 4.6: Examples with noise added. The scaling factor equals 0.5.

Gamma transformation. Next, the camera's gamma is applied. Figure 4.7 shows the outcome of using a gamma of 2.2. As can be seen, the low luminance values are boosted.

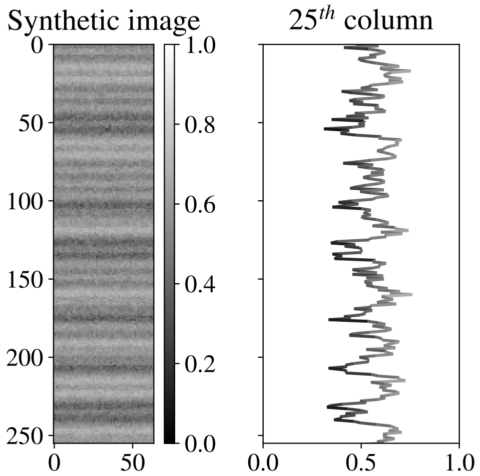


(a) Gamma equals 1.

(b) Gamma equals 2.2.

Figure 4.7: Examples after applying different gamma transformations. (Scaling factor=0.5, SNR=5dBs).

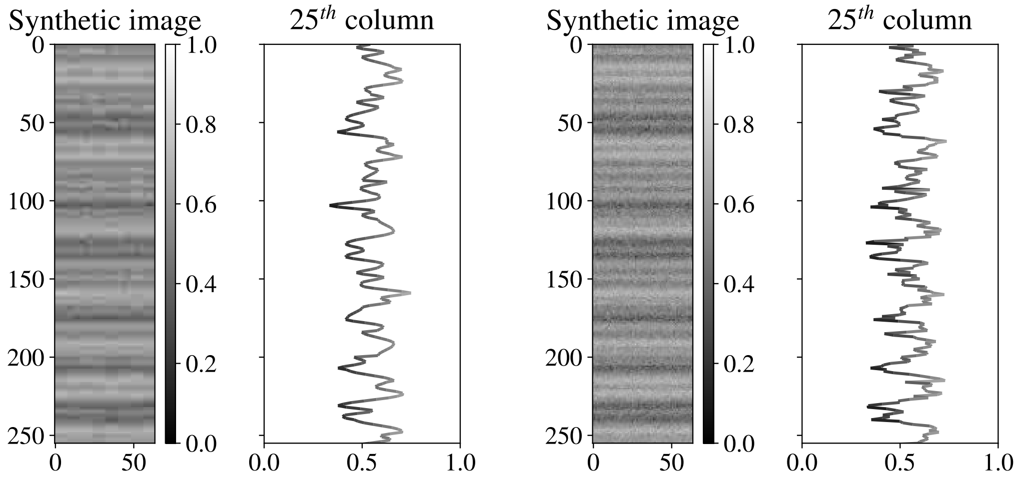
Quantization. The subsequent step quantifies the values depending on the pixel value resolution (8 bits). This is accomplished by multiplying the vector by $2^8 - 1$ (=255) and rounding its values to the nearest integer. Figure 4.8 shows the output result and the quantization errors obtained.



(a) Image quantized to 8 bits unsigned integer values.

Figure 4.8: Examples after the quantization step (Scaling factor=0.5, SNR=5dBs, Gamma=1.8).

JPEG compression. Finally, the image is compressed using a JPEG encoder with different configurable qualities and divided by 255 (for normalization). Figure 4.9 shows two examples for JPEG quality of 10 and 70. As it can be seen, the JPEG compression algorithm can greatly distort the received image. This distortion can negatively affect the SNR or alter the spectrum of the received signal. Also, since encoding is done on image patches, these distortions can unevenly affect different parts of the signal.



(a) JPEG quality equals 10.

(b) JPEG quality equals 70.

Figure 4.9: Examples with JPEG compression artifacts (Scaling factor=0.5, SNR=5dBs, Gamma=1.8).

Synthetic dataset

The samples used to train the network are not generated from scratch during training. This is extremely inefficient. Instead, they are drawn from a previously built dataset. However, this dataset does not contain the final inputs used in training. Actually, the dataset's samples are obtained after the (ii) RS acquisition step. The final post-processing steps will be applied on demand during training. The reason behind this is to speed up those parts of the synthetic generation that are computationally more demanding while leaving more freedom to vary on demand those parameters that do not have a high impact on the training time, such as the scale and the SNR. This way, the training samples can be adjusted to increase the network's performance. For example, random values for the scale and the SNR can be retrieved using different probability distributions.

4.1.2 Experimental setup

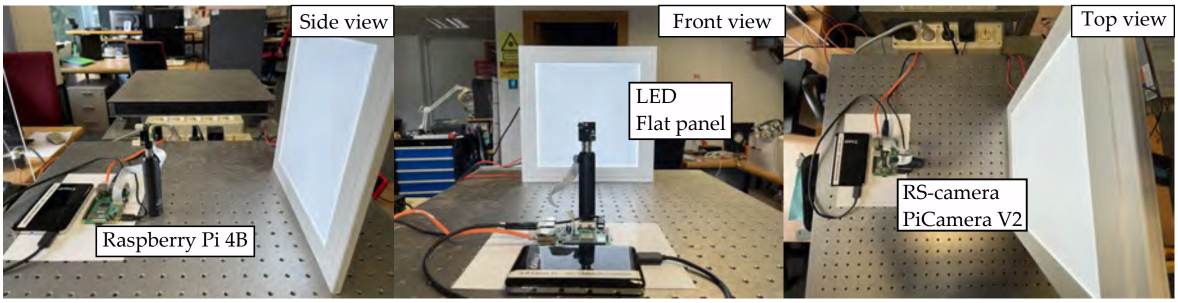


Figure 4.10: Experimental setup used for obtaining real-world image samples.

The validation dataset (Dataset 1 [27]) was generated using the experimental setup shown in Fig. 4.10. It consists of a flat-panel LED pointing towards an RS-camera separated at a distance of 20 cm. At this distance, the transmitter occupies the image entirely and illuminates the IS uniformly. The transmitted signal is generated using a arbitrary waveform generator (AWG) and a power supply. Table 4.1 summarizes the key parameters of the experiment necessary for its replicability.

Table 4.1: Experiment's key parameters.

Parameter	*Value*
Camera	
Camera model	Raspberry Pi Camera V2
IS	Sony IMX219 [127]
Aperture lens & relative focal length	f/2 | (3 mm)
Image resolution (Aspect ratio)	1920x1080 pixels (16:9)
Capture mode	(1) Partial AoV. No pixel binning.
Analog gain	1 to 12 (selected from 1 to 7)
Digital gain	1 to 12 (fixed to 1)
White balance (red and blue relative gains)	
Exposure times, t_{exp} (in μs)	20 to 3000 in steps of 19 (selected from 20 to 1500)
Sampling period, T_{S} (in μs)	18.9035
Transmitter	
Flat-panel model	Eglo connect Salobrena-C (Ref: 32546)
Electric power (in Watts)	16
Source voltage (VDC)	30 to 39
Source current (in mA)	420
Luminous flux (in Lumens)	2000
LED color ($T_{[\text{Kelvin}]}$)	Cold white (6500K)
Color rendering index (CRI), RA scale	$\geq$ 80
Surface dimensions (in mm)	230x230
Total dimensions (in mm)	300x300x11
AWG model	Agilent 33500
Power supply model	Yokogawa GS820
Symbol time, t_{sym} (in μs) — NPPS	75 to 135 — 4 to 8
Header, payload, stuffed, and trailer bits	5, 35, 12, 1
Random seed	31415

Table 4.2: Dataset's descriptors.

Descriptor	*Description*
sample_uuid	A universally unique identifier (UUID) for the image.
relative_path	The relative path where the images are stored.
dims	The image dimensions (1080,1920,3).
tclk	The reference clock period used as reference.
sampling_ticks	The number of clock ticks of the sampling period, T_S.
exposure_ticks	The number of clock ticks for the exposure time, t_{exp}.
txsymbol_ticks	The number of clock ticks for the symbol time, t_{sym}.
npps	The ratio between txsymbol_ticks and sampling_ticks.
esr	The ratio between exposure_ticks and txsymbol_ticks.
camera_model	The camera model used for reference.
analog_gain	The camera's analog gain.
digital_gain	The camera's digital gain.
awbg_gains	The vector containing the relative gains for the red and blue channels regarding the green channel. They are used to control the white image balance.
led_color	The identifier of the flat panel used.
power	The electrical power units of the signal.
modulation	The modulation's identifier. 'OOK' corresponds to an OOK modulated signal.
encoding	The encoding strategy's identifier. 'NRZ-Manchester' refers to a NRZ manchester encoding.
framing	The framing strategy's identifier. '11111-ABC C-0' refers to a packet with five one-bit headers, followed by bit A, bit B, and bit C, then the NOT version of bit C (stuffed bit), and a zero-bit trailer.
data_length	The payload's length (in bits).
data	The number of payload bits.

The process for obtaining the validation samples is fully automated. First, the signal is uploaded from the host computer to the AWG using the VXI-11 network protocol (over TCP/IP). Then the computer sends commands through secure shell (SSH) to the Raspberry Pi for the camera to take a photo. Next, the image is stored temporarily in local folders to be transferred through SSH to the host computer. Finally, the host computer creates a new entry for the sample with the corresponding sample descriptors. Capturing a complete image takes about 170 milliseconds on average, considering the camera warm-up time.

The generated validation dataset [27] contains more than 7000 image samples for different NPPS and ESR (Chapter 3). The NPPS values range from 4.0 to 7.0 in steps of 0.5 units, and the ESR from 1.0 to 7.0 in steps of 0.5 units. This dataset contains a CSV file with the sample's descriptors (columns) detailed in Table 4.2.

4.2 Data processing

Data processing can be divided into three phases: data preprocessing, training and validation of NN models, and final system assessment concerning communications performance. In addition, despite following a similar processing flow, the methods used for each task: Task A exposure time equalization and Task B link parameter estimation, may differ. This section details the general and specific methods for both cases.

4.2.1 Preprocessing

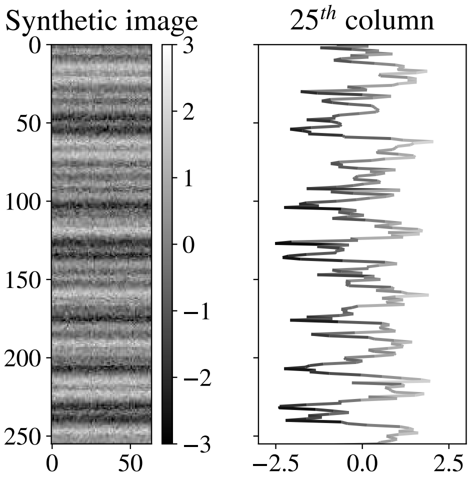


Figure 4.11: Examples after applying the z-score standardization (Scaling factor=0.5, SNR=5dBs, Gamma=1.8, JPEG Quality=70).

The input samples (images) delivered to the neural network, both in training (synthetic images), validation, and regular operation (real-world images), go through a preprocessing phase. The most critical processing step is the z-score standardization, which subtracts the images' mean and divides them by their standard deviation (Eq. 4.2).

$$z = \frac{x - \mu}{\sigma} \tag{4.2}$$

This step is essential since: (i) it makes both the synthetic and real-world images comparable, at least from the network's viewpoint and (ii) it pushes aside the need to consider the power of the received signal (and some camera parameters such as the analog and the digital gains) since two matching signals with different power levels will keep the same shape and acquire the same values (with an approximation error) after this standardization. For this reason, the synthetic generation algorithm does not consider a priori the expected signal power level at the reception and exclusively considers the temporal parameters of the link. To validate these statements, the comparison between the synthetic and real-world images is accomplished in the last two works of the compendium and detailed in Section 5.2 and Section 5.3. Figure 4.11 shows an example of the z-scoring normalization of a synthetic image sample.

4.2.2 Network training and validation

The steps followed in the supervised training of the neural models are as follows. First, a subset of samples of the synthetic dataset is extracted for training (the training dataset) and another for evaluating the model fit (the validation dataset). Second, in each training epoch, the post-processing steps, as seen in the synthetic generation section, are applied to each sample. In this part, the scaling factor, the signal noise, the gamma, and the JPEG quality are chosen randomly from their defined intervals following a uniform distribution. It is important to highlight that while the training and validation datasets do not change between epochs, the final samples delivered to the network will be different based on the random values chosen for those post-processing parameters. Third, the training samples are preprocessed (using the z-score standardization) and delivered in batches to the neural network until completing one epoch, i.e., when the network sees all the samples from the training dataset. Finally, the loss function is computed, and the errors are backward propagated, adjusting the weights and biases of the neural network. The loss function used in both tasks is the MSE. In Task A, the error is calculated between the network's delivered image and the expected reconstructed version. In Task B, it is computed between the network's estimated and expected parameters. After each epoch, the validation dataset is presented to the network to obtain an unbiased evaluation of the model fit. In addition, this work optimizes the network hyperparameters by using the hyperband searching algorithm [128]. This searching algorithm is generally used in place of other methods, such as grid search or random search, which involves testing a fixed set of hyperparameter configurations because it is more efficient and can find a better set of hyperparameters in many cases. One reason hyperband is more efficient is that it uses a simple and flexible structure that allows it to adapt to the particular characteristics of the optimization problem. It does this by using an iterative process to narrow down the search space, focusing on the most promising configurations and discarding the ones that are unlikely to be optimal. This contrasts with grid search, which tests all configurations in a pre-specified grid.

The network models considered in this thesis are a CAE for Task A, a CNN Task B, and a U-Net used in another related work [26] (Section 5.3.1). These network models are described in Chapter 3, and their network parameters are detailed in the corresponding Section 5.2 and Section 5.3 of Chapter 5.

4.2.3 Network assessment

The network assessment is carried out using the real-world dataset. Therefore, the metrics used in this evaluation differ from the error computed in the network training.

In Task A (the exposure equalization), the evaluation metric is the BER. The evaluation steps are detailed in Section 5.2 and can be summarized as follows. First, segments (with the network input's dimensions) are extracted from each real-world image and delivered (after preprocessing) to the neural network. Then, the reconstructed versions for each segment are concatenated to reassemble the entire signal, which is queued to be processed by the acquisition routines. The acquisition routines are extensively detailed in [23] and can be summarized as follows. First, a header template is created for the expected NPPS. Then, it is correlated with the entire image to retrieve the maximum Pearson correlation coefficient. A data packet is found if the Pearson coefficient exceeds a fixed threshold. Next, the signal is sampled at the estimated sampling instants to decode the final bits using correlation methods. Finally, the BER is computed by comparing the decoded and original bits.

In Task B (the link parameter estimation), the evaluation metrics are the root MSE (RMSE) and the relative error (RE) for each estimate (i.e., the NPPS and the ESR) regarding its target value. The evaluation steps are detailed in Section 5.3 and can be summarized as follows. First, random segments (with the network input's dimensions) are extracted from each real-world image in the dataset and delivered to the network. Finally, the evaluation metrics are computed with the network's estimates.

Chapter 5

Research outcomes

This chapter presents the research outcomes supporting this thesis' compendium. The original journal manuscripts are embedded in this chapter and sorted according to their publication date. In addition, each publication has a brief introduction section that describes the context of the work, the challenges addressed, the results obtained, and details its contributions and highlighted findings.

This chapter introduces a practical deployment of an OCC link in an artificial light-driven microalgae production plant in Section 5.1. In this work, a laboratory surveillance camera behaves simultaneously as an optical receiver of several smart-lighting photobioreactors transmitters. This work is considered the starting point of this thesis and constitutes its main thread. On the one hand, the requirement imposed on the surveillance camera to simultaneously watch the room without compromising data acquisition will lead to formulating the general objective GO1 and its technical objectives. Moreover, the experiment reveals (as discussed in Chapter 3) that the exposure time has harmful effects on the signal, constraining the attainable receiver bandwidth. Therefore, those effects must be faced, especially when the exposure is not adjustable. Section 5.2 presents a AI-assisted equalizer used to mitigate those harmful effects in different noise conditions. Furthermore, the fact that the laboratory uses generic surveillance cameras with almost no configurable settings revealed the need to develop new strategies for expanding the design of OCC links from fully-customizable laboratory cameras to the great diversity of cameras that floods the market. For this purpose, a novel AI-assisted architecture is proposed in Section 5.3. Each of these works raises and validates new hypotheses that are finally collected in Chapter 1.

5.1 Application of optical camera communication in industrial environments

Citation:
[20] C. Jurado-Verdu, V. Guerra, V. Matus, C. Almeida, and J. Rabadan, "Optical camera communication as an enabling technology for microalgae cultivation," *Sensors*, vol. 21, no. 5, 2021

This article addresses the general objective GO3 by analyzing the deployment of several OCC links in an industrial environment used to monitor intelligent photobioreactors. Moreover, it addresses the technical objective TO2.2, with the in-depth characterization of the camera that will be used in the subsequent experiments and with the design of the tools and methods for measuring the irradiance of different light sources, i.e., the bioreactor's outermost surface.

The research carried out in this article is related to a previous contribution included in Annex A[1]. This previous work developed a fully functional ad-hoc prototype for OCC that uses image correlation techniques as the primary tool for source detection (and tracking), channel equalization, and data decoding. The next step in this research involved deploying this prototype in a real environment. The chosen scenario consists of a pilot plant for the cultivation of microalgae. In this scenario, the OCC links will provide communication support to several bioreactors that feed microalgae with artificial LED light.

One of the main contributions of this work is a novel transparent-wall photobioreactor that has a custom intelligent lighting system with two simultaneous functions: (i) to provide light to the microalgae adapted to its growth and feeding cycles and (ii) to send optical codes containing data relative to the crop (i.e., temperature, pH levels, dissolved carbon, among others) to a surveillance camera. On the other hand, the surveillance camera sends images to a computer that performs three operations simultaneously: (i) to watch the laboratory to prevent the presence of intruders, (ii) to decode data embedded in the images sent by each photobioreactor, (iii) to estimate based on the received light crop's parameters such as its growth state, biomass concentration and, other parameters of interest.

Another contribution of this work is an in-depth analysis of the optimal deployment of photobioreactors (henceforth nodes) within an empty room. In any microalgae production driven with artificial light, space optimization is crucial[2], to offer a competitive solution against open-air plants, with the sun as the primary source of energy. Therefore, a feasible OCC deployment must efficiently exploit the available space while reserving some space free of obstructive elements for the camera to visualize all the nodes. Moreover, this space is required by laboratory technicians to carry out in-situ operations such as crop sampling or maintenance. This analysis extracted practical guidelines transferable to any indoor scenario where several transmitter nodes must be arranged to optimize the available space efficiently while providing a shared channel capacity using a single camera. This manuscript proposes a multidimensional metric to assess different deployments that consist of three variables: (i) N, the number of nodes monitored using a single camera, (ii) the space utilization ratio (SUR), which is the

---

[1] This contribution gives the student access to doctoral studies.

[2] Along with the optimization of the light quality (intensity profile, and light spectra) and delivery (frequency and duty cycles of on and off states).

ratio between the volume of the nodes and the volume of the room[3], and (iii) Q, a novel metric derived from the channel capacity equation proposed by Claude Shannon, that considers the optical irradiance of the light source and the projected size of the node in the image. This three-dimensional space allows evaluating various arrangements to choose the optimal configuration based on the requirements. The increase in N will imply a higger aggregate channel capacity, the SUR, a better use of space, and, finally, Q, a better communications performance.

Another contribution is the characterization of the light source in terms of irradiance, i.e., the outermost surface of the photobioreactor. This characterization is performed for increasing biomass concentrations, which results in higher attenuation and scattering. The results reveal that scattering can be beneficial, as it homogenizes the irradiance output from the surface, improving the binding quality for more obtuse viewing angles.

Finally, this work and the previous related work presented in a conference [24] brought to light the effect that the exposure time has on the signal, constraining the receiver bandwidth. After several experiments, the exposure time was fixed to 57 μs to (i) allow perceiving the surface of the container composed of sharp symbol bands while (ii) providing a maximum baud rate of 2151(Bd/reactor). However, at this exposure time, only the surface of the container is perceived in the image, as shown in Fig. 6[4] corresponding to the indoor experiment. Furthermore, Fig. 11 shows an ideal outdoor link, where the camera acts simultaneously as an imaging device and a receiver. As can be seen, the scene is correctly perceived in the image, while the colored symbol bands (as it uses three independent RGB channels for the communications) are sharp enough for decoding. The scene can be seen because the exposure time was increased to 100 μs, along with the analog gain, which introduces some noise. Consequently, the transmitter baud rate was halved to avoid the occurrence of ISI. This highlights the exposure time's impact on communications, limiting the attainable data rate. Moreover, the Xiaomi smartphone's camera used in this case has a greater sensitivity than the Raspberry Pi camera used in the indoor experiment.

---

[3] This metric can be replaced by the ratio of the surface area of the nodes to the total available surface area, if addressing a two-dimensional problem, e.g., when trying to fill a wall with several transmitter nodes.

[4] In that figure, no data is being transmitted. The LED is constantly "ON".

sensors

MDPI

Article

Optical Camera Communication as an Enabling Technology for Microalgae Cultivation †

Cristo Jurado-Verdu [1,*,‡], Victor Guerra [1,‡], Vicente Matus [1], Carlos Almeida [2] and Jose Rabadan [1]

[1] Institute for Technological Development and Innovation in Communications (IDeTIC), Universidad de Las Palmas de Gran Canaria (ULPGC), 35017 Las Palmas de Gran Canaria, Canary Islands, Spain; vguerra@idetic.eu (V.G.); vmatus@idetic.eu (V.M.), jrabadan@idetic.eu (J.R.)

[2] Spanish Bank of algae (BEA), Instituto de Oceanografía y Cambio Global (IOCAG), Fundación Canaria Parque Científico y Tecnológico, Universidad de Las Palmas de Gran Canaria (ULPGC), 35230 Las Palmas de Gran Canaria, Canary Islands, Spain; carlos.almeida@ulpgc.es

* Correspondence: cjurado@idetic.eu

† This paper is an extended version of our paper published in IEEE.

‡ These authors contributed equally to this work.

Abstract: Optical Camera Communication (OCC) systems have a potential application in microalgae production plants. In this work, a proof-of-concept prototype consisting of an artificial lighting photobioreactor is proposed. This reactor optimises the culture's photosynthetic efficiency while transmitting on-off keying signals to a rolling-shutter camera. Upon reception, both signal decoding and biomass concentration sensing are performed simultaneously using image processing techniques. Moreover, the communication channel's theoretical modelling, the data rate system's performance, and the plant distribution requirements and restrictions for a production-scale facility are detailed. A case study is conducted to classify three different node arrangements in a real facility, considering node visibility, channel capacity, and space exploitation. Finally, several experiments comprising radiance evaluation and Signal-to-Noise Ratio (SNR) computation are performed at different angles of view in both indoor and outdoor environments. It is observed that the Lambertian-like emission patterns are affected by increasing concentrations, reducing the effective emission angles. Furthermore, significant differences in the SNR, up to 20 dB, perceived along the illuminated surface (centre versus border), gradually reduce as light is affected by greater dispersion. The experimental analysis in terms of scattering and selective wavelength attenuation for green (*Arthrospira platensis*) and brown (*Rhodosorus marinus*) microalgae species determines that the selected strain must be considered in the development of this system.

Keywords: optical camera communications; visible light communications; microalgae cultivation; artificial lighting; light management; smart farming; Agriculture 4.0

check for updates

Citation: Jurado-Verdu, C; Guerra, V.; Matus, V.; Almeida, C.; Rabadan, J. Optical Camera Communication as an Enabling Technology for Microalgae Cultivation. *Sensors* **2021**, *21*, 1621. https://doi.org/10.3390/s21051621

Academic Editor: Luis Nero Alves

Received: 26 January 2021
Accepted: 22 February 2021
Published: 25 February 2021

1. Introduction

Microalgae culture has gained significant momentum during the last decade, triggered by the necessity of developing new and sustainable resources. They have become a promising alternative source for biofuels and biogas production, human and animal nutrition, cosmetics and bioactive supply for nutraceutical and pharmaceutical applications.

Microalgae biomass production is currently carried out using both open ponds (e.g., raceways) and closed photobioreactors [1]. The last ones are preferable at the laboratory and pilot-plant scales since, with the appropriate design, they can optimise growth conditions (nutrient levels, carbon concentration, temperature, acidity, among others). Among these parameters, light radiant energy is a capital factor that affects the photosynthetic efficiency, and therefore the overall productivity [2].

Generally, microalgae production plants are designed to take advantage of the Sun as the primary light source due to cost optimisation. However, the Sun's irradiance depends on several factors, such as weather conditions, latitude and day time. Furthermore, these open-air plants need vast extensions to be profitable since effective biomass production depends on the directly-exposed surface. Nonetheless, artificial lighting can provide advantages in photosynthetic efficiency (custom spectrum and intensity profiles) and the tight control they offer concerning microalgal biochemistry and growth, increasing industrial processes' reliability. Using this type of lighting, biomass production depends no longer on the plant's exposed surface, but on its equivalent volume. Although optimising light quality could considerably reduce energy consumption, another approach concerns how light is delivered to the culture: flashing (pulsed) light instead of continuous illumination.

In this context, pulsed illumination could serve as a communication link, enabling the re-utilisation of light sources as effective visible light communication (VLC) transmitters [3]. The low bandwidth transmission channel they provide (attaining the stated frequency restrictions) could be used for online monitoring of the culture's state conditions via deployable sensors within the photobioreactor module.

Nevertheless, there are still a few parameters that are either too complex to be measured in situ or imply invasive methods. These parameters are microalgae biomass concentration and its growth phase. In [4,5], different approaches based on the digital processing analysis of red-green-blue (RGB) images for low cost, fast and accurate quantification of biomass concentration were proposed and experimentally validated. Therefore, cameras could be used as sensing devices for these parameters.

On the other hand, the use of cameras as communication receivers for VLC links is a research topic that is receiving significant attention. Furthermore, there is a current standard specification by the working group IEEE 802.15.7 [6], which has finally integrated this new strategy known as optical camera communication (OCC) [7–10]. These devices are bandwidth-limited compared to traditional photodetectors, such as p-type, intrinsic, n-type (PIN) and avalanche photodiodes (PDs). However, due to the use of image-forming optics [11], cameras can receive light from multiple sources, providing inherent spatial multiplexing capabilities [12–14], which can be easily exploited for simultaneous monitoring within a microalgae production plant.

This work proposes the use of low-cost cameras for both remotely sensing microalgae culture parameters and for establishing a direct communication link with flat-panel photobioreactors. Compared to other radio or wired technologies, the use of OCC has the following advantages. First of all, this technology reuses the light that comes from the photobioreactors and makes better use of this excess of energy that would otherwise be wasted. Second, it replaces a generic communications coordinator with an intelligent camera that performs additional routines apart from establishing a communication link. On the one hand, it carries out continuous surveillance of the reactors and the personnel who access the room and can act as an early warning system. On the other hand, it simplifies the estimation of some culture parameters that are difficult to measure or involve intrusive procedures, such as the concentration of biomass or the strain's growth state. For example, for evaluating the biomass concentration, high-cost equipment must be used to automate the extraction of representative samples. In the worst case, this procedure must be carried out periodically by the staff. Furthermore, concerning communications security, an extra layer of protection is added thanks to the light confinement within the room. Finally, cameras' inherent spatial multiplexing capabilities significantly simplify link protocols, making communications more robust and less prone to errors.

This work presents part of the ATICCuA Project results, a multidisciplinary research project carried out by the Spanish Bank of Algae (BEA by its Spanish acronym) and the Photonics and Communications Division of the Institute for Technology Development and Innovation in Communications (IDeTIC). This project addresses the development of prototypes based on the visible light communication (VLC) and, more specifically, those used in underwater wireless optical communications (UWOC) for application in microalgae culture systems. Precisely, its main objective consists of the design of an LED-based dual-use system, which provides configurable lighting for the culture and production of microalgae and cyanobacteria and optical wireless communication capabilities for optical underwater channel characterisation.

This work introduces and discusses the microalgae cultivation restrictions that affect the communications' performance and the overall system's design. In addition, a comprehensive analysis based on geometrical constraints is carried out, providing results such as optimal camera positioning concerning a custom-defined metric that relates the aggregate data rate and effective space exploitation. From this analysis, a preliminary study of the plant distribution of a case study is carried out. Furthermore, several experiments that evaluate the system's communications performance are conducted in an indoor and outdoor scenario.

The remainder of the paper is structured as follows. Section 2 introduces the proposed OCC-based architecture in a bottom-up manner. Section 3 groups together the discussion of the channel model, the analysis of the data rate achievable for each container and its optimal distribution in the plant for the efficient deployment of this technology. Section 4 describes the methods, materials and procedures involved, on the one hand, in the study of the plant distribution for a real application case and, on the other hand, in the two experiments conducted for the evaluation of the prototype based on the quality of the optical signal received by the camera. Section 5 presents the results obtained, their interpretation and a discussion. The conclusions of this work are summarised in Section 6.

2. Proposed Architecture

The proposed architecture follows a many-to-one unidirectional network topology where photobioreactor nodes transmit sensor-related data to a receiver camera node (Figure 1).

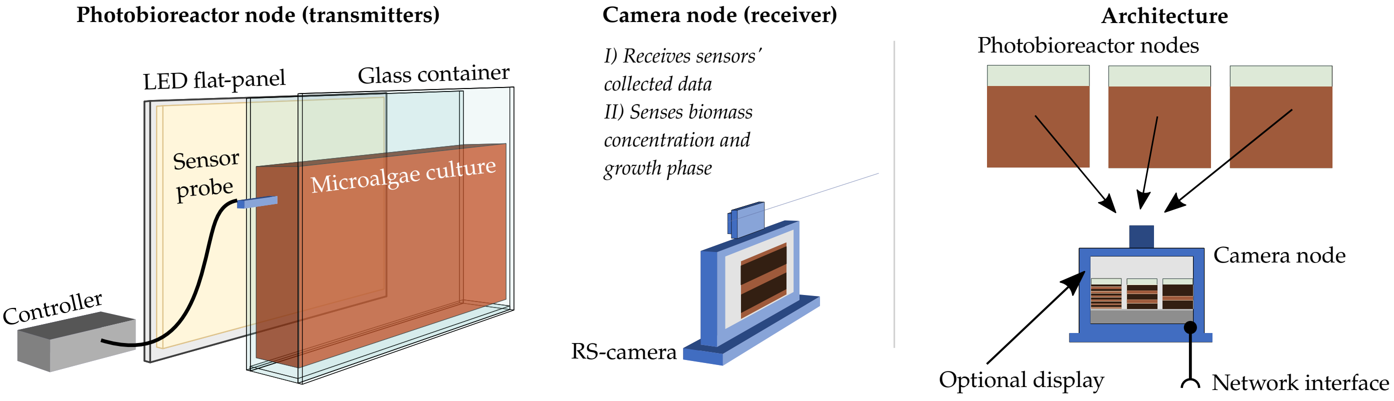


Figure 1. Proposed architecture based on flat-panel photobioreactors. RS, rolling-shutter.

The photobioreactor nodes comprise a uniform-radiance LED flat panel attached to a glass container that holds the microalgae culture. The receiver node consists of a rolling-shutter (RS) camera connected to the communication endpoint that exposes a control interface and a standard wired communication bus. The transmitter nodes use two different signals to establish a link: a well-known beacon signal that uniquely identifies the node and data packets composed of Manchester-encoded pulses. The receiver uses the beacon signal to perform three separate routines: establishing the communication link, estimating the channel response to complete the decoder training, and finally, sensing the culture by estimating the biomass concentration and growth phase of the microalgae species. Therefore, this system performs two functions. It establishes a monitoring link and directly senses essential culture parameters.

In the following sections, both nodes are presented, highlighting the influence of the microalgae cultivation needs and requirements in their design.

2.1. Photobioreactor Node

The transmitter is comprised two parts: the LED flat panel and the LED driver that generates the corresponding OCC driving signal to transmit sensor data provided by probes deployed inside the container (such as temperature, acidity, carbon and nutrient levels). This section is focused on light generation and transmission.

As was previously mentioned, the competitiveness of any artificial light-driven microalgal production requires improvements in photon harvesting and the conversion efficiency of light sources [2]. Hence, it is crucial to optimise light quality and delivery. In terms of light energy, the photosynthetic rate is directly related to the irradiance power, and excessive or insufficient incident light constrains optimal performance and may induce the photo-inhibition and photo-oxidation of the cells, eventually attaining photo-damage and even leading to culture death [15]. In terms of light spectra, the radiant energy absorbed by microalgae highly depends on the chemical nature of their native pigments, which have specific absorption bands in the visible and near-infrared spectra. Thus, better energy usage can be achieved by adjusting the light source's emission to match the absorption spectrum [16].

On the other hand, recent experimental studies have shown that combining short and intense light flashes with extended dark periods instead of continuous illumination might increase the culture's growth efficiency, as discussed in [17,18]. The light frequency and duty cycle are particular to each microalgae species and the expected biomass product result (lipids, carotenoids, etc.). It may vary between a few Hz up to tens of kHz.

Therefore, the transmitting source's design must consider all these restrictions: light quality (intensity profile, spectra) and light delivery (frequency and duty cycles) concerning the selected strain and the expected results.

2.2. Camera Node

The acquisition mechanism of image sensors inherently limits the available bandwidth for communications. In global-shutter (GS) cameras, the whole image sensor is exposed simultaneously. Thus, the achievable data rate is upper-bounded by their frame rate, restraining the light source's switching frequency considerably. On the other hand, RS cameras scan the scene sequentially row by row of pixels (usually on the shorter dimension of the sensor), allowing capturing different light states (intensity, colour variations, among others) within the lamp's source projection in the frame [19]. In this case, the theoretical bandwidth limit is imposed by the row shift time, commonly denominated the sampling time, t_s, which is the fixed duration between the start of a row scan and the consecutive one. Furthermore, these rows are not disjointedly exposed, but in an overlapping manner. This overlap duration depends on the configured exposure time, t_{exp}, which is the span of time during which each row of pixels is integrating light. In terms of communications, the effect of this overlap can be modelled (for a uniform light source) as the product of a weighted moving-average filter, with its corresponding transfer function Equation (1), which further restricts the effective bandwidth (cut-off frequency—Equation (2)).

$$H(w) = \frac{e^{-jw/2(N-1)}}{N} \frac{sin(wN/2)}{sin(w/2)} \tag{1}$$

where N is an integer that relates the exposure time to the sampling time ($N = t_{exp}/t_s$) and w is the normalized angular frequency in radians per sample ($w = 2\pi f/f_s$).

The cut-off frequency, w_{3dB}, defined as the half-power point's frequency, can be computed using the modulus squared function of the transfer function using Equation (2).

$$|H(w_{3dB})|^2 = \frac{1}{N^2} \frac{sin^2(\frac{w_{3dB}N}{2})}{sin^2(\frac{w_{3dB}}{2})} = \frac{1}{2} \tag{2}$$

Equation (2) does not have a general analytical solution. However, it is possible to rely on numerical methods such as Newton–Raphson's algorithm to determine the cut-off frequency that sets the system's available bandwidth (Equation (3)).

$$f_{tx} \leq f3\text{dB}/2 \tag{3}$$

Therefore, to increase the effective transmission bandwidth, it is necessary to minimise the camera's exposure time. Consequently, the PDs of the pixels exposed for a shorter time have a lower signal-to-noise ratio (SNR). In previous works, it was shown that the reduced received power due to the extinction along the path or due to low exposure times can be overcome using the analogue amplifier of the CMOS camera by increasing the gain it provides before the analogue-to-digital converter (ADC), ultimately improving the SNR [20,21].

3. Communications Modelling

The proposed architecture presents some specific characteristics from the communication system point of view. First, the light propagates through different media from the emitter to the camera, which modifies the received optical signal. On the other hand, the camera position affects the communications' performance since it determines the received power and the emitter image size. Furthermore, the distribution and size of the photobioreactors also impact the complete OCC system performance. In this section, all these aspects are addressed.

3.1. Communication Channel

The channel can be divided into different layers. Figure 2 details the layered version of the channel, which is composed of: the inner air gap layer, the inner glass layer, the microalgae suspension in the water layer and the outer glass layer that emits the light that reaches the camera. This work focuses on the analysis of these four primary layers of the channel. The effect of the link's air gap between the container and the camera was addressed previously in [22].

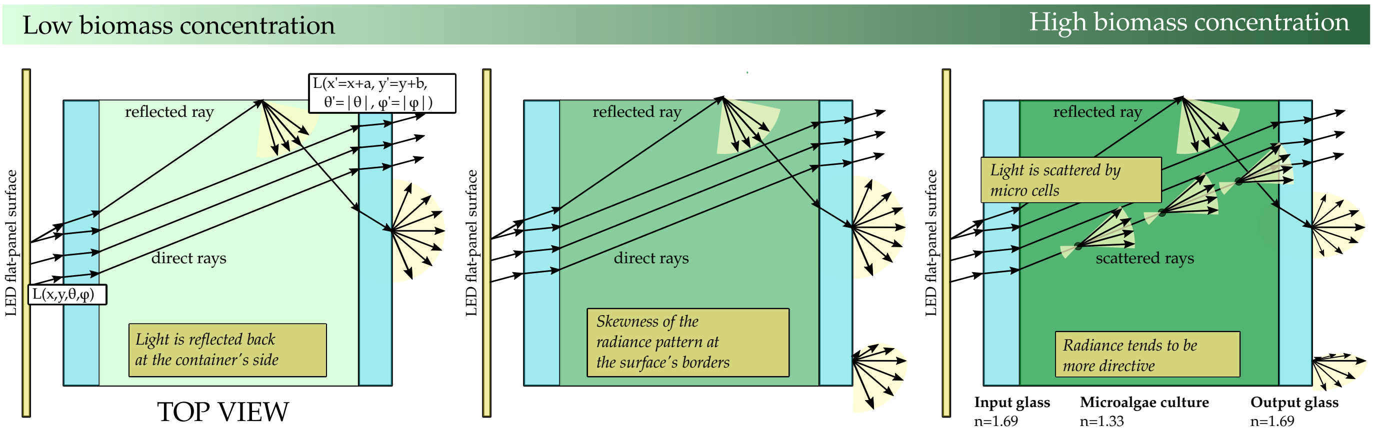


Figure 2. Layered version of the channel.

This preliminary analysis shows two critical interfaces: the inner glass/water and the outer glass/air interfaces that affect communications' performance and constrain harvesting optimisation.

Light rays emitted in a specific direction by the light source (ϑ_0, φ_0) change their trajectory upon reaching the surface of the inner glass ($\theta_{\text{ig}}, \varphi_{\text{ig}}$). Then, those rays reach the first critical interface (inner glass/water interface) and undergo partial or total internal reflection depending on the critical angle ($\Theta_{\text{iglass/water}} = 51.9°$) (considering refraction indices of $n_{air} = 1.0$, $n_{glass} = 1.69$ and $n_{water} = 1.33$ for the air, glass and water media, respectively). However, total internal reflection will never happen at this critical interface in the considered conditions. The reason is that as soon as θ_0 approaches 90 degrees, θ_{ig} tends to 31°, which is considerably lower than $\Theta_{glass/water}$. In the microalgae medium, light rays travel with direction (ϑ_{water}, φ_{water}). Those rays that reach the top, bottom, left or right side reflectors of the container are reflected. Figure 2 shows an example of one ray reaching the container's right side and being reflected.

Finally, within the outer glass, light propagates with (ϑ_{og}, φ_{og}) until it arrives at the second critical interface. The outer glass-air interface has its corresponding critical angle ($\Theta_{glass/air} = 36.28°$). Light rays with incident angles greater than $\Theta_{glass/air}$ are reflected, reducing the total optical power leaving the photobioreactor.

This preliminary analysis aims to better understand the behaviour of the radiance $L(\overrightarrow{r}, \overrightarrow{\theta})$ of the extended outer glass surface (Equation (4)); in other words, the radiant intensity ϕ, emitted from an infinitesimal unit surface, $d_{A_\perp}$, and contained within a unit of solid angle aligned normal to the direction of interest, $\overrightarrow{\theta} = (\vartheta, \varphi)$, for a particular location, $\overrightarrow{r} = (x, y)$.

$$L(x, y, z = z_0 \vartheta, \varphi) = \frac{d^2\phi(x, y, z = z_0, \vartheta, \varphi)}{d\Omega dA cos\vartheta} \tag{4}$$

After characterising the radiance of the surface, the total irradiance over a pixel is obtained by integrating the incoming radiance from any direction in the normal hemisphere that encloses the pixel area Equation (5).

As an example, Figure 3 shows the irradiance per unit of area of an infinitesimal portion, d_V, of a virtual surface that encloses the container keeping the same horizontal distance to the container's centre point.

$$E(x, y, z) = \int_\Omega L(x, y, z = z_0, \vartheta, \varphi) cos\vartheta d\Omega \tag{5}$$

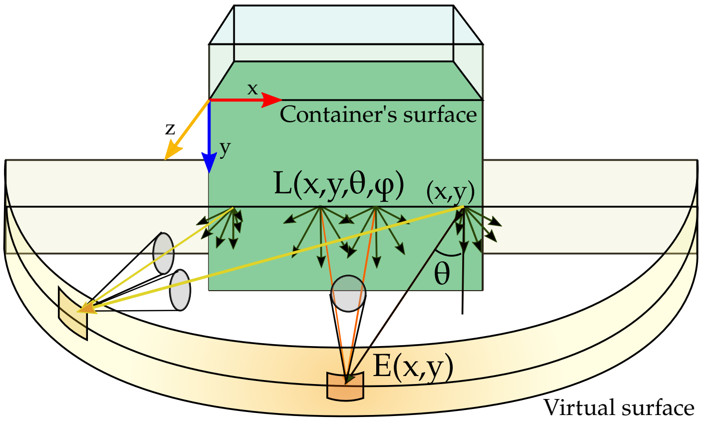


Figure 3. Virtual surface that encloses the container.

In the stated conditions, the radiance would be affected by different phenomena: the culture's absorption capacity, the shading effect between cells, and light scattering. When the amount of biomass is negligible compared to the volume of water in the container, these phenomena might be neglected, and radiance can be estimated using geometrical optics. Now, considering that the channel consists of symmetrically repeating parallel layers that start and end in the same air medium, light's incident angle ϑ_0, φ_0 coincides with exiting radiance angle $\vartheta_{out}, \varphi_{out}$ without any hard restriction as there is no total internal reflection. Figure 2 shows light example paths in 2D dimensions.

Now, considering that the bottom, top, left and right container's sides reflect light in a specular manner, the radiance of a point $(x', y', z = 0)$ of the external surface in any direction $\overrightarrow{\vartheta'}$ is a fraction of the radiance emitted by the light source in the same direction (or in a shifted direction, if the ray comes from a reflection in the walls) ($\overrightarrow{\vartheta'}$), but from a translated origin point, x,y, of the output point ($x = x' + a, y = y' + b$).

As a result, the output radiance from infinitesimal surfaces located at the container's centre tends to mimic the original lamp radiation pattern. However, the output radiance at the borders of the container tends to skew. This skewness occurs because there are no direct light contributions. This skewness effect of the radiance pattern significantly constrains the communications' performance, as the optical signal power recovered from specific viewpoints would not be enough to establish effective communication. As it can be extracted from preliminary experiments, this skewness can be partially reduced if the container's sides reflection has a perfect diffuse behaviour, causing the light to be uniformly spread in all directions. However, the skewness cannot be fully mitigated since the light rays that contribute to reducing this effect come from very steep entry angles and repeatedly bounce within the container boundaries, ultimately reducing its optical power.

Regarding light power exiting the surface, it is important to consider the fraction of the incident light reflected at each interface. Considering non-polarised light the effective reflection coefficient, for each input angle, ϑ_i, can be expressed by Equation (6).

$$R_{\text{eff}} = \frac{(R_s + R_p)}{2} \tag{6}$$

where R_s (Equation (7)) and R_p (Equation (8)) are the reflectances for s-polarized and p-polarized light, respectively.

$$R_s = \left| \frac{n_1 \cos\vartheta_i - n_2\sqrt{1 - (n_1/n_2 \cdot \sin\vartheta_i)^2}}{n_1 \cos\vartheta_i + n_2\sqrt{1 - (n_1/n_2 \cdot \sin\vartheta_i)^2}} \right|^2 \tag{7}$$

$$R_s = \left| \frac{n_1\sqrt{1 - (n_1/n_2 \cdot \sin\vartheta_i)^2} - n_2 \cos\vartheta_i}{n_1\sqrt{1 - (n_1/n_2 \cdot \sin\vartheta_i)^2} + n_2 \cos\vartheta_i} \right|^2 \tag{8}$$

where n_1 and n_2 are the refractive indexes of both mediums. Hence, light rays with acute entry angles have a significant amount of power reflected. This has some implications. At first, not all the optical power reaches the microalgae culture, and it will depend on the radiance pattern of the surface and the effective emission angle of the source. Finally, not all the optical power leaves the container's surface, which is not desirable for establishing an optical communication link despite being suitable for cultivation. Under this initial configuration, when the biomass concentration is above a threshold level, the scattering and the absorption phenomena cannot be neglected. In that case, Beer–Lambert's law can be used to describe the attenuation of light due to absorption by the biomass concentration (Equation (9)). This equation states that the attenuation of light over a distance is proportional to the light intensity, where C is the volumetric absorption coefficient. The latter is the product of the specific light absorption coefficient, $\alpha_{x,\lambda}$, and the biomass concentration (ρ). The integration of Equation (9) over the light path, taking into account the wavelength dependency, results in Equation (10).

$$\frac{dI_\lambda(x)}{d(x)} = -C_{x,\lambda} \cdot I = -\alpha_{x,\lambda} \cdot \rho \cdot I \tag{9}$$

$$I(x) = \sum_{\lambda=800}^{\lambda=300} I_\lambda(0) \cdot e^{-\alpha_{z,\lambda} \cdot \rho \cdot x} \tag{10}$$

However, as light travels through the photobioreactor, it is absorbed and scattered by the microalgae. The light intensity $I_\lambda(r, \overrightarrow{d})$ ($\mathrm{Wm^{-2}sr^{-1}}$) at a given location within the container, r, and in the direction $\overrightarrow{s}$ can be determined by solving the radiative transfer equation (RTE) [23], which represents an energy balance on the radiative energy travelling along a particular direction. In steady-state conditions, the RTE for non-collimated light, as in this case, can be expressed as Equation (11).

$$\overrightarrow{s} \cdot \nabla I_\lambda(r, \overrightarrow{d}) = -\beta_\lambda I_\lambda(r, \overrightarrow{d}) + \frac{\sigma_\lambda}{4\pi} \int_{4\pi} I_\lambda(r, \overrightarrow{d}) \cdot \Phi_\lambda(\overrightarrow{s_i} \to \overrightarrow{s}) d\Omega \tag{11}$$

where Φ_λ is the scattering phase function (SPF), which represents the angular distribution of the scattered light, in other words, the probability that radiation travelling in a given direction, $\overrightarrow{s_i}$, will be scattered to the direction, $\overrightarrow{s}$, of interest. This function is determined by the size, shape and refractive index distribution of the scattering particle. β_λ is the extinction coefficient (m^{-1}) composed by the scattering coefficient, σ_λ, and the absorption coefficient, κ_λ (m^{-1}). These two parameters can be expressed in terms of the averaged scattering C_{sca} and absorption C_{abs} cross-sections (m^2), respectively (Equation (12)).

$$\sigma_\lambda = C_{\text{sca}} \cdot N \quad \text{and} \quad \kappa_\lambda = C_{\text{abs}} \cdot N \tag{12}$$

where N is the microorganism concentration expressed in number of cells per m^3 of water suspension.

This RTE equation given by Equation (11) reveals that the scattering absorption cross-sections and the SPF have an important role in predicting light transfer in photobioreactors for simulation, design and optimisation purposes. However, these characteristics are interrelated and difficult to estimate from the electromagnetic wave perspective, given the microorganisms' complex morphology. Nonetheless, they can still be measured experimentally with more or less difficulty and related to some parameters of the cell, as detailed in [24]. Furthermore, these radiative properties of microalgae (the absorption cross-section, scattering cross-section and scattering phase function) vary significantly over time, depending on the strain's growth stage, as shown by the studies [25,26]. As all these radiative parameters depend on the wavelength of light and because they vary in such a significant way through time, it can be considered that the values obtained by an RGB camera would correctly identify not only the strain, but its growth's state.

Regarding the communications' performance, although the presence of these microorganisms attenuates the signal before it reaches the exposed container's surface, it may be beneficial in some cases because it distributes the optical power more evenly, uniforming the radiation pattern over the entire container's surface. In conclusion, microalgae particles' scattering phenomena can be exploited for communications in scenarios where the camera is not facing perfectly perpendicular to the transmitter surface.

3.2. Data Rate Analysis

The maximum achievable data rate for a photobioreactor depends on its projection's vertical size over the image; in other words, the total number of vertical samples (rows of pixels) recovered from the signal [27]. Consequently, it is necessary to consider the scenario's geometrical configuration, the relative positions between the camera and the photobioreactors and their size.

Equations (13) and (14) relate the projection dimensions (in meters) of a rectangular surface over the scene plane, H_p, W_p, with its corresponding pixels dimensions, h_{roi}, w_{roi}. Figure 4 shows all the geometry and camera parameters involved in a generalised scenario.

$$h_{roi} = H_p \cdot \frac{h_{\text{res}}}{2 \cdot \tan(\text{AoV}_\text{V}/2) \cdot \text{D}_\text{FoV}} \quad (pixels) \tag{13}$$

$$w_{roi} = W_p \cdot \frac{w_{\text{res}}}{2 \cdot \tan(\text{AoV}_\text{H}/2) \cdot \text{D}_\text{FoV}} \quad (pixels) \tag{14}$$

where h_{res} and w_{res} are the pixel dimensions of the image, AoV_h and AoV_w the horizontal and vertical angle of view (AoV) of the camera (in degrees), respectively, and D_{FoV} the distance (in meters) from the camera lens to the scene plane.

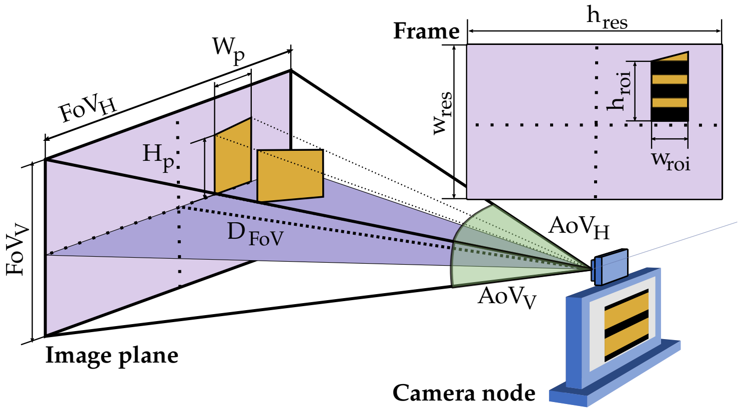


Figure 4. Geometrical parameters involved in the computation of the maximum achievable node's data rate.

Following Nyquist's criterion ($f_{\mathrm{tx}} \leq f_{\mathrm{3dB}}/2$), the minimum number of vertical samples of the signal required per symbol can be computed using Equation (15).

$$h_s = \frac{f_{\mathrm{s}}}{f_{\mathrm{tx}}} \leq \frac{2}{f_{\mathrm{3dB}} \cdot t_{\mathrm{s}}} \quad (\mathrm{samples}) \tag{15}$$

It must be remarked that communication takes place in a windowed manner. Leaving aside the reflections with the objects present in the scene [13,28], the signal has to be recovered mainly from the projection within the image of the light source where the signal quality is considerably better. Hence, during the acquisition, only a fraction of the total data sent is sampled by the sensor. While the sensor is not scanning the transmitter's surface, but another part of the scene, it will remain blind to the light changes of the transmission [19]. To overcome these blind periods, the data packets must be sent repeatedly (at least while the camera is acquiring two full frames). In addition, to avoid packet losses, the source's image projection, h_p, must fit at least two complete packets, as was detailed in the previous work [29]. With these restrictions (Equation (17)), the overall transfer rate (in bauds) is obtained from Equation (16).

$$R_b = S_{packet} \cdot \frac{\mathrm{fps}}{2} \tag{16}$$

$$h_{\mathrm{packet}} \geq 2 \cdot S_{\mathrm{packet}} \cdot h_s \tag{17}$$

where S_{packet} is the equivalent number of symbols per data packet. On the other hand, the projection's width in pixels (number of columns) also plays an important role in communications. Pixels located in the same row are exposed to light simultaneously, and they can be used to filter out noise, thus strengthening the signal-to-noise ratio. Moreover, wider areas aid source discovery and tracking and considerably ease the decoding routine. As a consequence, a minimum pixel width must be selected as a design requirement.

Finally, it is worth mentioning that the proposed system's achievable capacity considerably exceeds the requirements for accurate monitoring of the culture parameters. To obtain a preliminary idea of these requirements, the following considerations are taken into account. The selected parameters to be measured are acidity, temperature, light conditions, carbon, nutrients, inhibitors presence and O_2 degassing. The packet's payload allocates twenty bytes per parameter, which is a much larger allocation than necessary. Furthermore, given that the culture's temporal evolution is considerably slow, samples of the culture can be taken reliably every 5 min, which offers a substantially acceptable temporal resolution. Based on these considerations, the required capacity is about four bits per second. Therefore, the hundreds of bauds per second that can be sent using the proposed system, as specified in Section 4.1, satisfy the stated capacity requirements.

3.3. Plant Distribution

The distribution of the photobioreactor nodes across the room plays an essential role in communications. To establish a link, the receiver should visualise each node, and consequently, it is necessary to reserve some space free of obstructive interference between the transmitters and the camera. This reveals the importance of analysing different possible solutions for placing the nodes until finding the one that best suits the project's initial requirements, either in terms of better link quality, higher capacity or more efficient space exploitation.

In a previous work [22], a metric was proposed to compare the performance of different solutions. Nevertheless, the complex nature of this problem, which involves several variables and a wide range of possible initial requirements, highlighted the need to modify this classification strategy. This work proposes replacing the original metric based on a single value with a modified multidimensional metric, Equation (18), allowing a more flexible classification of the arrangements. This new metric offers a comparison tool in which it is left to the designer's discretion to select a specific application design's priorities.

$$F : \{N, Q, SUR\} \tag{18}$$

$$\text{where :} \quad \mathrm{SUR} = \frac{\mathrm{V_{cont}}}{\mathrm{V_{room}}} \tag{19}$$

It consists of three independent variables: the number of simultaneous photobioreactors monitored by a single camera, N, a communications performance metric, Q, and the space utilisation ratio (SUR). The first two variables account for the overall achievable data rate of the setup. The last term relates the total equivalent volume of the containers bound to biomass harvesting, $\mathrm{V_{cont}}$, and the minimum room volume needed, $\mathrm{V_{room}}$. The optimisation of this term has direct implications in the reduction of production costs and, therefore, the viability and competitiveness of cultivation plants based on artificial lightning and vertical racks.

The proposed Q metric is derived from the well-known Shannon–Hartley's equation [30] for estimating the channel capacity for each container (Equation (20)).

$$C = BW \cdot log_2(1 + S/N) \tag{20}$$

where C is the capacity in bits per second, BW is the bandwidth of the channel in Hertz and S/N is the SNR, expressed as a linear power ratio. As was previously mentioned, communications happen in a windowed manner. In other words, the transmission effectively takes place during a fraction of the time to acquire a frame, t_{frame}. This fraction of time, or channel availability, τ_{rx}, depends on the geometrical configuration, t_{geo}, and on the elapsed time between when the camera finishes capturing one frame and starts with the next one t_{inter} (Equation (22)). However, this last term is considerably lower than t_{frame} and can be neglected.

Adding this factor to Equation (20), the effective capacity is obtained for each container, as Equation (21) shows.

$$C = BW \cdot log_2(1 + S/N) \cdot \tau_{\mathrm{rx}} \tag{21}$$

$$\text{where :} \quad \tau_{\mathrm{rx}} = \frac{t_{\mathrm{geo}}}{t_{\mathrm{frame}} + t_{\mathrm{inter}}} \approx \frac{h_{roi}}{h_{res}} \tag{22}$$

The proposed metric Q is then defined as the relationship between the channel's capacity for a particular arrangement and container, C, and the ideal capacity, C_{ideal}, Equation (23).

$$Q = \frac{C}{\mathrm{C_{ideal}}} = \frac{BW}{BW_{\mathrm{ideal}}} \cdot \frac{log_2(1 + S/N)}{log_2(1 + (S/N)_{\mathrm{ideal}})} \cdot \frac{\tau_{\mathrm{rx}}}{\tau_{\mathrm{rx,ideal}}} \tag{23}$$

The vast number of parameters involved in this metric's computation makes this metric unreasonable for a practical analysis of different plant distributions, even more so if there are plenty of configurations to be compared. Hence, for this metric to be a useful tool for this purpose, it is necessary to assume a series of coarse approximations. These approximations would greatly simplify the practical comparison of two cases without incurring harsh penalties. First, it is important to remark that the camera's hardware and configuration would not change. Consequently, the bandwidth, which depends on the sampling frequency and the image sensor's exposure time, remains constant ($BW' = BW_{\text{ideal}}$). On the other hand, the maximum capacity will be achieved when the transmitting source occupies the image entirely in the scanning dimension (usually from top to bottom). In this way, the link availability is kept while capturing a frame ($\tau_{\text{rx,ideal}} = 1$). With these considerations, Equation (23) could be reduced to Equation (24).

$$Q = \frac{log_2(1 + S/N)}{log_2(1 + (S/N)_{\text{ideal}})} \cdot \tau_{\text{rx}} \tag{24}$$

Now, using the first order Taylor approximation of the logarithmic function (at x = 0), Equation (25), results in Q as given by Equation (26).

$$\lim_{x \to 0} log_2(1 + x) = \frac{x}{ln(2)} \tag{25}$$

$$Q \approx \frac{S/N}{(S/N)_{\text{ideal}}} \cdot \frac{h_{\text{roi}}}{h_{\text{res}}} \tag{26}$$

In this way, the measurement of the capacity of a placement can be approximated to the comparison of the SNR with the best case and the size of its projection in the image. However, assumptions regarding the relationship between the SNR under study and the best-case SNR can simplify the analysis. These assumptions are detailed below. First, the container radiance varies smoothly through the entire surface, or at least in the area were the signal is recovered. Not the whole surface of the container is utilised for signal detection, but a fraction of it. The radiance at any given direction is expressed with respect to the maximum radiance L_{max} (Equation (27)).

$$L(\vartheta, \varphi) = L_{\text{max}} \hat{L}(\vartheta, \varphi) \tag{27}$$

The pixel FoV is small enough to assume that the pixel irradiance's contributions come from the same emitter radiance's output angles. Furthermore, the use of image-forming optics compensates the power loss due to spherical propagation with the projected size of the optical source on the image sensor [31]. Therefore, the power received by a pixel (Equation (28)) in a given direction can be approximated as the ratio of the total received power in the direction of maximum radiation, and it can be related to the ratio of the emitted radiance.

$$S_{\text{pixel}}(\vartheta, \varphi) = S_{\text{pixel,max}} \cdot \hat{L}(\vartheta, \varphi) \tag{28}$$

Furthermore, assuming that the region of interest (ROI) where the signal will be recovered is constant in width and height, then the total received power would be the aggregation of all contributions from the ROI's pixels. Therefore, the optical power received within the ROI is approximated by Equation (29).

$$S_{\text{roi}}(\vartheta, \varphi) = S_{\text{roi,max}} \cdot \hat{L}(\vartheta, \varphi) \tag{29}$$

Regarding the noise power, no external interfering optical sources were considered for simplicity. The primary noise sources are thermal noise (σ_{th}), shot noise (σ_{shot}) and quantisation noise (σ_{adc}), which is generated by the ADC of the camera. Besides, thermal noise does not depend on the signal power, whilst shot noise is affected by the received optical power. Nonetheless, considering the application scenarios of this work, it is expected that the shot noise contribution could be neglected with respect to thermal noise. Hence, the noise power is the same regardless of the receiver's position.

Therefore, comparing the radiation at different viewing angles can give a good approximated idea about the difference in the SNR of different container's arrangements.

$$\frac{(S/N)}{(S/N)_{\text{ideal}}} \approx \hat{L}(\vartheta, \varphi) \tag{30}$$

$$\text{where :} \quad \hat{L}(\vartheta, \varphi) = \frac{L(\vartheta, \varphi)}{\max\{L(\vartheta, \varphi)\}} \tag{31}$$

In conclusion, the final approximation of Q (Equation (32)) is obtained by combining Equations (24) and (30).

$$Q \approx \hat{L}(\vartheta, \varphi) \cdot \tau_{\text{rx}} \tag{32}$$

As mentioned, this metric is intended to be easy to compute and practical when guiding the engineer to distribute the plant most optimally. The approximations stated above greatly simplify the comparison of two cases without incurring non-affordable penalties. This Q parameter allows indirectly approximating the maximum transmission rate achievable by each photobioreactor node (Equation (17)).

In this work, the minimum value of Q ($Q|^{\text{min}} = min\{Q_{c,1}, Q_{c,2}, ..., Q_{c,N}\}$), of all the containers, is used to evaluate a particular arrangement. The reason is that as an initial requirement, all the containers must share the same data rate. To achieve this, all the emitters will adapt their transmission to the minimum available in the scene. It should be clarified that this metric varies between zero and one, where one implies the maximum theoretical performance (optimised use of channel capacity).

4. Methodology

In this work, a preliminary analysis of the plant distribution of a case study was carried out. In addition, several experiments to evaluate the system's signal reception in different scenarios were conducted. In the following sections, these experiments are described separately following the scheme: description, materials and resources, methods and data analysis.

4.1. Plant Distribution Study

This section describes the preliminary study of the best plant distribution for a real case study. In this case, it is intended to locate 12 custom containers of 150 × 50 × 9 cm (height by width by depth) in a 300 × 200 × 300 cm empty room. These containers must be placed in individual racks capable of holding up to 4 containers, two at the top and two at the bottom.

In terms of communication performance, those cases must satisfy the following derived constraints. The top photobioreactors will fill the upper half of the image and the lower ones the lower half. Consequently, the camera should always point to the shelf's vertical centre, and it will be aligned with respect to the centre (in the vertical dimension). In this way, the transfer rate is ensured to be equal for each reactor, with the number of vertical pixels equal to half of the image's vertical resolution (Equation (33)). The minimum pixel width of any reactor was set to 30 pixels (Equation (34)). This value was selected after conducting preliminary experimental tests. Finally, it must be highlighted that the sensor's aspect ratio relates both the vertical and the horizontal FoV, aspect$_{\text{ratio}}$; thus, altering one of them will influence the other.

To properly address this analysis, it is necessary to select a camera as an example. The camera selected is the PiCamera v2, detailed in Table 1. Using this camera with an exposure time of 57 µs and attending to the previously stated conditions, the maximum baud rate achievable is approximately 717 (Bd/reactor) per available channel. Considering three independent communication channels, red, green and blue, the maximum baud rate is 2151 (Bd/reactor).

This value is obtained by combining Equations (15), (16), (17) and (33) into Equation (36). The cut-off frequency was computed using Newton–Raphson's algorithm and is approximately 8213 Hz.

$$h_{roi} = H_p \cdot \frac{h_{\text{res}}}{\text{FoV}_V} = \frac{h_{\text{res}}}{2} \quad (\text{pixels}) \tag{33}$$

$$w_{roi} = W_p \cdot \frac{w_{\text{res}}}{\text{FoV}_H} \geq 30 \quad (\text{pixels}) \tag{34}$$

$$\text{where}: \quad \text{AoV}_H = \text{AoV}_V \cdot \text{aspect}_{\text{ratio}} \tag{35}$$

$$R_{\text{b}} = S_{\text{packet}} \cdot \frac{fps}{2} = \frac{h_{\text{roi}}}{h_{\text{s}} \cdot 2} \cdot \frac{fps}{2} = \frac{h_{\text{res}}}{4} \cdot \frac{f_{3\text{dB}} \cdot t_{\text{s}}}{2} \cdot \frac{fps}{2} \tag{36}$$

Table 1. Resources and equipment. * (wavelengths (nm): 630 (red).

Photobioreactor Node		Camera Node	
Part	*Parameters*	**Part**	*Parameters*
LED Lamp	Eglo Tunable White - RGB connect - 1 white cold LED (6500K) - 1 white warm LED (2700K) - 1 RBG LED * 530 (green), 475 (blue))	**Camera**	PiCamera Version 2 - Image sensor: Sony IMX586 [32] - Aperture lens: f/2 - Focal length (equivalent) (mm): 3 - Image resolution (px): 3280 × 2464 - Sampling time t_{s} (µs): 18.904
Container	Square glass panels (custom) - Dimensions (cm) : 50 × 50 × 9	**Receiver**	Raspberry Pi 3 Model B

Taking into account all these starting requirements, three different cases are proposed. Those cases are shown in Figure 5. In the first case, Case I, the camera is perfectly aligned with the shelf's centre. In Case II, the camera is shifted on the y-axis, and it views the rack from the side. In the last case (Case III), another rack is included, forming a corridor, and the camera is aligned to its centre.

In the remainder of this section, the procedure for calculating the metrics for each case is detailed. In all the cases, the restrictions mentioned above were used to compute the camera's location by optimising the SUR quantity.

For clarification, Figure 5 shows the geometrical definitions, variables and relations over the scenario's top and side view. The variables $d_{cam,i}$ are the camera's relative distance from the shelf, h_{shelf}, l_{shelf} the shelf's height and length, respectively, and h_{cont}, l_{cont} the container's height and length, respectively.

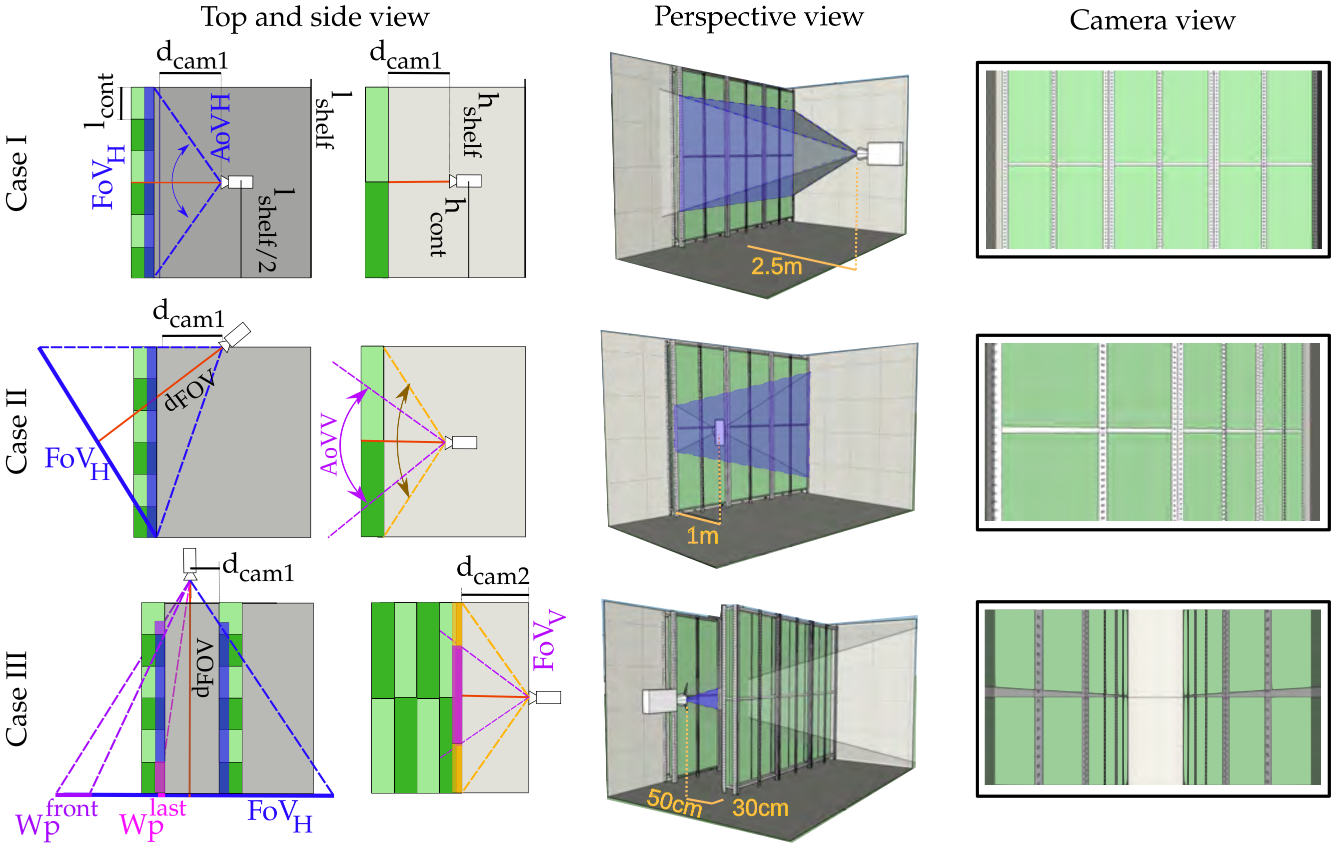


Figure 5. Three different cases proposed for the plant distribution study.

4.1.1. Case I

In this case, to attain the initial restrictions (Equations (33), (34) and (35)), the horizontal FoV, FoV_H, of the camera must completely cover the entire shelf's length (Equation (37)).

The other restriction is that the camera's vertical FoV, FoV_V, must be lower than or equal to the shelf's height (Equation (38)). Otherwise, the floor and the ceiling will be visible within the image, reducing the photobioreactors' vertical size.

$$FoV_H = 2 \cdot \tan\left(\frac{\mathrm{AoV}_H}{2}\right) \cdot d_{\mathrm{cam},1} \geq l_{\mathrm{shelf}} \tag{37}$$

$$FoV_V = 2 \cdot \tan\left(\frac{\mathrm{AoV}_V}{2}\right) \cdot d_{\mathrm{cam},1} \leq h_{\mathrm{shelf}} \tag{38}$$

To resolve the stated equations, AoV_V must be set to its maximum possible value. As design criteria, it was established as 70°. In Section 5, the reason behind not selecting a higher $AoV_V$is discussed.

Using Equations (37) and (38), the camera distance, d_{camera}, is obtained, and consequently the SUR. The selected worst-case viewing angle corresponds to the container located further to the right.

4.1.2. Case II

The translation of the camera to the side has the advantage that its distance is considerably reduced. However, as can be seen in Figure 5, the horizontal projection of the reactors shrinks relative to its distance to the camera, and the angle of view also reduces the optical power received by the image sensor. Therefore, Equation (34) becomes an important restriction. In this case, FoV_H is defined in Equation (39).

$$FoV_H = 2\tan\left(\frac{\mathrm{AoV}_H}{2}\right)\cdot d_{\mathrm{FoV}_H} \tag{39}$$

$$\text{where :}\quad d_{FoV_H} = \frac{d_{\mathrm{cam},1}\cdot\cos(\mathrm{AoV}_H/2)}{\cos(\mathrm{AoV}_H)}$$
$$AoV_H = \arctan\left(\frac{l_{\mathrm{shelf}}}{d_{\mathrm{cam},1}}\right)$$

In this case, utilizing the maximum $AoV_H = 70^\circ$ restriction, w_{roi} considerably exceeds the restriction imposed in (34); thus, this configuration is still viable.

We highlight that, in this particular case, the camera does not visualise the whole surface of the nearest containers, but merely a region of their base (as shown in Figure 5). In contrast, the furthest containers are fully scanned. Notwithstanding this uneven configuration, the image projection for each container is preserved (as shown in the example frame). Both the nearest and the furthest containers are projected over the same number of vertical pixels within the image. Therefore, all the containers share the same image area available for communications.

4.1.3. Case III

In this setup, part of the image belongs to the end wall, reducing the available area left for the photobioreactors. In this case, also the horizontal projection constrains the arrangement. If AoV_H is too wide, the last reactor's horizontal size will not reach the minimum imposed (Equation (34)) and would not be visible within the image. Otherwise, if AoV_H becomes too narrow, the first reactor, which is partially visible, could disappear from the image. The size of the horizontal projection for the first container and the last can be computed using Equations (40) and (41), respectively.

$$W_p^{last} = \frac{d_{\mathrm{cam},1}\cdot l_{\mathrm{cont}}}{l_{\mathrm{shelf}} + d_{cam,2} - l_{\mathrm{cont}}} \tag{40}$$

$$W_p^{first} = (l_{\mathrm{shelf}} + d_{cam,2}) \quad\cdot\quad \left[\tan\left(\frac{\mathrm{AoV}_H}{2}\right) - \frac{d_{\mathrm{cam},1}}{(l_{\mathrm{cont}} + d_{cam,2})}\right] \tag{41}$$

The location of the camera y_{cam} and $d$cam is obtained by reducing the horizontal size of both containers (first and last) within the image, attaining the imposed minimum size (Equation (40)).

4.2. *Experiments*

Two experiments were conducted at the BEA facilities to evaluate signal reception and deterioration due to the increase in biomass concentration. The first experiment evaluated the container surface's radiance in a controlled indoor environment, whilst in the second experiment, the system's performance in an outdoor scenario was assessed. The shared materials, resources and equipment are summarised in Table 1.

The following subsections describe each scenario separately to facilitate understanding the motivation, methods and data analysis for each experiment.

4.2.1. Indoor Experiment

In this experiment, the radiance emitted by specific regions from the container's surface is evaluated. The aim is to determine how the radiance profile changes as the biomass concentration increases for different camera viewing angles. At the same time, the SNR of the channel is measured, and the bit error rate (BER) is estimated. The selected species for this experiment was *BEA 0007B Arthrospira platensis* (*A. platensis*).

Regarding the evaluation methods, the experiment was based on acquiring nine images at different viewpoints with respect to the azimuthal angle ϑ, keeping the link range constant. The radiance for vertical angles was not evaluated because camera translations in the vertical axis were not considered in the theoretical analysis, and for symmetry reasons, a similar behaviour was expected. For each discrete angle, the camera was aligned in the direction to the centre of the container. Figure 6a shows the experimental setup. The biomass concentration was increased in steps of $20\,\mathrm{mg\,l^{-1}}$, starting from $0\,\mathrm{mg\,l^{-1}}$ and reaching $100\,\mathrm{mg\,l^{-1}}$.

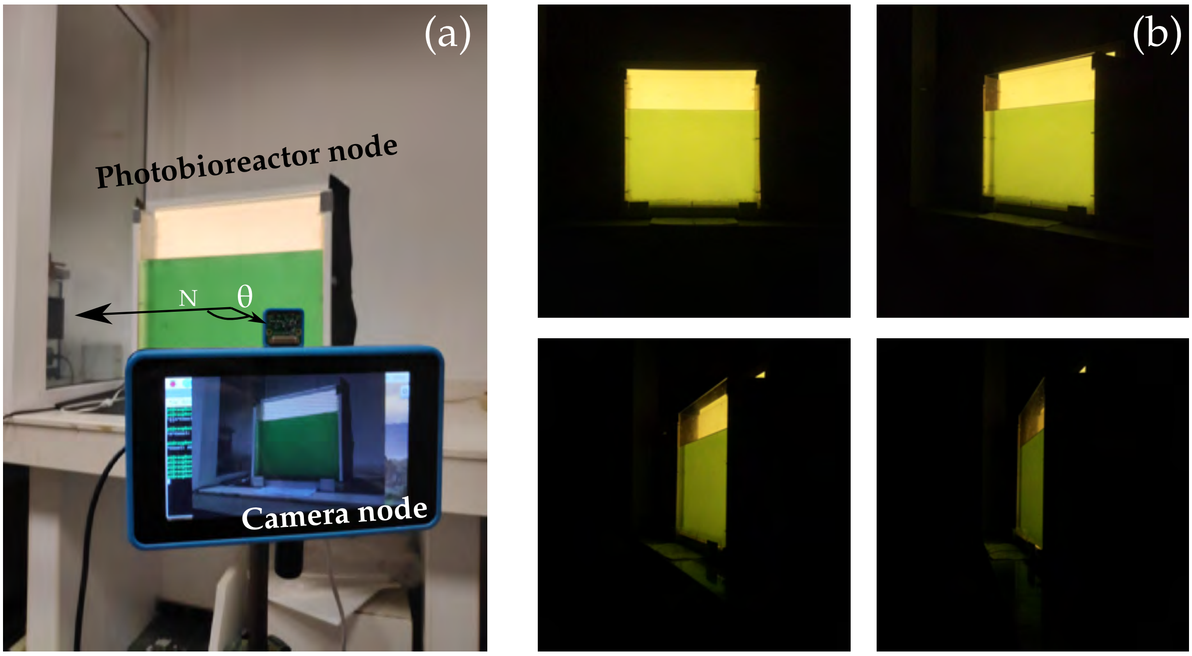


Figure 6. Pictures of the indoor experiment. (**a**) shows the experimental setup, and (**b**) depicts exemplary images captured at different angles for different biomass concentrations.

After capturing the image samples, they were processed offline to obtain the radiance and SNR, from two different regions, the centre and the boundaries of the container's surface, to analyse the skewness phenomenon in the outermost regions of the surface.

The radiance was measured indirectly by averaging the pixel values in a tiny window no higher than 50 × 50 pixels. For the area projected in this window and considering the distance and the size of the container, it can be assumed that the radiance pattern varies very smoothly within this region. The original horizontal and vertical window's length was established at the start of the experiment when $\vartheta = 0^\circ$. Afterwards, the horizontal size must decrease as a function of the cosine of the angle, ϑ. Consequently, only the radiance contribution of the same original area was evaluated for each angle. The vertical length would remain unaltered as there were no vertical translations. We highlight that because the camera points to the centre of the containers, the windowed area located at the boundaries would decrease not only as a function of the cosine of the view angle, but also as a function of the distance, which varies very slightly for more acute angles. However, the area reduction factor due to this distance increment is considerably lower than the factor due to the view angle, and it can be neglected.

In addition, the SNR was computed within the same window by measuring the mean, μ_{chan}, and the standard deviation, σ_{chan}, of each independent colour channel, Equation (42), taking into account that there was no other source of light during the experiment. Furthermore, the expected theoretical BER for an on-off keying (OOK) signal can be estimated from Equation (43) using the complementary error function $\mathrm{erfc}(\cdot)$ [3].

$$SNR = 20 \cdot log_{10}\left(\frac{\mu_{chan}}{\sigma_{chan}}\right) \tag{42}$$

$$BER = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{SNR}{2}}\right) \tag{43}$$

The key parameters of this experiment are summarised in Table 2. Figure 6b shows some image samples for different biomass concentrations. It can be anticipated by the images obtained that at the borders, the radiance decreases abruptly for internal view angles (in the case of low biomass concentration). This suggests that the initial assumption about the radiance's skewness phenomenon at the surface's borders is valid.

Table 2. Indoor experiment key parameters.

Parameter	***Value***
Horizontal view angle (ϑ)	0° to 80° in steps of 10 degrees
Distance	1.5 m
LEDs	Warm and cold white LEDs
Camera	PiCamera version 2
Microalgae	
Genus	Arthrospira
Species	*BEA 0007B Arthrospira platensis*
Biomass concentration ($\mathrm{mg\,l^{-1}}$)	0 to 100 in steps of 20

4.2.2. Outdoor Experiment

This experiment aimed to evaluate how the presence of external light sources interferes with the signal received by the camera, considering both the light that is reflected on the container's transmission surface and the light that enters through the sides of the container. In this experiment, the camera node faced just one photobioreactor, which repeatedly transmitted a beacon packet through different biomass concentrations (from low to high). The selected microalgae species was *BEA 1286B Rhodosorus marinus* (*R. marinus*) (brown algae) to evaluate the communication restrictions that imply the use of microorganisms with different absorption curves. The beacon signal is comprised five sequential pulses (green, red, blue, cold white, warm white) followed by a dark guard. The key parameters of this experiment are summarised in Table 3.

Table 3. Outdoor experiment key parameters.

Parameter	***Value***
Optical signal	Beacon
Chip duration T_{chip}	1/8400
Distance (m)	2
Microalgae	
Genus	Rhodosorus
Species	*BEA 1286B Rhodosorus marinus*
Biomass concentration (mg/L)	(estimated) 75, 195, 430
Camera	
Model	Mi 9T Pro
Image sensor	Sony IMX586
Image resolution (px)	4000 × 3000
Focal length (mm)	4.8 mm
Aperture value	F/1.7
ISO speed rating	450, 490, 670
Exposure time (µs) t_{exp}	100

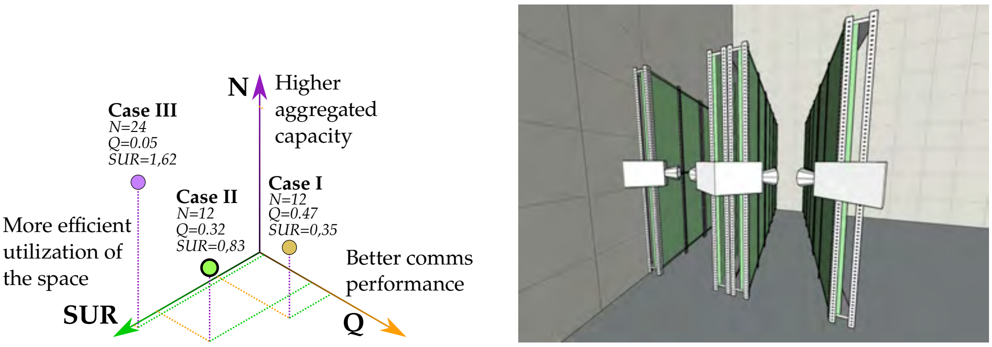
5. Results

5.1. Plant Distribution Study

Table 4 shows the metrics calculated for each case. For illustrative purposes, Figure 7 plots every case in the space defined by the proposed multidimensional metric F.

Table 4. Metric values obtained for the three arrangements.

Case	V_{room} (m^2)	V_{cont} (m^2)	h_{min}	$L(\vartheta,\varphi)$	N	SUR	Q
I	23.31	0.81	0.5	0.95	12	0.035	0.47
II	9.81	0.81	0.5	0.65	12	0.083	0.32
III	11.68	1.62	0.5	0.12	24	0.139	0.05



(**a**) (**b**)

Figure 7. Illustration of the results of the plant distribution study. (**a**) Space generated by the multidimensional metric F; (**b**) example of Case II's replication process.

The first configuration, Case I, has a relatively low SUR. Approximately only four percent of the available space is utilised for harvesting. Therefore, this proposal should be discarded, even though the conditions for establishing a communication link are practically ideal. This SUR could be increased if a greater AoV_V were selected. Increasing AoV_V causes the FoV_H to expand, and therefore, the camera can get closer to the shelf, reducing the distance considerably and minimising the required space. However, increasing AoV_V without limits would intensify the distortion effects introduced by the lens, bending the vertical lines in the image (barrel distortion), especially for fish-eye lenses. This distortion would not modify the shape of the symbol bands. The symbol bands will always be horizontal (or vertical) straight lines because they result from the RS acquisition method regardless of the lenses' optics properties. However, it does affect the containers' shape, reducing its vertical size. Furthermore, this non-linear distortion cannot be mitigated using image processing techniques; doing so would reflect the distortion on the signal bands.

In Case III, however, just the opposite happens. The SUR is almost four times higher than in Case I. However, the optical received power from the most distant containers is so low that it reduces the effective data rate and increases the complexity at reception. This reveals the importance of the camera viewing view. Those acute view angles will significantly affect the total received optical power. They will make it difficult not only to establish the link, but also to estimate the microalgae culture's biomass precisely.

In conclusion, Case II becomes the optimal choice for this application. It is the solution that achieves a balance in all parameters and has greater replicability. This replicability allows easily creating new plant distributions, such as the one shown in Figure 7. In that example, it is shown how just by adding a second camera on the original shelf and installing a new identical rack that holds the first camera, a corridor similar to Case III can be created. Hence, this configuration makes better use of space at the expense of adding a single low-cost camera. In that case, both the SUR and N double their original values, maintaining the communication performance.

However, despite this being the chosen solution, there is still a discussion to be made. The fact that the camera scans only a fraction of the first containers and performs the full scan of those that are further away (due to the image's viewing perspective) gives rise to new challenges. For example, the signal-to-interference-plus-noise ratio (SINR) is different for each container, especially under the presence of tiny air bubbles in motion within the reactor, which are utilised generally to aerate the organisms and induce a continuous movement. The projection of these small air bubbles increases in size for the nearest containers, generating more noticeable interference phenomena. Furthermore, in terms of biomass sensing, the scanning of just a fraction of the culture can provide reasonable estimations only if the biomass is correctly distributed within the recipient.

On the other hand, analysing the SUR values, it is observed that they are relatively small. The maximum aggregated container's volume is approximately 13% of the entire required space. However, it is important to notice that, in all vertical cultivation plants, it is necessary to reserve some room for the technicians to execute periodic control routines, handle the extraction of samples with ease and react quickly to warning alarms. Conventionally, it is recommended to reserve up to one meter of separation between racks. This separation coincides with the optimal camera distance in Case II.

5.2. Indoor Experiment

Figure 8 shows the intensity radiation pattern emitted from the two selected regions at the container's surface. The columns represent the area's location, from left to right: the centre and the border. The rows represent each independent image's channel, from top to bottom: red, green and blue. Each graph shows the radiation pattern normalised to the maximum radiant intensity for azimuth angles between five and 175 degrees. The radiation patterns for the different concentrations are grouped following a colour gradient from lighter to darker as the concentration increases. Finally, for reference purposes, the Lambertian radiation patterns with n = 1 and n = 2 are shown.

In the container's central region, the radiation pattern follows a Lambertian with n = 1 for red and green channels when no microorganisms are suspended in the water. In the blue channel, it deviates minimally for wider angles. It can be seen that as the biomass increases, the emission becomes more directive. Actually, in the case of the red and green channel, at the point of maximum concentration, it follows a perfect Lambertian pattern with n = 2. The reason behind this has a connection with *A. platensis*'s SPF and its absorption and scattering cross-section coefficients. *A. platensis* is a planktonic filamentous cyanobacterium. Its cylindrical morphology gives the cell a highly directive SPF [33] (depending on the orientation). The absorption cross-section measured experimentally is moderately high compared to other microalgae organisms.

Therefore, the absorption coefficient, measured experimentally in [25], contributes more to the light extinction than the scattering. This absorption causes the outgoing light at steep angles, which had travelled longer distances than the direct rays, to undergo a significant attenuation, which ultimately produces this directivity on the radiation pattern.

On the other hand, in the blue channel, the radiation intensity decreases faster with the increase in biomass than the red and the green channel. This is to be expected given that the absorption cross-coefficient curve as a function of the wavelength is not flat, but intensifies by almost 1.7 times in the 350–450 nm region of the spectrum [25].

Regarding the emission from the edges, the skewness effect mentioned in Section 3.1 is observed. This emission also becomes more and more directive as the biomass increases.

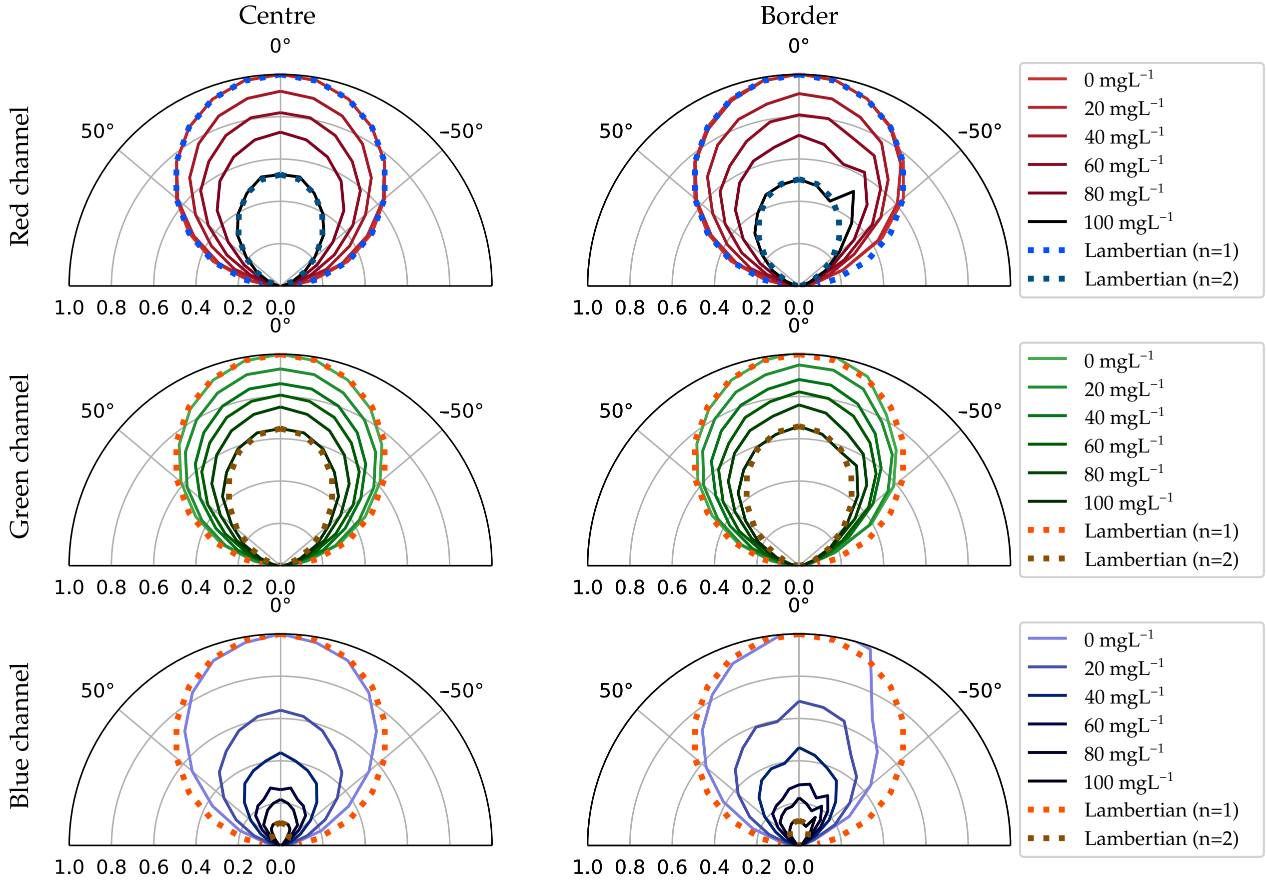


Figure 8. Red, green and blue radiation diagram of small areas located at the centre and the border of the container.

To better illustrate how the camera perceives the differences in the radiation between both regions (centre and border) at a certain angle and to understand the reason behind selecting these two regions for the analysis, Figure 9 shows some examples of the images captured during this experiment. In this figure, framed in a blue rectangle, the interface that abruptly separates the centre and the border region is highlighted (Figure 9a). Furthermore, this region's evolution with respect to the increase in concentration is shown in the boxes below.

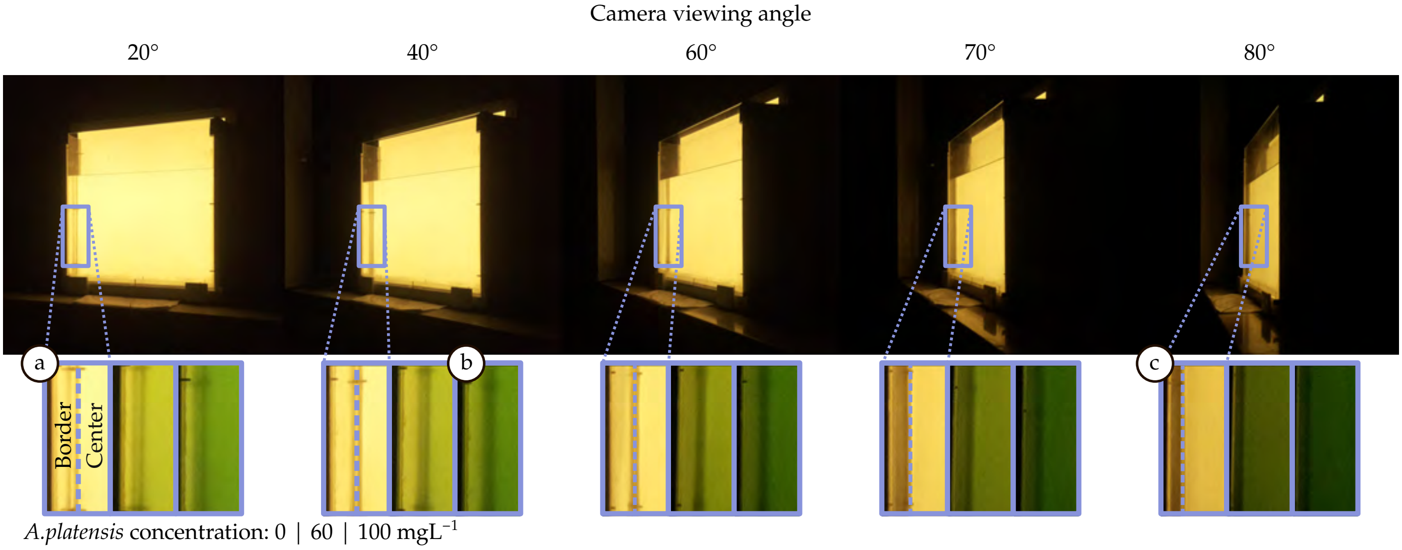


Figure 9. Illustrative example of the results obtained in the indoor experiment.

The location of this interface on the surface depends on the container's geometric construction and moves with the point of view. In thicker containers, this interface gets closer to the centre, and consequently, the border region increases significantly in size. On the other hand, as the container is viewed from a tighter angle, this interface moves towards the edges (Figure 9c). In this case, the central region extends smoothly towards the borders without reaching them. This interface's location should be taken into account in the design of the communications link because it delimits two areas with notable differences in light emission that will ultimately impact the SNR. In this setup, the power received from the edges is generally less than from the centre. Therefore, if the sampling occurs at the edges, the signal would be affected by a lower SNR. However, it can be seen that the relative intensity contribution for angles above 30 degrees is higher in the borders than the centre (Figure 8). For these angles, the radiation intensifies relatively. This effect is also observed in Figure 9b, where the edge region appears brighter than the centre. This outcome is related to the SPF of the *A. platensis* and the internal shape of the container.

Regarding the SNR, Figure 10 shows the SNR for different viewpoints either when the signal is received from the centre (dashed lines) or the side (dotted lines). The columns represent the biomass concentration, increasing from left to right. The rows represent the image channel, from top to bottom: red, green and blue. Each graph shows the SNR against the camera viewing angle (from zero to 85 degrees).

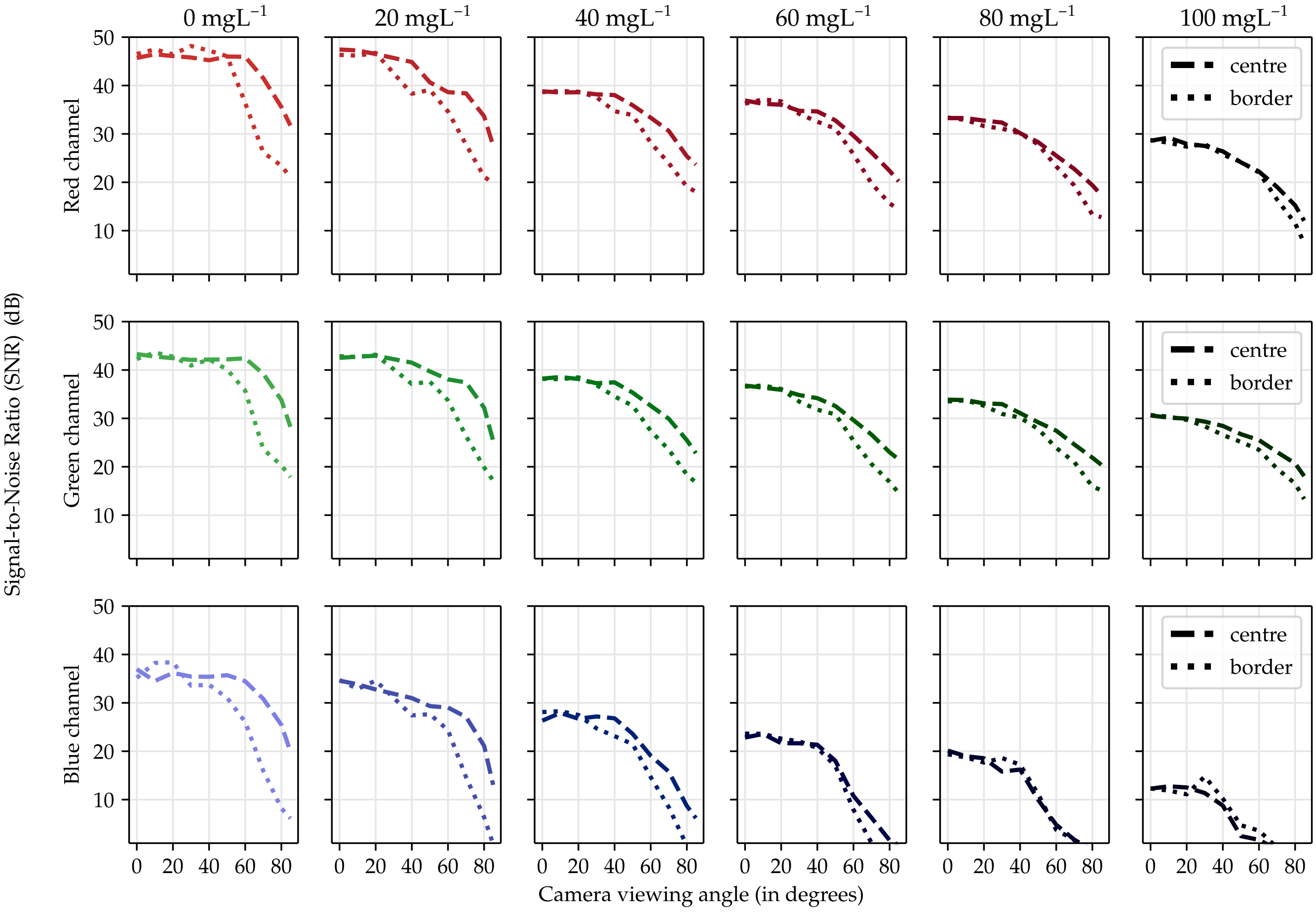


Figure 10. Red, green and blue SNR for small areas located at the centre and the border of the container.

The most evident result extracted from these graphs is that the SNR in the blue channel decreases much faster than in the other channels. Furthermore, the SNR starts to decay at 40 degrees at the edges, approximately 20 degrees earlier than in the red and green channels. This was discussed previously in terms of the light intensity emitted by the surface. Nevertheless, another result can be extracted. The differences between the central and outer SNRs are reduced progressively as the biomass concentration increases. *A. platensis* scattering slowly contributes to distributing the optical signal more evenly across the surface. Besides, the difference in power also gradually decreases, at least in the red and green channels. Therefore, slow attenuation combined with increasing scattering is beneficial, mainly because it reduces the gap between the two regions, increasing the effective signal reception area. Finally, regarding the sensing of the biomass concentration and culture's growth state, it is concluded that it slightly depends on the viewing angle. The relative differences in attenuation per angle for each channel as a function of concentration are non-linear. As the concentration increases, the radiant intensity diagram varies slightly in shape, relatively different for each channel. Therefore, to estimate the biomass correctly using only the pixel values inside a given area, it is necessary to consider the viewpoint. However, this radiation's behaviour provides valuable information that can be favourably exploited for more accurate crop parameters' estimation. The larger the variations, the better the estimation will be. In this sense, Case II, which proved to be the ideal candidate for a real deployment, has the added advantage that it analyses the strain from different angles. If all containers are interconnected through pipes and the strain is transferred from container to container periodically, its sensing will be more accurate.

5.3. Outdoor Experiment

The results are presented in Figure 11. This figure displays a snapshot taken for three concentration densities: low (a), medium (b) and high (c). The white rectangle encloses the detected beacon signal. The graphs shown on the picture's right side represent the red, green and blue pixel values from a one-pixel column located within this rectangle. Furthermore, Figure 11d represents the reference case where the signal is extracted without being affected by the microalgae channel. The signal recovered corresponds to the sequentially pulsed LEDs: white cold, white warm, dark guard (no pulse), green, red and blue.

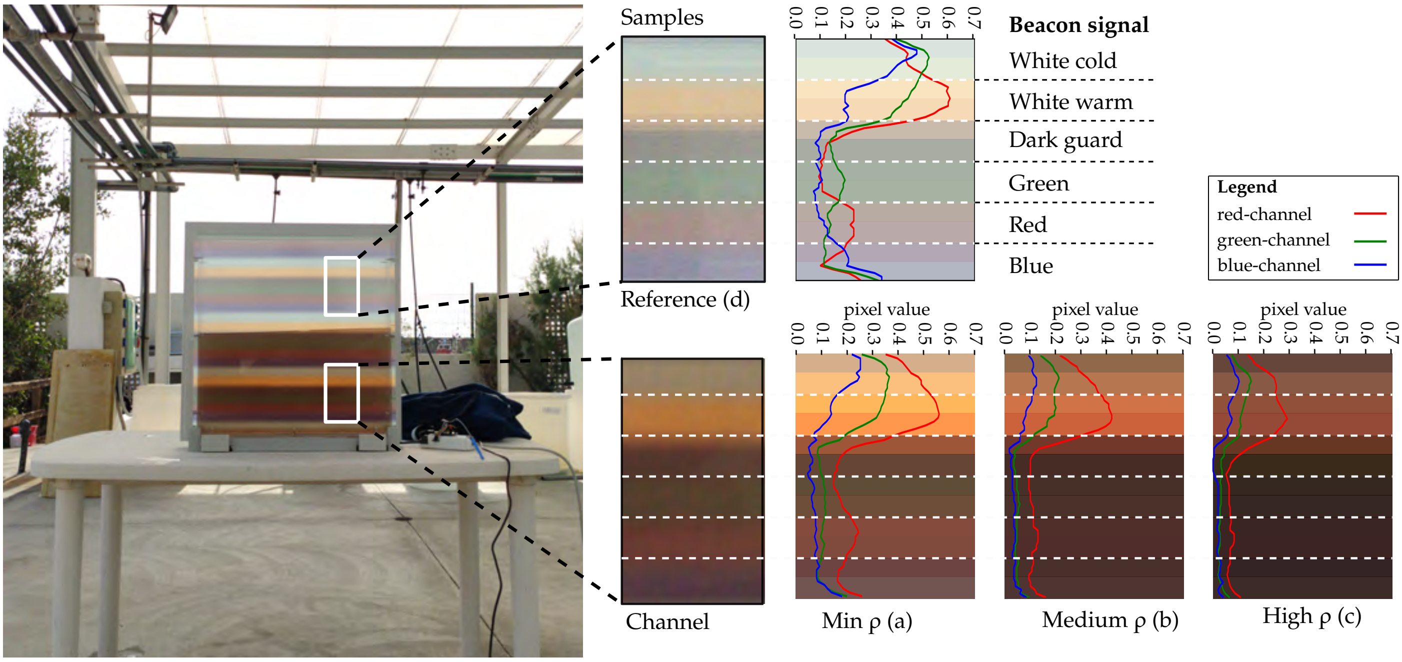


Figure 11. Experimental transmission of a well-known beacon signal.

It can be observed that this species produces significant attenuation in the blue and green portions of the spectrum (attaining the Bayer filter's spectrum response of the camera). Despite that the green, blue and red pulses cannot be distinguished in Figure 11c (due to low optical transmitting power), it is still possible to differentiate the white cold and white warm spectral signatures. Therefore, those LEDs are viable for communication purposes.

Finally, analysing the reference signal (Figure 11d), it is observed that the sunlight increases the light power received at the container's surface. It sets an offset value and consequently reduces the dynamic range available for communications. However, as the concentration increases, this offset value decreases because of the attenuation of the microalgae. This implies that the external light that enters through the container's sides is attenuated, just like the signal of interest. Therefore, in high biomass concentration conditions, increasing the photobioreactor's light emission power will expand the available dynamic range. Pixel values obtained from dark pulses will tend gradually to zero, while the signal power is increased on purpose.

6. Conclusions

In this work, OCC was proposed as a suitable communication technology for monitoring microalgae production plants based on artificial lighting. The theoretical channel model was introduced alongside the analysis of the parameters that significantly impact the achievable data rate, such as the nodes' geometrical configuration and the camera sampling frequency and exposure time.

This research highlighted the importance of optimising plant distribution in terms of link quality, channel capacity and efficient space exploitation. A multidimensional metric was designed for this purpose and tested in the conducted case study, where three different node arrangements were classified. This study's main result was that the configuration in which the camera observes the reactor shelf from the side is the preferred solution. On the one hand, it provides comparable link qualities between all containers without incurring detrimental losses of the signal strength due to the camera's viewing angle. On the other hand, it better exploits the scarce space available than a camera watching the containers frontally.

In addition, the experimental evaluation of the proposed flat-panel photobioreactor prototype was carried out in indoor and outdoor environments for two different microalgae species: *A. platensis* and *R. marinus* (green and brown algae, respectively). In the indoor experiment, the container surface's radiation pattern was measured for each image channel at different concentrations. This analysis reveals significant differences in the SNR between channels due to the algae absorption spectrum. Moreover, notable differences can be observed between the borders and the central part of the surface. These differences tend to reduce as the concentration of microalgae increases due to the phenomenon of scattering. Despite that the viewpoint must be taken into account when designing the system, it additionally provides valuable information for in situ sensing the algae's biomass and its growth state. In the outdoor experiment, the blue and green channels were highly attenuated, which compelled discarding these wavelengths for data transmission. Furthermore, in this scenario, the sunlight decreased the available dynamic range considerably for signal transmission, especially for low biomass concentration. Finally, examining the obtained results for the two microalgae species with different absorption profiles determined that the selected strain and its temporal evolution must be considered in the development of the optical link.

Future research should further develop and confirm these initial findings by extending the SNR's evaluation to the actual achievable BER for this deployment in both indoor and outdoor scenarios. It should examine more deeply how the presence of aeration bubbles might affect the signal reception. It could also focus on comparing and developing novel framing strategies for sensor data, reducing packet overhead, easing transmitter discovery and data synchronisation and equalisation. In terms of culture sensing, artificial intelligence for biomass estimation may constitute the object of future studies.

In conclusion, it was demonstrated that microalgae production plants are a potential use case for OCC. This technology provides simultaneous node monitoring capabilities using a cost-effective deployment, paving the way to develop smart farming strategies within the microalgae cultivation field.

Author Contributions: Conceptualization, C.J.-V. and V.G.; data curation, C.J.-V.; formal analysis, C.J.-V., V.G. and J.R.; funding acquisition, J.R.; Investigation, C.J.-V., V.M. and C.A.; methodology, C.J.-V. and V.G.; project administration, J.R.; resources, C.A. and J.R.; software, C.J.-V.; supervision, V.G. and J.R.; validation, C.J.-V.; visualization, C.J.-V.; writing—original draft, C.J.-V.; writing—review and editing, C.J.-V., V.G., V.M., C.A. and J.R. All authors read and agreed to the published version of the manuscript.

Funding: This project received funding from the Canary Island Regional Government (Project ATICCuA ProID2017010053). This project has received funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement No. 764461.

Acknowledgments: C.J.-V. thanks the technical support given by Carlos Guerra Yanez at IDeTIC—Universidad de Las Palmas de Gran Canaria.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

References

1. Acién Fernández, F.G.; Fernández Sevilla, J.M.; Molina Grima, E. Photobioreactors for the production of microalgae. *Rev. Environ. Sci. Biotechnol.* **2013**, *12*, 131–151. [CrossRef]
2. Nwoba, E.G.; Parlevliet, D.A.; Laird, D.W.; Alameh, K.; Moheimani, N.R. Light management technologies for increasing algal photobioreactor efficiency. *Algal Res.* **2019**, *39*, 101433. [CrossRef]
3. Ghassemlooy, Z.; Alves, L.N.; Zvanovec, S.; Khalighi, M.A. *Visible Light Communications: Theory and Applications*; CRC Press: Boca Raton, FL, USA, 2017.
4. Sarrafzadeh, M.H.; La, H.J.; Lee, J.Y.; Cho, D.H.; Shin, S.Y.; Kim, W.J.; Oh, H.M. Microalgae biomass quantification by digital image processing and RGB color analysis. *J. Appl. Phycol.* **2014**, *27*, 205–209. [CrossRef]
5. Murphy, T.E.; Macon, K.; Berberoglu, H. Multispectral image analysis for algal biomass quantification. *Biotechnol. Prog.* **2013**, *29*, 808–816. [CrossRef] [PubMed]
6. *IEEE Standard for Local and Metropolitan Area Networks–Part 15.7: Short-Range Optical Wireless Communications-Redline*; IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011)-Redline; IEEE: New York, NY, USA, 2019; pp. 1–670.
7. Le, N.T.; Hossain, M.A.; Jang, Y.M. A survey of design and implementation for optical camera communication. *Signal Process. Image Commun.* **2017**, *53*, 95–109. [CrossRef]
8. Saeed, N.; Guo, S.; Park, K.H.; Al-Naffouri, T.Y.; Alouini, M.S. Optical camera communications: Survey, use cases, challenges, and future trends. *Phys. Commun.* **2019**, *37*, 100900. [CrossRef]
9. Almadani, Y.; Plets, D.; Bastiaens, S.; Joseph, W.; Ijaz, M.; Ghassemlooy, Z.; Rajbhandari, S. Visible Light Communications for Industrial Applications—Challenges and Potentials. *Electronics* **2020**, *9*, 2157. [CrossRef]
10. Saha, N.; Ifthekhar, M.S.; Le, N.T.; Jang, Y.M. Survey on optical camera communications: Challenges and opportunities. *Iet Optoelectron.* **2015**, *9*, 172–183. [CrossRef]
11. Kuroda, T. *Essential Principles of Image Sensors*; CRC Press: Boca Raton, FL, USA, 2017.
12. Teli, S.R.; Zvanovec, S.; Perez-Jimenez, R.; Ghassemlooy, Z. Spatial frequency-based angular behavior of a short-range flicker-free MIMO–OCC link. *Appl. Opt.* **2020**, *59*, 10357–10368. [CrossRef]
13. Hassan, N.B.; Ghassemlooy, Z.; Zvanovec, S.; Biagi, M.; Vegni, A.M.; Zhang, M.; Luo, P. Non-Line-of-Sight MIMO Space-Time Division Multiplexing Visible Light Optical Camera Communications. *J. Light. Technol.* **2019**, *37*, 2409–2417. [CrossRef]
14. Teli, S.R.; Matus, V.; Zvanovec, S.; Perez-Jimenez, R.; Vitek, S.; Ghassemlooy, Z. Optical Camera Communications for IoT–Rolling-Shutter Based MIMO Scheme with Grouped LED Array Transmitter. *Sensors* **2020**, *20*, 3361. [CrossRef] [PubMed]
15. Carvalho, A.P.; Silva, S.O.; Baptista, J.M.; Malcata, F.X. Light requirements in microalgal photobioreactors: An overview of biophotonic aspects. *Appl. Microbiol. Biotechnol.* **2011**, *89*, 1275–1288. [CrossRef] [PubMed]
16. de Mooij, T.; de Vries, G.; Latsos, C.; Wijffels, R.H.; Janssen, M. Impact of light color on photobioreactor productivity. *Algal Res.* **2016**, *15*, 32–42. [CrossRef]
17. Abu-Ghosh, S.; Fixler, D.; Dubinsky, Z.; Iluz, D. Flashing light in microalgae biotechnology. *Bioresour. Technol.* **2016**, *203*, 357–363. [CrossRef] [PubMed]
18. Schulze, P.S.; Guerra, R.; Pereira, H.; Schüler, L.M.; Varela, J.C. Flashing LEDs for Microalgal Production. *Trends Biotechnol.* **2017**, *35*, 1088–1101. [CrossRef]
19. Aoyama, H.; Oshima, M. Line scan sampling for visible light communication: Theory and practice. *IEEE Int. Conf. Commun.* **2015**, *2015*, 5060–5065. [CrossRef]
20. Matus, V.; Guerra, V.; Zvanovec, S.; Rabadan, J.; Perez-Jimenez, R. Sandstorm effect on experimental optical camera communication. *Appl. Opt.* **2021**, *60*, 75–82. [CrossRef]
21. Matus, V.; Eso, E.; Teli, S.R.; Perez-Jimenez, R.; Zvanovec, S. Experimentally Derived Feasibility of Optical Camera Communications under Turbulence and Fog Conditions. *Sensors* **2020**, *20*, 757. [CrossRef] [PubMed]

22. Jurado-Verdu, C.; Guerra, V.; Matus, V.; Rabadan, J.; Perez-Jimenez, R.; Luis Gomez-Pinchetti, J.; Almeida, C. Application of Optical Camera Communication to Microalgae Production Plants. In Proceedings of the 2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Porto, Portugal, 20–22 July 2020; pp. 1–6.
23. Pilon, L.; Berberoĝlu, H.; Kandilian, R. Radiation transfer in photobiological carbon dioxide fixation and fuel production by microalgae. *J. Quant. Spectrosc. Radiat. Transf.* **2011**, *112*, 2639–2660. [CrossRef]
24. Svensen, Ø.; Frette, Ø.; Erga, S.R. Scattering properties of microalgae: The effect of cell size and cell wall. *Appl. Opt.* **2007**, *46*, 5762–5769. [CrossRef] [PubMed]
25. Ma, C.Y.; Zhao, J.M.; Liu, L.H. Experimental study of the temporal scaling characteristics of growth-dependent radiative properties of Spirulina platensis. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *217*, 453–458. [CrossRef]
26. Ma, C.Y.; Zhao, J.M.; Liu, L.H.; Zhang, L. Growth-dependent radiative properties of Chlorella vulgaris and its influence on prediction of light fluence rate in photobioreactor. *J. Appl. Phycol.* **2019**, *31*, 235–247. [CrossRef]
27. Shahjalal, M.; Hasan, M.K.; Jang, Y.M.; Nguyen, V.T.; Chowdhury, M.Z. Performance Analysis and Improvement of Optical Camera Communication. *Appl. Sci.* **2018**, *8*, 2527.
28. Lain, J.K.; Jhan, F.C.; Yang, Z.D. Non-Line-of-Sight Optical Camera Communication in a Heterogeneous Reflective Background. *IEEE Photonics J.* **2019**, *11*. [CrossRef]
29. Jurado-Verdu, C.; Matus, V.; Rabadan, J.; Guerra, V.; Perez-Jimenez, R. Correlation-based receiver for optical camera communications. *Opt. Express* **2019**, *27*, 19150. [CrossRef] [PubMed]
30. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
31. Goto, Y.; Takai, I.; Yamazato, T.; Okada, H.; Fujii, T.; Kawahito, S.; Arai, S.; Yendo, T.; Kamakura, K. A New Automotive VLC System Using Optical Communication Image Sensor. *IEEE Photonics J.* **2016**, *8*. [CrossRef]
32. Sony Corporation. *IMX219PQH5-C, Diagonal 4.60 mm (Type 1/4.0) 8 Mega-Pixel CMOS Image Sensor with Square Pixel for Color Cameras, Datasheet*; Sony Corporation: Tokyo, Japan, 2014.
33. Komine, M.; Takakura, T. Analysis of Light Intensity Distribution in a Suspension of Spirulina Platensis by the Radiative Transfer Equation. *Environ. Control. Biol.* **1998**, *36*, 41–51. [CrossRef]

5.1.1 Related contributions

Citation 1:
[24] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, R. Perez-Jimenez, J. Luis Gomez-Pinchetti, and C. Almeida, "Application of optical camera communication to microalgae production plants," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2020

This manuscript is a preliminary version of this work, which was presented at the International Symposium on Communication Systems, Networks, and Digital Signal Processing (CSNDSP) 2020.

Citation 2:
[25] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, "Barcolits: Barcodes using led tags and optical camera communications," in *2022 IEEE 18th International Conference on Factory Communication Systems (WFCS)*, pp. 1–8, 2022

During this work, it became clear that the successful implementation of this technology depended on providing a communications solution that (i) was extensible and adaptable to the wide variety of cameras available and (ii) whose deployment was not perceived as a disruptive approach to well-established industry processes and operations, but as compatible and interoperable with them. Therefore, following the GO3 objective, the search for a less disruptive communication solution led to present a novel system presented at the international conference on Factory Communication Systems (WFCS) 2022 (included in Annex B). This system, called Barcolits, uses traditional barcode coding schemes to encode optical signals emitted with OCC transmitters, ensuring interoperability with conventional barcode scanners. Furthermore, this coding strategy contributes to decoupling the camera (hardware) from the reception routines (software), following the scheme proposed in Section 5.3.

5.2 Artificial intelligence-assisted exposure equalization

Citation:

[21] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, and R. Perez-Jimenez, "Convolutional autoencoder for exposure effects equalization and noise mitigation in optical camera communication," *Opt. Express*, vol. 29, pp. 22973–22991, Jul 2021

This work addresses the general objective GO1 (along with the technical derived objectives), proposing an AI-assisted solution for simultaneous visualization and data detection using RS cameras.

As discussed in Chapter 3, cameras, as imaging devices, must adjust their exposure time according to the ambient light. In general, the exposure time must be set to relatively longer values, especially in indoor environments, compared to the IS's row sampling time (Chapter 3). However, in long exposures, the pixels accumulated the irradiance of several consecutive symbols, producing a detrimental ISI in the received signal. Consequently, the camera's exposure severely restricts the attainable reception bandwidth, behaving as a low-pass filter.

Therefore, the choice of exposure time sets a tradeoff. It should be: (i) long to improve scene display and the receiver's sensitivity and (ii) short to increase the link throughput.

This work proposes a novel AI-assisted equalization stage to mitigate the detrimental ISI associated with long exposures in cases where (i) the receiver's sensitivity needs to be increased to improve the SNR (or the visualization of the scene) or (ii) the camera exposure cannot be set. This proposal aims to validate hypothesis H2[5]. This hypothesis can be reformulated in more depth with the following statement: *"A neural network model can exploit the attenuated frequency components of a bandlimited received signal that falls beyond the cutoff frequency, to reconstruct the original signal as if it were acquired with a broader bandwidth".*

The validation of this hypothesis is the main contribution of this work and has important relevance in the field of communication systems. As follows from the reformulated hypothesis, the results of this work can be transferred to any communication system. It should be noted that, in this case, the bandwidth limitation derives from long exposures.

On the other hand, this work also aims to validate hypothesis H3[6]. This hypothesis arises from an unprecedented challenge: obtaining the network training samples. Those training samples consist of a set of two images: (i) overexposed images, used as the network's inputs, and (ii) ground-truth images, used as the network's outputs (Chapter 4). In other words, bandlimited signals (input) and ground-truth signals (output).

Obtaining these samples using a camera-based experimental testbed is exceptionally complex and time-consuming. First, the lack of synchronization between the camera and the transmitter makes it difficult to perfectly match the training inputs with their corresponding outputs, dramatically impacting the network performance. Some authors have partially solved this problem using matched filtering techniques that are less robust as the exposure time increases (Chapter 4). Second, the number of adequately characterized cameras to obtain worthwhile samples is small. Consequently,

---

[5] It is possible to enhance the received signal bandwidth using equalizers based on AI techniques

[6] It is possible to train an OCC-based AI system with a synthetic dataset.

the network will be biased by the cameras used during training and the experimental conditions, limiting the generalizability of the results. In short, using real-world samples for training requires a complex training stage before the system's deployment, making this solution impracticable and device/scenario dependent in most cases.

Alternatively, using synthetic samples for training enables the equalizer to operate with a vast range of cameras (i.e., any camera whose parameters fit into the training space). The procedure for the synthetic generation is presented in this manuscript, along with the detailed modeling of the RS acquisition mechanism in Chapter 4. Its most distinctive feature is that it uses exclusively temporal parameters of the OCC link: the transmitter symbol time and the camera's row sampling and exposure times. Remarkably, the implemented algorithm is independent of other signal parameters, such as the transmitted power. This is achievable thanks to the pre-standardization stage introduced before the NN. The z-score standardization of the input samples makes the synthetic (training) and real-world (validation) images comparable, at least from the network's point of view, and eliminates the necessity of estimating the received signal power. In simpler terms, the z-score standardization bounds the signal samples (synthetic and real) to similar values, regardless of the original signal power. The comparison between synthetic and real-world samples is also assessed in this manuscript.

Regarding the network architecture, the equalizer consists of a 2D CAE, explained in more detail in the manuscript. This two-dimensional model extract features from the signal samples (image rows), exploiting the redundancy present in nearby columns efficiently to reconstruct a denoised version in which the effects of the exposure time are highly mitigated.

The equalizer's performance is evaluated for different exposure and noise conditions. The exposure time is set relative to the symbol time. The ratio between both times is called the ESR (Chapter 3). In the extreme case, the ESR equals 7, which means that each pixel is exposed during the transmission of 7 consecutive symbols, accumulating the irradiance of all of them. In addition, the equalizer's capacity to operate with slightly longer and shorter exposure times than the selected one for training the network is assessed.

The results reveal that these equalizers can mitigate the ISI produced with ESR equals 7 while keeping the BER well below the FEC limit in moderate SNR conditions (12dB to 18dB). This represents an improvement of around 14 times the attainable receiver bandwidth compared to non-equalized receivers. In addition, these equalizers can withstand up to 11% deviation in the exposure time (compared to the reference trained exposure time) without compromising the performance. These results indirectly validate the proposed algorithm for the synthetic generation of samples and hypotheses H2 and H3. In conclusion, this work is of great relevance in communications systems. It demonstrates that using AI-assisted equalization stages, trained with synthetically generated samples, significantly increases the receiver bandwidth in real links, which cannot be achieved using traditional linear signal-processing techniques. This equalization becomes crucial in cases where the exposure time cannot be adjusted, or the camera's sensitivity must be increased, e.g., to operate as an imaging device.

Research Article
Vol. 29, No. 15 / 19 July 2021 / *Optics Express* 22973
Optics EXPRESS

Convolutional autoencoder for exposure effects equalization and noise mitigation in optical camera communication

Cristo Jurado-Verdu,* **Victor Guerra,** **Vicente Matus,** **Jose Rabadan,** **and Rafael Perez-Jimenez**

[1] *Institute for Technological Development and Innovation in Communications (IDeTIC), Universidad de Las Palmas de Gran Canaria (ULPGC), 35017 Las Palmas de Gran Canaria, Canary Islands, Spain*
[*] *cjurado@idetic.eu*

Abstract: In rolling shutter-based optical camera communication (OCC), the camera's exposure time limits the achievable reception bandwidth. In long-exposure settings, the image sensor pixels average the incident received power, producing inter-symbol interference (ISI), which is perceived in the images as a spatial mixture of the symbol bands. Hence, the shortest possible exposure configuration should be selected to alleviate ISI. However, in these conditions, the camera produces dark images with impracticable light conditions for human or machine-supervised applications. In this paper, a novel convolutional autoencoder-based equalizer is proposed to alleviate exposure-related ISI and noise. Furthermore, unlike other systems that use artificial neural networks for equalization and decoding, the training procedure is conducted offline using synthetic images for which no prior information about the deployment scenario is used. Hence the training can be performed for a wide range of cameras and signal-to-noise ratio (SNR) conditions, using a vast number of samples, improving the network fitting and the system decoding robustness. The results obtained in the experimental validation record the highest ISI mitigation potential for Manchester encoded on-off keying signals. The system can mitigate the ISI produced by exposure time windows that are up to seven times longer than the transmission symbol duration, with bit error rates (BER) lower than 10^{-5} under optimal SNR conditions. Consequently, the reception bandwidth improves up to 14 times compared to non-equalized systems. In addition, under harsh SNRs conditions, the system achieves BERs below the forward error correction limit for 1dB and 5 dB while operating with exposure times that are 2 and 4 times greater than the symbol time, respectively.

1. Introduction

Optical camera communication (OCC) is considered an extension of visible light communication (VLC), which replaces high-bandwidth photodiodes with image sensors (IS) to detect time and spatial variations in light intensity for enabling data communications. This technology emerges from the interest of reusing conventional cameras embedded in an increasing number of end-user devices (such as mobile phones, laptops, vehicle dashcams) to capture intensity-modulated (IM) light signals from a wide range of lighting sources, ultimately paving the way for VLC to break the market's entry barriers imposed by utilizing specific hardware. Furthermore, it has recently been included in the IEEE 802.15.7 [1] standard, which reveals the interest in this technology.

However, the handicap of IS-based receivers is their relatively low bandwidth inherently limited by the camera's frame rate [2–4], which makes them ideal for applications with low data rates, such as internet of things (IoT) applications, smart farming, indoor location, advertising, or vehicle-to-vehicle (V2V) communications among others. This restriction affects differently depending on the camera's acquisition mechanism. In global shutter (GS) cameras, the whole IS is exposed simultaneously. Therefore, the light signal is sampled once per acquired frame.

#433053
Journal © 2021
https://doi.org/10.1364/OE.433053
Received 3 Jun 2021; revised 25 Jun 2021; accepted 25 Jun 2021; published 6 Jul 2021

Consequently, the data rate is upper-bounded by the highest frame rate, restraining the transmission frequency, f_{tx}. Hence, the switching speed of light sources is constrained, producing, in some cases, a noticeable flicker that must be mitigated to prevent discomfort and health-related issues in human users [5]. This flickering can be alleviate by using under-sampled modulation (USM) schemes, such as under-sampled frequency shift on-off keying (UFSOOK) or under-sampled phase shift OOK (UPSOOK) [6] at the expense of decreasing the data rate. On the other hand, rolling shutter (RS) cameras scan the image progressively row by row of pixels. Each row of pixels is activated sequentially, sampling the light source at different instants during the frame capture. This acquisition mechanism produces different illuminated bands for the transmitted symbols within the image [7]. In this case, the sampling period (significantly lower than GS cameras) coincides with the time that elapses between the activation of two consecutive rows [8], which is also limited, albeit indirectly, by the camera's frame rate. Another parameter that further restricts the signal bandwidth is the exposure time (row exposure time for RS cameras), the duration in which a pixel remains exposed to light. During this exposition, the pixel integrates light, acting as a low pass filter, producing significant inter-symbol interference (ISI). This ISI is perceived as a spatial mixture of the symbol bands within the image, and its effects begin to be relevant after the exposure time is longer than half the symbol time.

Therefore, from a communications perspective, the exposure must be as short as possible to prevent ISI on high-speed signals at the cost of reducing the received signal strength. This trade-off must be addressed in any OCC system design [7,9]. However, improving the receiver bandwidth by reducing the exposure time will eventually result in impracticable light conditions for either human or machine-supervised applications. As it can be seen in Fig. 1 short exposure times would produce dark images, in which objects cannot be acceptably recognized (Mandrill picture). This energy impairment due to the reduction of the integration window can be mitigated by increasing the camera's analog gain, which can significantly improve the signal-to-noise ratio (SNR), as concluded in previous works [10–12].

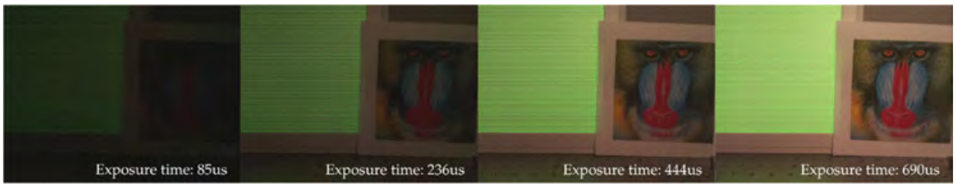


Fig. 1. Effects of increasing the exposure time in RS-cameras.

To alleviate this ISI effect, in [13] authors proposed a one-dimensional artificial neural network (ANN) equalizer with promising results. The neural network performance has been validated against Manchester encoded on-off Keying (OOK) [13], and constant power 4-PAM symbols [14]. In [13] this equalization allows to decode data with bit error rates (BER) below the forward error correction (FEC) limit of 3.8×10^{-3}, for exposure times up to 4 times greater than the symbol time. Translated in terms of bandwidth, it can recover signals whose bandwidth exceeds up to approximately nine times the low-pass filter's cutoff frequency that models the effect of the exposure time. However, the proposed multilayer perceptron (MLP) network performance was evaluated exclusively under optimal signal-to-noise ratio (SNR) levels. Furthermore, the network training was conducted online, with the receiver operating under the deployment scenario conditions. Up to the authors' knowledge, only these works try to mitigate the exposure-related ISI in RS cameras by using artificial neural networks.

However, artificial networks are acquiring a relevant role in solving other OCC technology challenges. In [15] the use of a Logistic Regression Machine Learning (LRML) algorithm is proposed for decoding signals transmitted with the backlight of advertising panels. In [16] a

1D-ANN architecture is proposed for the same purpose. Both works aim to decode signals affected exclusively by the interference produced by the frontal image content of the panel. Therefore, the camera is configured with short exposure times for the optimal reception of the transmitted symbols. In [17], a convolutional neural network (CNN), which combines convolutional layers with a fully connected classification network, is used for source detection and pattern recognition of LED-based sources in V2V communications. This network decodes spatially multiplexed streams under partial occlusion and/or harsh weather conditions. However, in this case, the RS acquisition mechanism is not exploited to increase the data rate, and hence the signal does not need prior equalization of the exposure-related effects. Instead, the transmitted symbol time is longer than the frame acquisition duration, so it can be considered that the system operates under GS conditions. In contrast, [18], uses a CNN for pattern detection and classification in V2V relying on the RS mechanism, recovering data from car rear taillights. Moreover, in [19], the use of CNNs is proposed for RS-symbol decoding. However, in previous works, the receiver operates on the premise that the exposure-related ISI is negligible since the cameras are configured with exposure times much shorter than the transmission symbol time.

This work proposes a novel two-dimensional convolutional autoencoder (CAE) for simultaneous exposure-related ISI equalization and noise mitigation, in which the training is conducted offline using synthetically generated images. These images are produced using exclusively time-related parameters from any chosen camera and transmitter: the configurable exposure time, the sampling period, and the transmission symbol time. These three parameters produce relevant training samples that, after prior standardization, enable the network to decode real captured images. This standardization makes the synthetic and real signals comparable from the point of view of the artificial network, regardless of the temporal average power received, as long as the camera's gamma transformation is precisely compensated. Hence, the training of the network becomes independent of the deployment scenario. These training samples can also be synthetically corrupted by a zero-mean additive white Gaussian noise (AWGN) to train the system for its operation under harsh SNR conditions. Therefore, the training can be carried using different synthetic noise levels and with a significant amount of samples stored in large databases, which considerably increases the robustness of the network.

Moreover, the use of CAE is justified since it has proven particularly useful in image denoising [20–22] presenting outcomes that outperform the capabilities of MLP architectures in this task [23], either in terms of efficiency and performance. This is, in part, because of the use of convolutional layers, but also because of the operation's nature of a CAE, which consists in extracting a latent representation or feature maps (generally of lower dimensionality) from the input (encoding part), and reconstructing it at the output using this representation (decoding part). In this encoding-decoding paradigm, noise-corrupted inputs might even be beneficial since they allow the network to deinterlace hidden useful features from the input. Furthermore, the use of two-dimensional inputs helps noise mitigation in RS acquired signals. Since the IS columns sample the light at the same sampling instants, the received signal is replicated across all the columns. This redundancy can be exploited to increase the SNR, thus facilitating the network task of noise mitigation.

The remainder of the paper is organized as follows. Section 2 introduces the RS theoretical modeling used for the synthetic image generation. Section 3 presents and describe the CAE-based proposed receiver system. Section 4 details, in the first place, the network training, including the synthetic image generation routines and the metrics used to assess the synthetic image similarity with real images and the network goodness of fit. In the second place, this section presents the procedures and metrics to evaluate the system's communication performance, including a detailed description of the experimental setup. Section 5 presents the results, to be compared in section 6 with those reported in recent literature. Finally, the conclusions of this work are summarized in section 7.

2. Rolling-shutter modelling

RS cameras scan the scene by progressively triggering each row of pixels rather than exposing the whole IS at once, as GS cameras do. Hence, light variations can be sampled up to H times per frame, where H is the IS's vertical resolution (height). This sampling produces different illuminated bands within the image that corresponds to the transmitted symbols. Figure 2(a) shows the temporal scheme of this acquisition mechanism. The IS starts exposing the first row of pixels to light during a configurable exposure time, t_{exp}. Then, the subsequent rows are activated sequentially every T_s seconds, which is the row sampling period. This acquisition can be modeled using a system with two separable stages: a filtering and a sampling stage, as shown in Fig. 2(c). The filtering part models the effects of the exposure time on the received signal for each pixel. The following stage manages the sampling instants for each pixel based on its position within the IS. Figures 2(b,d) illustrates how the incident light is converted into the final discrete values for each pixel (for a given column). Figure 2(b) represents the incident power reaching the j-th pixel over time, t, $P(t,j)$. On the one hand, Fig. 2(d) shows the pixel signals after the filtering stage, $v(t,j)$ (colored dashed lines). On the other hand, it depicts the one-dimensional discrete pixel values obtained after the sampling stage, $v[n]$ (black dotted vertical lines). It should be highlighted that, as it is shown in Fig. 2(b), the temporal evolution of the incident power follows the same shape for all the pixels. Nonetheless, the received signals are affected by different factor depending on several link parameters, such as the source's radiation pattern, the relative configuration between the transmitter and the camera, the camera lenses, the channel losses, and the scenario's reflections, among others. In this figure example (Fig. 2(b)), the average incident power gradually decreases from top to bottom. Moreover, the signal power is not enough to provide a suitable SNR for decoding in some cases. For this reason, in OCC, the light source projection within the image is generally considered as the ROI because it corresponds to the image area where the SNR is significantly higher. However, data can also be recovered from reflections as examined in [24–26]. Finally, it is important to mention that the following modeling is presented for a generic IS column. Therefore if the incident power for the pixel located at the i-th column and the j-th row is expressed with $\mathbf{P}(t,i,j)$, $P(t,j)$ satisfies the relation $P(t,j) = \mathbf{P}(t,m,j) = P^{(m)}(t,j)$, where $m \in [0,W)$ (the selected m column), and W is the IS's horizontal resolution (width).

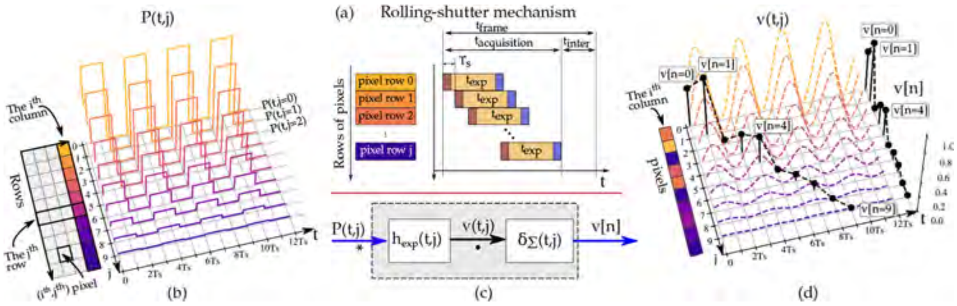


Fig. 2. RS acquisition mechanism. **(a)** RS Temporal scheme. **(b)** Normalized optical power $P(t,j)$ reaching each pixel of the i-th IS column. **(c)** RS system modeling (impulse response, h_{exp} and sampling function, δ_{Σ}). **(d)** Filtered curves, $v(t,j)$, and the discrete signal, $v[n]$ after sampling.

2.1. *Filtering stage*

The output value for the j-th pixel in the i-th column, $v(t,j)$, depends on the accumulated charge on the photodiode during the time it is exposed to light, the exposure time, t_{exp} (direct integration) [4]. This time extends from reset, in which the pixel's photodiode is biased with reverse voltage, until its readout. During readout, in passive pixel sensors (PSS), the charge is transferred to a floating diffusion amplifier (FDA) (shared for all the IS columns), where the charge is converted to voltage with a conversion gain of G_{conv}. Finally, this voltage is amplified at the general output amplifier and quantized by the analog-to-digital converter (ADC). In active pixel sensors (APS), the charge-to-voltage conversion occurs at the pixel level, and the voltage is transferred to the output amplifier using source follower amplifiers. In short, the output voltage of the pixel at the ADC input is given by Eq. (1).

$$v(t,j) = \frac{G}{C_{\mathrm{f}}} \int_{t}^{\mathrm{t+t_{exp}}} P(t,j) \cdot \Re(\mathrm{j}) dt \tag{1}$$

where $P(t,j)$ is the incident optical power, $\Re(j)$ and C_{f}, the equivalent photodiode's responsivity and capacity, respectively. The latter is approximately equal to the FDA's capacitor, and G the output amplifier's gain. This windowed integration of the input signal over the exposure time, can be modeled with a finite impulse response (FIR) low pass filter, h_{exp} given by Eq. (2), with its corresponding transfer function (Eq. (3)) [27].

$$v(t,j) = P(t,j) \circledast h_{\mathrm{exp}}(t,j) \quad \text{where:} \quad h_{\mathrm{exp}}(t,j) = h(t) = \frac{G}{C_{\mathrm{f}}} \cdot \Big(u(t + t_{\mathrm{exp}}) - u(t)\Big) \tag{2}$$

$$H(w) = \mathcal{F}\{h(t)\} = t_{\mathrm{exp}} \frac{G}{C_{\mathrm{f}}} \frac{sin(w \cdot t_{\mathrm{exp}}/2)}{w \cdot t_{\mathrm{exp}}/2} e^{jwt_{\mathrm{exp}}} \tag{3}$$

where $u(t)$ is the unit step function. From Eq. (3) it follows that the filter DC gain is proportional to the exposure time. Regarding the available reception bandwidth, to compute the cutoff frequency, it is necessary to rely on numerical methods such as Newton-Raphson's algorithm. However, to get an approximate idea of how the reception bandwidth is related to the exposure time, the first null frequency can be examined, which is inversely proportional to the exposure time. Therefore, a trade-off between the gain and the available bandwidth must be considered for the configuration of the camera's exposure settings. Light signals captured with shorter exposure times are affected by lower ISI, but also the received power decreases, as shown in Fig. 1. In those cases, it is still possible to improve the received signal quality by increasing the analog gain G [10,11].

2.2. Sampling stage

The family of curves obtained after the filtering stage, $v(t,j)$, shown in Fig. 2(d), is ideally sampled using a two-dimensional Dirac delta train function, $\delta_{\sum}$ (Eq. (4)), generating a one-dimensional discrete signal, $v[n]$.

$$\begin{aligned} & v[n] = v(t,j) \cdot \delta_{\sum}(t,j) \qquad \text{where:} \\ & \delta_{\sum}(t,j) = \sum_{n=0}^{\infty} \delta\Big(t - \Big\lfloor \frac{n}{H} \Big\rfloor \cdot t_{\mathrm{Frame}} - \mathrm{mod}(n,H) \cdot T_{\mathrm{s}}, \mathrm{mod}(n,H)\Big) \end{aligned} \tag{4}$$

where $\lfloor \cdot \rfloor$ is the floor function, $\mathrm{mod}(a,b)$, the modulo operation that returns the remainder of the division a/b. The floor division ($\left\lfloor \frac{n}{H} \right\rfloor$) returns the number of generated frames from the start. The modulo division ($\mathrm{mod}(n,H)$) returns the pixel index (j-th) that contributes to the n-th sample of the discrete signal $v[n]$. This equation can be further simplified in Eq. (5) under the assumption that t_{inter} is zero and the scanning operation is continuous. In other words, there are no periods in which the sensor becomes blind to transmission [7].

$$\delta_{\sum}(t,j) = \sum_{n=0}^{\infty} \delta\Big(t - n \cdot T_{\mathrm{s}}, \mathrm{mod}(n, H)\Big) \tag{5}$$

Furthermore, the Eq. (6) introduced in [27] can be derived from Eq. (5) under the condition that the IS pixels are affected by the same signal power. Therefore, the discrete signal, $v_{\mathrm{eq}}[n]$ (Eq. (6)) can be interpreted as the signal that would be acquired from a single equivalent pixel.

$$v_{\mathrm{eq}}[n] = v(t) \cdot \sum_{i=0}^{\infty} \delta(t - n \cdot T_{\mathrm{s}}) \tag{6}$$

The equations Eq. (4) and Eq. (5) indicate that each sample, $v[n]$, depends on the signal for the j-th pixel activated at the sampling instant and, consequently, the pixel position within the IS. Therefore, the sampling function relates the evolution of the signal over time with different image locations. In other words, this function express mathematically the space-time duality of OCC systems.

Finally, following the ideal sampling theory, the number of pixels (samples) per transmitted symbol (using the nomenclature introduced in [14]), N_{pps}, can be computed knowing the symbol time, t_{sym} and the sampling period, T_{s} ($N\mathrm{pps} = t_{\mathrm{sym}}/T_{\mathrm{s}}$).

3. Communications scheme

The proposed system architecture, and the functional blocks, are shown in Fig. 3. Regarding the transmitter, it emits non-return to zero (NRZ) Manchester encoded pulses to avoid flickering. Pseudo-random data sequences are grouped into packets with a header consisting of five consecutive ones and a zero-bit trailer. A redundant bit is inserted every three bits to prevent a header sequence from appearing within the payload. This stuffed bit is set to one if the preceding bit is zero and zero otherwise. This coding strategy eliminates the use of forbidden codes for synchronization, reducing the system's complexity and easing error detection at reception. The symbol time, t_{sym} is selected according to the camera's row sampling period T_{s}, to generate the desired N_{pps} ($N_{pps} = 5$). The transmitting source consists of a 20x20cm RGB LED flat panel that uniformly distributes the light across its surface. The operating link distance will depend exclusively on the lamp's size in the image and not on the optical emitted power (as long as the projection of the lamp occupies more than one pixel). As detailed in [28] the use of image-forming optics compensates the power losses due to spherical propagation with the projected size of the optical source on the IS. The receiving side consists of a RS-camera attached to a computing unit that performs the following routines for data acquisition as shown in Fig. 3. First, it selects M columns from the image's central region where the source is expected to be located (M equals 16). Next, it performs an equalization procedure to adjust the pixel values along the vertical dimension. In this work, no prior equalization is conducted. Then, the ROI is segmented into s overlapping windows according to the CAE's input dimensions (256x16 pixels) (with s equals 6). These image segments are standardized using the z-score function, which subtracts the image's mean μ to each sample x_i and divides the difference by the image's standard deviation, σ. This standardization is essential since it allows the images captured with the camera to be comparable with the synthetic training images, as will be discussed in section 4.1. The outputs generated by the CAE (with the exact dimensions of the input) for each segment are merged using a linear combination at the edges where overlap occurs. This linear merging helps mitigate the slight edge imperfections that appear near the top and the bottom of the output images. The size of this overlap depends on the number of selected segments. Increasing the number of segments will lightly improve the system's performance, but it will increase the computational load. Experimentally, it has been concluded that the imperfections affect a small area with 10 pixels height under the worst conditions (long exposure settings). Hence splitting a 1080-pixel image into six segments, generating 38-pixel overlaps for 256-pixel windows, is a reasonably conservative solution. Finally, the packet synchronization within the reconstructed image is conducted using the Pearson correlation with a header searching template.

The fundamental element of the proposed system is the CAE that performs both the equalization and denoising of the ROI. An autoencoder is a neural network that attempts to reconstruct the original input using a lower-dimensional latent representation [29]. It consists of a trained encoding network (encoder) that extracts relevant features from the input whilst its counterpart (decoder) is tuned to reconstruct the original input from this representation through the minimization of a loss function and a back-propagation algorithm for updating the weights of the architecture. This process is mathematically described in Eq. (7).

$$\bar{\mathbf{x}} = \mathcal{D}\left(\mathcal{E}\left(\mathbf{x}\right)\right) \tag{7}$$

where $\mathbf{x}$ is the input signal, which can be multi-dimensional, $\bar{\mathbf{x}}$ is the autoencoded version of $\mathbf{x}$, $\mathcal{E}(\cdot)$ is the encoding operation, and finally $\mathcal{D}(\cdot)$ is the decoding procedure. In this work, the loss training function $L(\mathbf{x}, \bar{\mathbf{x}})$ is the L_2-norm (mean squared error) with a regularization term to prevent over-fitting (Eq. (8)).

$$L(\mathbf{x}, \bar{\mathbf{x}}) = E\left[(\mathbf{x} - \bar{\mathbf{x}})^2\right] + \lambda R\left(\mathcal{E}, \mathcal{D}\right) \tag{8}$$

where $E\left[\cdot\right]$ denotes expected value, λ is the regularization coefficient, and $R(\cdot)$ is the regularization function, which in this work corresponds to a combination of the $L1$ and the $L2$ weights regularization penalties. On the other hand, a denoising autoencoder (DAE) is a specific type of AE that exploits the presence of noise in inputs to de-interlace useful properties, eventually mitigating the noise corruption in the output. In this case, it minimizes Eq. (9).

$$L\left(\mathbf{x}, \mathcal{D}\left(\mathcal{E}\left(\tilde{\mathbf{x}}\right)\right)\right) \tag{9}$$

where $\tilde{\mathbf{x}}$ is a copy of $\mathbf{x}$ that has been corrupted, in this case, by an a zero-mean additive white Gaussian noise (AWGN). Finally, a CAE uses convolutional layers (*Conv*) and transposed convolutional layers (*TConv*) [23] to encode and decode the input, respectively.

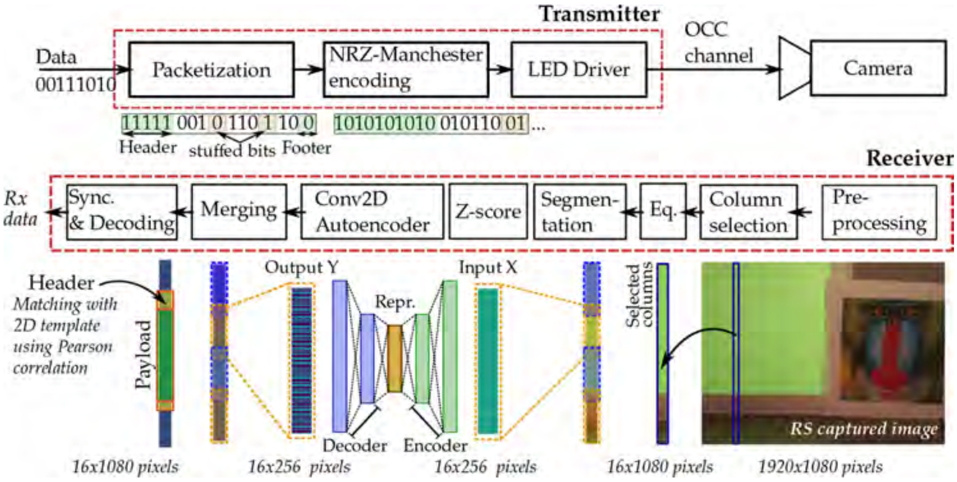


Fig. 3. Proposed system based on 2D convolutional autoencoder (CAE).

The *Conv* layers utilize a set of convolutional filters, which generates feature maps, F (one per filter), or activation maps by applying the discrete convolution operation. Considering the case in which the input consists of 2-dimensional gray-scale images, I, the discrete convolution uses two-dimensional kernels, K. The convolution result, $Z[i,j]$, at the i,j position for each filter is computed using Eq. (10).

$$Z[i,j] = (I \circledast K)[i,j] = \sum_{\mathrm{m}=0}^{k_{\mathrm{w}}-1} \sum_{\mathrm{n}=0}^{k_{\mathrm{h}}-1} I[m,n] \cdot K[i-m, j-n] \tag{10}$$

However, when working with RGB images, the input to the *Conv* layer consists of a three-dimensional tensor, where two dimensions are used for pixel position (width and height) and the last dimension for the three RGB color channels (depth). In this case, the convolution kernel is three-dimensional. Hence, generalizing for a number of D channels, the kernel's depth, k_{d} will match the number of channels of the input tensor, and the convolution result for each filter at the i,j position is computed using Eq. (11).

$$Z[i,j] = (I \circledast K)[i,j] = \sum_{\mathrm{m}=0}^{k_{\mathrm{w}}-1} \sum_{\mathrm{n}=0}^{k_{\mathrm{h}}-1} \sum_{\mathrm{l}=0}^{k_{\mathrm{d}}-1} I[m,n,l] \cdot K[i-m, j-n, l] \tag{11}$$

It should be highlighted that the filter translation over the image happens exclusively on the vertical and horizontal dimensions, summing up all the weighted contributions for all the channels to generate a two-dimensional tensor. The number of trainable weights per kernel will depend on its vertical and horizontal size and the input tensor channels. The result of the convolution, $Z[i,j]$, is then biased ($B[i,j]$) and transformed using a non-linear activation function, ψ, generating the corresponding features map, $F[i,j]$ (Eq. (12)).

$$F[i,j] = \psi(Z[i,j] + B[i,j]) \tag{12}$$

The nonlinear activation functions used in this work are the Sigmoid and the Rectified Linear Unit (ReLU) functions that work optimally in this type of architecture as demonstrated experimentally in [30].

The total trainable parameters of the l-th layer is the sum of the kernel's weights and biases considering all the filters. The latter coincides with the number of this layer's outputs, which can be computed knowing the horizontal and vertical dimensions of the output matrix, $O[i,j,l]$ using the Eq. (13).

$$dim(O[i,j,l]) = \left(\left\lfloor \frac{n_{\mathrm{H}} + 2p_{\mathrm{H}} - k_{\mathrm{H}}}{s_{\mathrm{H}}} \right\rfloor + 1, \left\lfloor \frac{n_{\mathrm{W}} + 2p_{\mathrm{W}} - k_{\mathrm{W}}}{s_{\mathrm{W}}} \right\rfloor + 1, D \right) \tag{13}$$

where n_{H}, n_{W} are the vertical and horizontal lengths of the input, p is the number of padding values added at boundaries (to control the output size), k_{H}, k_{W} are the vertical and horizontal lengths of the filter's kernel, and s the stride, the step translation of the kernel when traversing the input, and D, the number of filters.

In this CAE architecture, *Conv* layers are usually followed by a pooling layer, which replaces the layers' outputs in specific locations with a statistical summary of the outputs at the vicinity. In this model, max-pooling layers (*MaxPool*) are used, which return the maximum output of a rectangular group of outputs. This ultimately contributes to increasing the non-linearity of the output (in addition to the nonlinear activation functions) and reduces the total number of network parameters.

Conv layers have proven especially effective for extracting useful features from images, and they are widely used in object detection and classification as well as image segmentation and denoising. This convolution operation can help improve the efficiency of deep learning systems.

Furthermore, it allows reducing the number of network parameters by making better use of the spatial similarities in the vicinity of an input sample (sparse connectivity). In addition, in this type of network, the same filer kernel's weights would be applied across all the inputs (parameter sharing), tying the weights for different samples. This is contrary to what would happen in a dense network, in which each neuron assigns a specific weight for each input, and consequently, a separate set of parameters for every location is generated. Instead, in *Conv* layers, just a single set of parameters is learned (those concerning the filters). In this way, trained kernels would search for shared activation patterns across the image. *Conv* networks are thus dramatically more efficient than dense networks, reducing the total trainable parameters significantly.

On the other hand, the *TConv* layer reverses the spatial transformation produced by a *Conv* layer. Even though it is also (wrongfully) known as a deconvolutional layer, it does not perform the deconvolution operation. Instead, it carries out a regular convolution on an upsampled version of the input tensor to obtain an output tensor with the dimensions of the expected input of its reverse *Conv* layer. In this CAE architecture, these layers are usually followed by two-dimensional upsampling layers (*UpSampling*) that expand the input tensor by repeating samples at each position. This ensures that the output tensor has the exact dimensions of the input.

4. Methodology

This section starts by describing the CAE training procedures, including the generation of training synthetic image datasets. Next, the algorithm utilized for optimizing the network's hyperparameters and the selected search space is detailed. Finally, the communications performance evaluation procedure and metrics are introduced alongside the details of the experimental setup. This evaluation aims to demonstrate the ability of the system as a whole to equalize and decode overexposed signals with robustness in low to moderate SNR conditions, allowing simultaneous data acquisition and image visualization. Furthermore, that the network can, once it has been trained for a specific exposure time, adapt to slightly longer or shorter exposures. Finally, it aims to validate the use of synthetic (scenario-independent) images for network training.

4.1. Network training

The supervised training of the CAE is performed offline only once, using synthetically generated images. The four parameters considered for the generation are the exposure time, the sampling period, the symbol time, and the SNR. The procedure for generating a synthetic image, shown in Fig. 4, is described below.

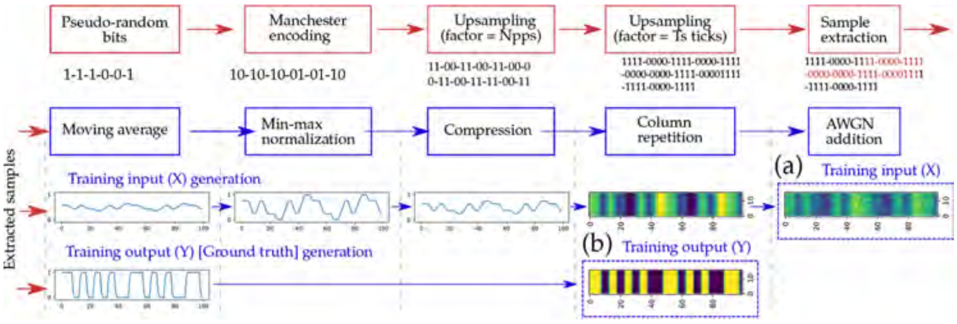


Fig. 4. Synthetic image generation.(a) Training input (X). (b) Training output (Y).

First, pseudo-random, one-dimensional bit sequences are encoded using the NRZ-Manchester line code. Then, they are upsampled by two factors: the number of row pixel samples (N_{pps}) expected per symbol and the number of ticks per sampling period T_s. The number of ticks depends on the selected resolution time. In this work, the selected time equals the clock period, which is approximately 100 nanoseconds. Hence since the sampling period is $18.9 \mu s$, its corresponding number of ticks is $T_s^{\text{ticks}} = 189$. Then, a section of the signal is extracted following a random starting offset to simulate non-perfect synchronization between the transmitter and the camera. This offset varies uniformly between zero (perfect synchronization) and one symbol time. Next, a moving average window is applied to model the effect of the camera's exposure time. The length of this window would depend on the requested exposure time. Particularly is equal to the number of ticks of the exposure time (for $t_{\text{exp}} = 444\mu s$, $t_{\text{exp}}^{\text{ticks}} = 4440$). The output is then normalized using the min-max normalization, resulting in a signal with values between zero and one. Then, it is compressed with a constant factor (0.5) to prevent clipping effects after adding the noise. The obtained one-dimensional signal is repeated along the horizontal dimension to generate a two-dimensional image. The resulting image has the dimensions of the CAE's input layer (256 rows and 16 columns). Finally, a zero-mean additive white Gaussian noise (AWGN) is added.

The training dataset collects sets of two synthetic images for a given random binary sequence: the input (X) and output (Y) images. The input image is made using the selected training exposure time. Figure 4(a) shows the procedure for generating the input image. The output image, that represents the ground truth, is generated similarly but selecting the shortest possible exposure time according to the time resolution (this time must be at least shorter than half of the symbol time). In this procedure, the min-max normalization, the compression, and the noise addition routines are discarded. Figure 4(b) shows the generation of the output image. The datasets generated in this work contain 35500 sets per exposure time (71000 images). From these datasets, 10% of the images are reserved for validation, while the remaining 90% are used for training.

Regarding the network training, the standardization of the input images is important. The z-score standardization applied to both the training and the real images makes them comparable from the point of view of the CAE. This eliminates the necessity to consider the expected average received power (and some camera parameters such as the analog and the digital gains) for the generation of the synthetic images. Therefore, the training is independent of the deployment scenario. However, it is mandatory to perform a prior compensation of the spatial power differences and the camera's gamma transformation. The similarity between synthetic and real images is measured using the Pearson's correlation coefficient. The training goodness-of-fit is quantified with the mean square error cost obtained for the training and the validation datasets.

Finally, regarding the network model, two different topologies are used in this work. The first topology has two stages composed of one *Conv* and one pool layer for the encoder part, and one *UpSampling* and *TConv* layer for the decoder part. The second topology adds another stage (with a total of three stages). For both topologies, efficient optimization of the network hyperparameters is conducted, following the hyperband algorithm detailed in [31] using the training exposure time of 444 μs. The considered parameters that constitute the search space are summarized in Table 1. The best architecture for each topology is used later for the system evaluation.

Table 1. Hyperparameter's search space

Hyperparameter	Value
Number of stages (*Conv*/*MaxPool* or *UpSampling*/*TConv*)	2, 3
Number of *Conv* (or *TConv*) filters	8, 16, 24, 32, 40, 48, 56, 64
Conv (or *TConv*) kernel's width and/or height	2, 3, 4, 5
MaxPool (or *UpSampling*) kernel's width and/or height	1, 2, 4
Learning rate	0.01, 0.001

4.2. Performance evaluation

Regarding communications, the CAE's performance is evaluated using real images captured with a conventional camera. The evaluation under different SNRs is performed by adjusting the light transmitted power through the control of the voltage source. Images taken under these conditions are used to estimate the SNR. It should be highlighted that since the received signal is affected by the camera's exposure time, it is not feasible to estimate the SNR by analyzing the image mean and variance. Hence, the following procedure is used. First, the pixel rows are averaged across all the image columns (1920 columns). This averaging increases the SNR by a factor of $N = 1920$ [14] (assuming that images are corrupted with AWGN). Next, the obtained averaged signal, $\bar{s}$, affected by a significantly low noise power, is subtracted from the signal at the desired decoding column, s, resulting in a noise signal, n. Finally, the signal power, S, is estimated using the maximum value of the autocorrelation of s (the same procedure is used for estimating the noise power N).

The selected metric to evaluate the communications' performance is the BER. In addition, to quantify the degree to which the signal is affected by exposure-related ISI, a new metric is introduced, the exposure-to-symbol ratio (ESR), the ratio between the cameras' exposure time, and the symbol time. For example, an ESR of 7 indicates that the exposure time exceeds seven times the symbol duration.

Figure 5 depicts the experimental setup used to capture the real images. It consists of an RGB flat panel pointing towards an RS-Camera separated by a distance of 50 cm. At this distance, the transmitter occupies approximately 3/4 of the image's vertical size. The transmitter signal is generated using an arbitrary signal generator and a power supply to control the voltage level of the light source.

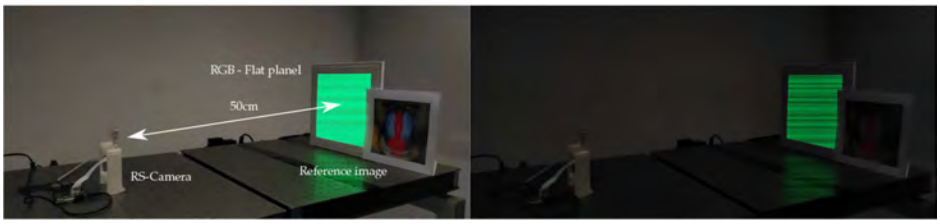


Fig. 5. Experimental setup

Table 2 summarizes the key parameters of the experiment setup.

Table 2. Experiment's key parameters

Parameter	Value
Camera	
Hardware	PiCamera version 2 (Sony IMX586) [32]
Aperture lens | Focal length (equivalent)	f/2 | 3 mm
Image resolution	1920x1080 pixels (Video mode - 3)
Clock time, t_{clk}	10 MHz
Sampling period, T_s	18.904 μs (Measured)
Exposure times, t_{exp}	from 85 μs to 1500 μs in steps of 19 μs
Transmitter	
Color channel used	Green channel
Voltage	from 25V to 36V in steps of 0,5V
Symbol time, t_{sym} (N_{pps})	94, 5 μs (5 pixels per symbol)
Packet's header, payload and trailer lengths	5, 42, 1 bits
Random seed	31415

5. Results

This section presents the results obtained for the training and the generation of synthetic images and the communications performance.

5.1. Network training

Figure 6 shows synthetic and real images for different exposure times to provide a visual comparison between them. Despite being placed horizontally, each segment corresponds to a vertical rectangle extracted from the image. The 24 examples are arranged into four groups based on the selected exposure time (161, 312, 444, or 520 μs). In each group, the reference template corresponds to the ground truth, the expected theoretical signal if the exposure time were infinitely short, and the incident power for all pixels, the same. The filtered template is generated for each exposure time using the reference template and normalized with the min-max function. This template is used to quantify the degree of similarity between the synthetic and real images. The following two examples correspond to the synthetic and real images without preprocessing as captured by the camera. As it can be seen, in those images, it is hard to distinguish light variations. This occurs because as the exposure time increases the dynamic range for the pixel values decays abruptly, reducing the ratio between the largest and smallest values. The last two images correspond to the standardized version (z-score) of synthetic and real images used as inputs for the network. As it was aforementioned in section 4, under the right conditions, both standardized images are comparable. Differences in the averaged received light in real images will minimally affect the standardization. Nevertheless, the camera's gamma transformation must be precisely compensated.

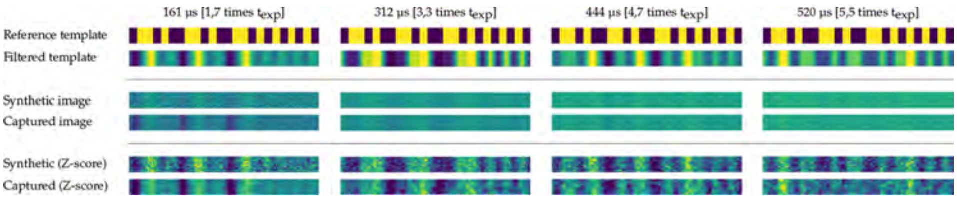
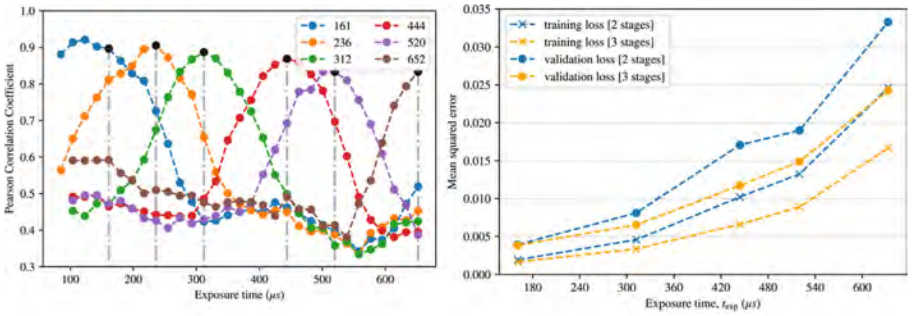


Fig. 6. Visual comparison of synthetic generated images versus real captured images

The degree of similarity between the synthetic and real samples is quantified by the maximum Pearson's correlation coefficient between the filtered template and the real images. Figure 7(a) presents the results obtained after matching the synthetic templates (shown in the legend) with images taken with a wide range of exposures. It should be mentioned that since unavoidable noise in the real images, the Pearson correlation coefficient does not reach its maximum value (1.0). However, it exceeds 0.85 in all cases in which both the synthetic and real exposures times are the same. As shown in this graph, as the exposure time slightly increases or decreases, the correlation coefficient rapidly decays to values around 0.4. The apparent symmetry of these curves reveals that non-similarities in the vicinity have a similar impact on the correlation. This could imply that the CAE could face approximately similar challenges when equalizing longer and shorter exposure times than the selected for the training.



(a) (b)

Fig. 7. (a) Pearson correlation coefficient between the filtered templates and the real images. (b) Training and validation losses for different training exposure times.

Table 3 summarizes the best model's parameters for both topologies as described in section 4. It details the number and type of layers (with their corresponding activation functions), the number of filters and kernel sizes, and, finally, the shape of the outputs and the total trainable parameters for each layer. In all cases, the learning rate that performed best was 0.001.

Figure 7(b) shows the training and validation losses for both topologies for different training exposure times. The use of a three-stage topology reduces losses for higher training exposures by an improvement factor of approximately 25%.

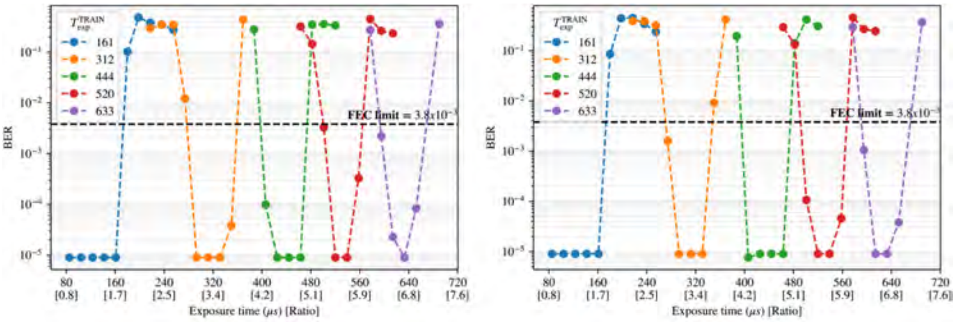
5.2. Performance evaluation

Figure 8 presents the BER results obtained after training the CAE with different exposure times of 161, 312, 444, 520, 633 μs. The x-axis shows the exposure time set by the camera (in microseconds) when taking images. In addition, the ESR (section 4.2), the ratio between the exposure time and the symbol time (94, 5 μs) is shown between brackets below its corresponding exposure. From these results, it can be extracted that the CAE can easily adapt to exposure times that are up to 10% higher or lower than the selected one for training, maintaining the BER below 10^{-5}. However, after this margin, the BER increases steeply. An exceptional case occurs when the training time is almost twice the symbol time ($T_{\text{exp}}^{\text{TRAIN}} = 161 \approx 1.8 \cdot 94.5\mu s$), as it can be seen in Fig. 8. In this particular case, the system can decode signals for all the exposures that are lower than the selected for training, maintaining the BER below 10^{-5}. The graph shown in Fig. 7(a) helps to explain this result. In this graph, it is observed that the correlation of the training template (161 μs) with images exposed with shorter exposure times remains relatively high. Hence, the neural network can extract from the training images a set of features that differ minimally from those of the images affected by lower ISI. Finally, the comparison between both

Table 3. Optimized model's summaries for the training dataset of 444 μs.

	Two stages topology				Three stages topology			
Layer (activ.)	Kernel	Filters	Out.shape	Params.	Kernel	Filters	Out.shape	Params.
Encoder								
Conv (ReLU)	(4,5)	16	(256,16,16)	336	(5,5)	24	(256,16,24)	624
Maxpool	(2,2)	16	(128,8,16)	0	(2,2)	16	(128,8,24)	0
Conv (ReLU)	(5,3)	56	(128,8,56)	13496	(5,4)	48	(128,8,48)	23088
Maxpool	(2,2)	56	(64,4,56)	0	(1,2)	48	(128,4,48)	0
Conv (ReLU)	*Not applicable*				(4,3)	32	(128,4,32)	18464
Maxpool	*Not applicable*				(2,2)	56	(64,2,32)	0
Decoder								
TConv (ReLU)	(5,3)	56	(64,4,56)	47096	(5,3)	32	(64,2,32)	12320
Upsamling	(2,2)	56	(128,8,56)	0	(2,2)	32	(128,4,32)	0
TConv (ReLU)	(4,5)	16	(128,8,16)	17936	(4,5)	48	(128,4,48)	30768
Upsamling	(2,2)	56	(256,16,16)	0	(2,2)	48	(256,8,48)	0
TConv (ReLU)	*Not applicable*				(4,5)	24	(128,8,24)	28824
Upsamling	*Not applicable*				(2,2)	56	(256,16,24)	0
Conv (Sigm.)	(5,3)	1	(256,16,1)	321	(5,3)	1	(256,16,1)	601
	Total trainable parameters			79185	**Total trainable parameters**			114689

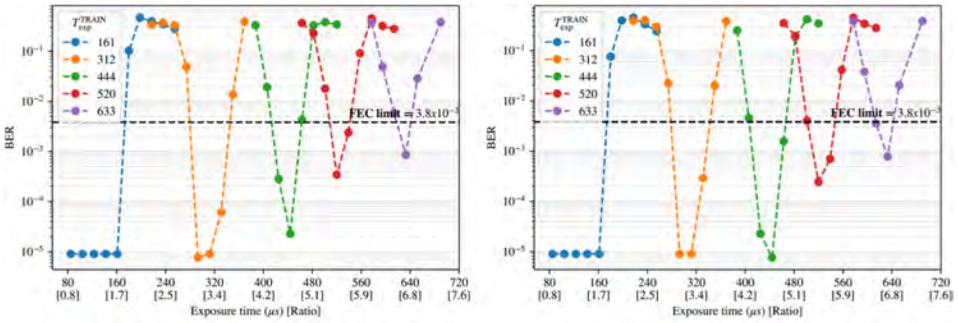
topologies agrees with the results shown in Fig. 7(b). The BER decreases less steeply for the three-stage topology, especially for the high exposures.



(a) Two stages (b) Three stages

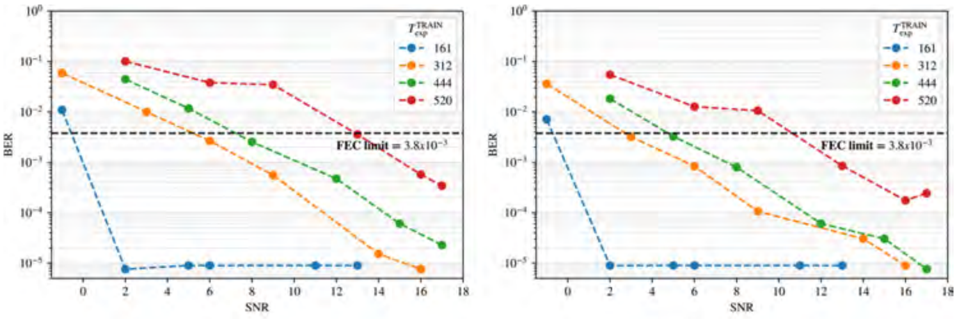
Fig. 8. BER results for training exposures, $T_{\text{exp}}^{\text{TRAIN}}$, under high SNR conditions.

Figure 9 shows the results obtained under harsh SNR conditions (between 12 and 18 dB). In this case, using a 3-stage topology is justified for cases where the ESR is greater than 3. Under these conditions, it is possible to decode signals with BER below the FEC limit for ESR values up to 7. Fig. 10 shows the behavior of the system under different SNR conditions. As it can be seen, the BER decreases approximately linearly with the SNR (in dB), with a comparable slope for all cases (except for the blue curve). This slope is approximately 200 dB^{-1}. In addition, it indicates that as the training is conducted longer exposure times, the CAE cannot correctly minimize the error at the output, regardless of the noise level. The exceptional case occurs when the ESR is approximately equal to 2. In this case, the network has succeeded at deinterlacing the hidden features of the training signal from the added noise, enabling the system to achieve BERs below 10^{-5} for SNR greater than 3 dB. Figure 11 shows a visual representation of the received signals for exposures of 444 and 520 μs, under the maximum (yellow curve) and minimum SNR (purple curve) conditions, with different average incident power. The estimated SNR is shown within the legend for each curve. To help the accurate visualization of the estimated SNR, a non-scaled version of the filtered template (black dashed curve) is included within the graph. It corresponds to the expected signal to be received in the absence of noise. Clarify that the average power level of the signal changes since it is being adjusted through the voltage source of the light to vary the SNR.



(a) Two stages (b) Three stages

Fig. 9. BER results for training exposures, $T_{\text{exp}}^{\text{TRAIN}}$, under moderate SNR (12 to 18 dB).



(a) Two stages (b) Three stages

Fig. 10. BER results for different training exposures, $T_{\text{exp}}^{\text{TRAIN}}$ under SNR conditions.

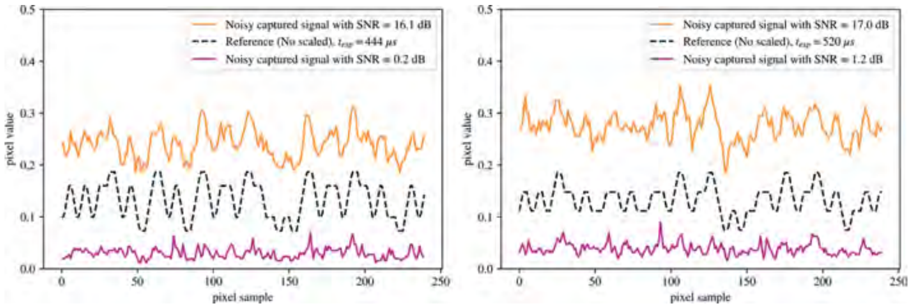


Fig. 11. Real signal examples obtained under different SNR conditions.

6. Discussion

In this section, the experimental results shown before are discussed. The proposed system is evaluated in the context of the available literature up to the authors' knowledge. Finally, a method to estimate the theoretical achievable data rate of the state-of-the-art systems reported is detailed and used to compare their performance based on the parameters of their experimental setups.

Table 4 shows the results comparison against the state-of-the-art. In this table, the fairest comparison results are the ESR (section 4.2) and the bandwidth improvement ratio. The latter refers to the ratio between the signal bandwidth and the cut-off frequency restriction imposed by the camera's exposure time. The proposed system improves the results in both cases.

Table 4. Comparison with the state-of-the-art.

Architecture	Output	Training	Max. ESR [a]	Max. BW ratio[a]	Min. SNR (ESR ≈ 4)[a]	Trainable params.	Theor. data rate[a]
1D-MLP[b]	1	Real	≈ 5 times	≈ 9 times	> 30 dB	100200[c]	2.584 kbps
2D-CAE	256x16	Synthetic	≈ 7 times	≈ 14 times	≥ 5 dB	79185	3.072 kbps

[a] With BER below the FEC limit.
[b] Younous *et al.* [13]
[c] Considering 500 input, 200 hidden and 1 output neurons (non-biased). [14]

Regarding the SNR conditions, this system can decode data with BER below the FEC, with ESR around 4, under SNR as low as 5 dB. However, it can also decode signals with ESR equals 2, under SNR as low as 2 dB, with BER below 10^{-5}. Regarding the number of training parameters, a notable reduction is also observed, which improves network efficiency by up to 20%.

The last column detail the achievable theoretical data rate. To conduct a fair comparison, both systems must meet the following requirements: the number of frames per second, fps is 30 fps, and the vertical resolution of the sensor, H, 1024 pixels. The assumptions considered for estimating the theoretical achievable data rate are detailed below. First, the whole vertical resolution of the IS is utilized for data recovery. Second, perfect synchronization between the transmitter and the receiver is assumed, without blind times at the reception, which means that the camera is operating continuously without stopping between frames. Finally, preamble or postamble bits are not considered. Under these assumptions, a fair comparison of the achievable data rate is computed using Eq. (14) (notice the factor of two since the signal is encoded using Manchester). In the case of 1D-MLDP, the data rate has been computed based on the sampling frequency of the Thorlabs CMOS camera model DCC1645C (detailed in the datasheet [33]) (13,315 kHz), and the maximum transmission frequency reported in [13] (2,240 kHz), resulting in $N_{\text{pps}} = 5.94 \approx 6$.

Under these conditions, the proposed CAE reaches a binary rate that is up to 500 bps higher than the achieved by the 1DMLP. Furthermore, the camera used in this work has a higher vertical resolution (1080 instead of 1024) and can be configured with 60 fps, so the maximum achievable rate is 6,480 kbps.

$$Rb = \frac{H}{2N_{\text{pps}}} \cdot \text{fps} = H \cdot \frac{T_{\text{s}}}{2t_{\text{sym}}} \cdot \text{fps} = H \cdot \frac{f_{\text{tx}}}{2f_{\text{s}}} \cdot \text{fps} \tag{14}$$

However, if previous requirements are not fulfilled, to ensure the successful packet detection within a frame, it is necessary to send the packet repeatedly, at least during the acquisition of two consecutive frames, and to restrict the packet length (in pixels) to at most half of the vertical size of the ROI [34]. Thus, the theoretical data rate must be divided by a factor of 4. In addition, preambles and postambles must be included in experiments, and the ROI is generally a fraction of the IS vertical resolution and might be considered in the computation of the data rate.

Finally, a series of comments should be added concerning the output size. Since the MLP network has a single output neuron, it is necessary to sweep the image from top to bottom pixel by pixel, which is computationally expensive. Besides, it is necessary to reserve m samples before (or after) with respect to the output sample, depending on whether the equalization weights the inputs backward or forwards (or a mixture of both). This would affect the equalization at the edges, introducing some artificial errors that further reduces the exploitable region for recovering data. In this work, the CAE's output has the same dimensions as the input, allowing equalizing the image by segments (not by pixels). Furthermore, the undesirable effects on the edges are effectively mitigated by the neural network that conveniently evaluates both backward and forward pixel values.

7. Conclusions

This work demonstrated that using a 2D CAE for exposure-related ISI mitigation outperforms the state-of-the-art one-dimensional MLP networks. The proposed system could decode signals with BER below 10^{-5} for exposure times that exceed seven times the transmission symbol time (ESR$\geq$ 7) in optimal SNR conditions. This implies a bandwidth improvement at the reception of approximately 14 times compared to a non-equalized receiver. Moreover, the system can decode signals under low SNR conditions. For example, BER values lower than the FEC limit can be obtained for SNR greater than 9, 5, or 1 dB for ESR of 7, 4, or 2, respectively. In addition, the network is capable of decoding signals with exposure times 10% longer or shorter than the selected one for training, which favors its flexibility to operate with different camera devices. On the other hand, the proposed architecture reduces up to 20% the total trainable network parameters. Finally, the procedure for generating synthetic RS training images was validated. The network's input standardization allows the synthetic procedure to rely exclusively on time-related parameters independent of the deployment scenario: the camera's exposure time, the sampling period, and the transmitted symbol time. These images can also be distorted with artificial noise. Therefore the network training can be conducted offline, for an endless number of cameras and SNR conditions, using vast training databases, ultimately improving the network fitting and the overall decoding robustness.

Funding. This work was funded by the Spanish Research Administration (MINECO project: OSCAR, ref.:TEC 2017-84065-C3-1-R). This project has received funding from the European Union's Horizon 2020 Marie Skłodowska-Curie grant agreement No. 764461.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. I. of Electrical and E. Engineers, "IEEE standard for local and metropolitan area networks–part 15.7: Short-range optical wireless communications," IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011) pp. 1–407 (2019).
2. W. Liu and Z. Xu, "Some practical constraints and solutions for optical camera communication," Phil. Trans. R. Soc. A **378**(2169), 20190191 (2020).
3. N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, "Optical camera communications: Survey, use cases, challenges, and future trends," Phys. Commun. **37**, 100900 (2019).
4. T. Kuroda, *Essential Principles of Image Sensors* (CRC Press, 2017).
5. S. Rajagopal, R. D. Roberts, and S.-K. Lim, "IEEE 802.15.7 visible light communication: modulation schemes and dimming support," IEEE Commun. Mag. **50**(3), 72–82 (2012).
6. P. Luo, M. Zhang, Z. Ghassemlooy, S. Zvanovec, S. Feng, and P. Zhang, "Undersampled-based modulation schemes for optical camera communications," IEEE Commun. Mag. **56**(2), 204–212 (2018).
7. H. Aoyama and M. Oshima, "Line scan sampling for visible light communication: Theory and practice," in *2015 IEEE International Conference on Communications (ICC)*, (IEEE, 2015), pp. 5060–5065.
8. N. Saha, M. S. Ifthekhar, N. T. Le, and Y. M. Jang, "Survey on optical camera communications: challenges and opportunities," IET Optoelectron. **9**(5), 172–183 (2015).
9. N. T. Le, M. Hossain, and Y. M. Jang, "A survey of design and implementation for optical camera communication," Signal Proc.: Image Commun. **53**, 95–109 (2017).
10. V. Matus, V. Guerra, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, "Sandstorm effect on experimental optical camera communication," Appl. Opt. **60**(1), 75–82 (2021).
11. V. Matus, E. Eso, S. R. Teli, R. Perez-Jimenez, and S. Zvanovec, "Experimentally derived feasibility of optical camera communications under turbulence and fog conditions," Sensors **20**(3), 757 (2020).
12. V. Matus, V. Guerra, C. Jurado-Verdu, S. R. Teli, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, "Experimental evaluation of an analog gain optimization algorithm in optical camera communications," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, (IEEE, 2020), pp. 1–5.
13. O. I. Younus, N. B. Hassan, Z. Ghassemlooy, P. A. Haigh, S. Zvanovec, L. N. Alves, and H. Le Minh, "Data rate enhancement in optical camera communications using an artificial neural network equaliser," IEEE Access **8**, 42656–42665 (2020).
14. O. I. Younus, N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, L. N. Alves, and H. Le-Minh, "The utilization of artificial neural network equalizer in optical camera communications," Sensors **21**(8), 2826 (2021).
15. Y.-C. Chuang, C.-W. Chow, Y. Liu, C.-H. Yeh, X.-L. Liao, K.-H. Lin, and Y.-Y. Chen, "Using logistic regression classification for mitigating high noise-ratio advisement light-panel in rolling-shutter based visible light communications," Opt. Express **27**(21), 29924–29929 (2019).
16. K.-L. Hsu, C.-W. Chow, Y. Liu, Y.-C. Wu, C.-Y. Hong, X.-L. Liao, K.-H. Lin, and Y.-Y. Chen, "Rolling-shutter-effect camera-based visible light communication using RGB channel separation and an artificial neural network," Opt. Express **28**(26), 39956–39962 (2020).
17. A. Islam, M. T. Hossan, and Y. M. Jang, "Convolutional neural networkscheme–based optical camera communication system for intelligent internet of vehicles," Int. J. Distributed Sens. Networks **14**(4), 155014771877015 (2018).
18. M. R. Soares, N. Chaudhary, E. Eso, O. I. Younus, L. N. Alves, and Z. Ghassemlooy, "Optical camera communications with convolutional neural network for vehicle-tovehicle links," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, (IEEE, 2020), pp. 1–6.
19. L. Liu, R. Deng, and L.-K. Chen, "47-kbit/s RGB-LED-based optical camera communication based on 2D-CNN and XOR-based data loss compensation," Opt. Express **27**(23), 33840–33846 (2019).
20. V. Jain and S. Seung, "Natural image denoising with convolutional networks," Advances in neural information processing systems **21**, 769–776 (2008).
21. L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* pp. 241–246 (2016).
22. M. Nishio, C. Nagashima, S. Hirabayashi, A. Ohnishi, K. Sasaki, T. Sagawa, M. Hamada, and T. Yamashita, "Convolutional auto-encoder for image denoising of ultra-low-dose CT," Heliyon **3**(8), e00393 (2017).
23. V. Turchenko, E. Chalmers, and A. Luczak, "A deep convolutional auto-encoder with pooling-unpooling layers in caffe," arXiv preprint arXiv:1701.04949 (2017).
24. W.-C. Wang, C.-W. Chow, L.-Y. Wei, Y. Liu, and C.-H. Yeh, "Long distance non-line-of-sight (NLOS) visible light signal detection based on rolling-shutter-patterning of mobile-phone camera," Opt. Express **25**(9), 10103–10108 (2017).
25. N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, P. Luo, and H. Le-Minh, "Non-line-of-sight 2 × N indoor optical camera communications," Appl. Opt. **57**(7), B144–B149 (2018).
26. J.-K. Lain, F.-C. Jhan, and Z.-D. Yang, "Non-line-of-sight optical camera communication in a heterogeneous reflective background," IEEE Photonics J. **11**(1), 1–8 (2019).
27. X. Li, N. B. Hassan, A. Burton, Z. Ghassemlooy, S. Zvanovec, and R. Perez-Jimenez, "A simplified model for the rolling shutter based camera in optical camera communications," in *2019 15th International Conference on Telecommunications (ConTEL)*, (IEEE, 2019), pp. 1–5.
28. Y. Goto, I. Takai, T. Yamazato, H. Okada, T. Fujii, S. Kawahito, S. Arai, T. Yendo, and K. Kamakura, "A new automotive VLC system using optical communication image sensor," IEEE Photonics J. **8**(3), 1–17 (2016).

29. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016). [Online; accessed 1-June-2021] Available: www.deeplearningbook.org.
30. M. Khalid, J. Baber, M. K. Kasi, M. Bakhtyar, V. Devi, and N. Sheikh, "Empirical evaluation of activation functions in deep convolution neural network for facial expression recognition," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, (2020), pp. 204–207.
31. L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," The Journal of Machine Learning Research **18**(1), 6765–6816 (2017).
32. Sony Corporation, *IMX219PQH5-C, Diagonal 4.60 mm (Type 1/4.0) 8 Mega-Pixel CMOS Image Sensor with Square Pixel for Color Cameras, Datasheet* (Sony Corporation, 2014).
33. Thorlabs Scientific Imaging, "Thorlabs DCx camera functional description and SDK manual," (2018). [Online; accessed 1-June-2021] Available: www.thorlabs.com/thorProduct.cfm?partNumber=DCC1645C.
34. C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, "Correlation-based receiver for optical camera communications," Opt. Express **27**(14), 19150–19155 (2019).

5.3 Artificial intelligence-assisted link parameter estimation

Citation:
[22] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, "Deep learning for signal clock and exposure estimation in rolling shutter optical camera communication," *Opt. Express*, vol. 30, pp. 20261–20277, Jun 2022

This work addressed the general objective GO1, proposing a novel receiver architecture for the simultaneous operation of cameras as visualization and data acquisition devices. In addition, it addresses the general objective GO2 with: (i) the design of an experimental platform for obtaining real-world samples, (ii) the improvement of the previously proposed algorithm for producing synthetic images, and (iii) the generation and publication of open datasets. Moreover, a later related work published at the CSNDSP conference further improves the proposed architecture, making it even more versatile and adaptable. At the end of this introduction, the contributions of the related publication, which is included in the Annex B, will be briefly described.

The main contribution of this work is the development of a generic AI-assisted architecture that operates indistinctly with a vast number of cameras, whose internal configuration (exposure time and row sampling period) is unknown in advance. Furthermore, this architecture is designed to work exclusively with image streams regardless of their origin, which can be cameras or multimedia files. Therefore, data acquisition routines are decoupled from the cameras' hardware and settings and can be deployed in a cloud infrastructure.

The central element in this architecture is the estimator, a CNN that estimates from images those link parameters required by the rest of the blocks in the reception chain, which, in this work, are the NPPS, and the ESR (Chapter 3). The former relates the transmitter's symbol rate with the camera's sampling rate (understood as the inverse of the row sampling period). Hence, it is related to the signal clock frequency (in terms of samples per symbol). The latter is the ratio between the exposure time and the symbol time.

In this work, both estimates are required to select the optimal equalizer from a bank of pre-trained exposure-related equalizers (Section 5.2). It should be noted that a single equalizer supports an ESR deviation (in the input images) of up to 11% with respect to the training ESR. Therefore, several overlapping pre-trained equalizers are required for the receiver to operate under different exposure times. The reason behind storing in memory the computed weights and biases of those equalizers instead of training on-demand a single equalizer lies in the required training time and resources. Longer training times, as is the case, would intensely occupy the receiver causing link disruptions and unacceptable latencies. The related contribution will solve this challenge by dramatically reducing the training time and resources by using transfer learning techniques. Moreover, the NPPS is also required in later stages to recover the signal clock.

As mentioned above, this estimator arises from the need to obtain these parameters for the correct reception when they are unknown beforehand. For example, the exposure setting is unknown when the camera (i) does not reveal its internal settings, for example, on firmware-protected devices (such as Android and iOS smartphones), (ii) does not allow to set it on-demand or (iii) dynamically adapts it based on the ambient illumination. Additionally, the IS's row sampling time is virtually unknown on most camera devices since revealing its internal clocks makes no practical sense. This parameter is specific to the hardware used and does not significantly impact the images from a photographer's point of view. Finally, although the transmitter symbol rate can be set as a design criterion, this is not desirable. Restricting the symbol rate forces the transmitter to either (i) be unable to dynamically adapt the data throughput based on link requirements or (ii) use complex protocols to notify the receiver of its current status.

This estimator works with profoundly impaired signals (i.e., dramatically affected by different exposure times), reassembling the functions of a traditional clock recovery system but with a significantly enhanced performance. Furthermore, it is not exclusively trained for discrete expected NPPS and ESR values. Instead, it covers all the possibilities within the interval designated for those parameters. This is achieved by training the network with thousands of representative synthetic images generated using an enhanced version of the previously presented generation procedure (Section 5.2).

This version adds three optional stages that make synthetic images even more realistic (Chapter 4). The first stage scales the signal and applies different camera gamma transformations. The second stage quantifies the pixel values based on the camera's bit resolution (in this case, 8 bits). This quantization maps the pixel float values to discrete integer values, adding the quantization noise. The final stage introduces JPEG compression artifacts. It should be highlighted that artifacts produced by other compression algorithms can also be explored, making the parameter estimation even more robust.

The evaluation of the estimator reveals relative errors for the NPPS and the ESR lower (on average) than 1.3% and 3%, respectively, with maximum values of 3% and 9%. These results guarantee that the proposed architecture will maintain a BER below the FEC limit based on the results obtained in the previous work. Therefore, they validated that this estimator can enable data reception without prior knowledge of the camera's hardware and configuration. Furthermore, the estimator can be used alternatively to characterize the camera exposure time and row sampling time by tuning the transmitter frequency. Accordingly, the estimator is also a laboratory tool for precisely characterizing cameras.

The main implication of this work is that it radically changes the design strategy of new OCC links. Rather than restringing the link design to particular transmitter-camera sets, it invites to search for methods, procedures, and subsystems that decouple image streams from data reception routines, shifting the focus from working with cameras to working with images. The estimator proposed in this work contributes to this purpose while allowing, at the same time, the simultaneous operation of cameras as data acquisition and imaging devices.

Another contribution of this work is the description of the iterative process carried out to improve the estimator training, reducing the gap between the validation and training errors through the expert adjustment of the synthetic training datasets.

Research Article Vol. 30, No. 12 / 6 Jun 2022 / *Optics Express* 20261

Optics EXPRESS

Deep learning for signal clock and exposure estimation in rolling shutter optical camera communication

CRISTO JURADO-VERDU,[1,*] **VICTOR GUERRA**,[2] **JOSE RABADAN**,[1] AND **RAFAEL PEREZ-JIMENEZ**[1]

[1] *Photonics and Communications Technology Division, Institute for Technological Development and Innovation in Communications (IDeTIC), Universidad de Las Palmas de Gran Canaria (ULPGC), Polivalente II, Planta 2, Las Palmas de Gran Canaria, PC: 35017, Spain*
[2] *Pi Lighting Sarl, Sion, Switzerland*
[*] *cjurado@idetic.eu*

Abstract: In rolling shutter (RS)-based optical camera communication (OCC) links, selecting the appropriate camera's exposure time is critical, as it limits the reception bandwidth. In long exposures, the pixels accumulate over time the incoming irradiance of several consecutive symbols. As a result, a harmful intersymbol interference corrupts the received signal. Consequently, reducing the exposure time is required to increase the reception bandwidth at the cost of producing dark images with impracticable light conditions for human or machine-supervised applications. Alternatively, deep learning (DL) equalizers can be trained to mitigate the exposure-related ISI. These equalizers must be trained considering the transmitter clock and the camera's exposure, which can be exceptionally challenging if those parameters are unknown in advance (e.g., if the camera does not reveal its internal settings). In those cases, the receiver must estimate those parameters directly from the images, which are severely distorted by the exposure time. This work proposes a DL estimator for this purpose, which is trained using synthetic images generated for thousands of representative cases. This estimator enables the receiver operation under multiple possible configurations, regardless of the camera used. The results obtained during the validation, using more than 7000 real images, registered relative errors lower than 1% and 2% when estimating the transmitter clock and the exposure time, respectively. The obtained errors guarantee the optimal performance of the following equalization and decoding receiver stages, keeping bit error rates below the forward error correction limit. This estimator is a central component of any OCC receiver that operates over moderate exposure conditions. It decouples the reception routines from the cameras used, ultimately enabling cloud-based receiver architectures.

1. Introduction

Optical camera communication (OCC) is a branch of visible light communication (VLC) [1], in which the optical receivers are the pixels of an image sensor. The interest in this technology lies in the reuse of embedded cameras in a wide range of end-user devices (e.g., smartphones, vehicle dashcams, laptops). In this way, it is intended to break the market's entry barriers imposed on VLC due to the need of using dedicated reception hardware (i.e., photodiode-based receivers) [2].

Notwithstanding, the actual reuse of rolling-shutter (RS) cameras for simultaneous data acquisition and scene visualization is challenging. Using RS-cameras as receivers requires the camera to be optimally configured to achieve the highest link throughput [3]. Its exposure time should be minimized, otherwise, it will restrict the available reception bandwidth (acting as a low pass filter [4,5]). However, under short exposure conditions, the camera delivers dark images

#458538 https://doi.org/10.1364/OE.458538
Journal © 2022 Received 15 Mar 2022; revised 13 May 2022; accepted 16 May 2022; published 23 May 2022

with impracticable light conditions for human or machine-supervised applications (i.e., users cannot perceive objects in the scene) [4].

Consequently, a tradeoff appears in selecting the suitable exposure time for a particular application. To obtain an in-depth understanding of how the exposure time affects communications is necessary to examine the RS-cameras' image acquisition mechanism [6]. In contrast to global shutter (GS) cameras that expose the whole image sensor during acquisition, RS cameras expose it sequentially row-by-row of pixels from top to bottom as shown in Fig. 1. In this acquisition, each row is activated after transcurred a fixed interval, known as the row sampling time, T_S. After the activation, the row's pixels remain exposed to light during a configurable amount of time, known as the exposure time, t_{exp}. As a result, each row samples the light at different sampling instants, generating an image composed of different intensity bands depending on the illumination state of the transmitter [6–9].

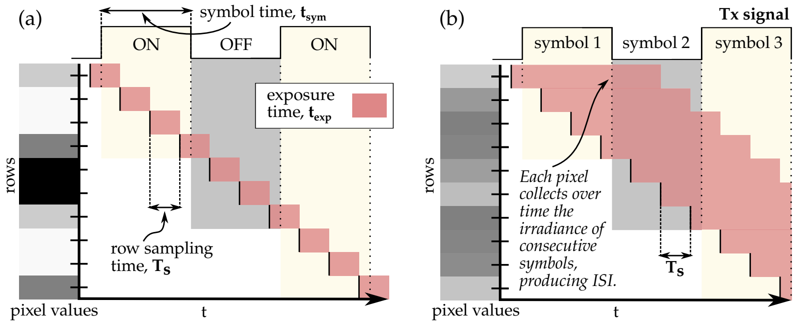


Fig. 1. Rolling shutter image acquisition mechanism.

Figure 2(a) shows an example of the generation of the symbol bands (or stripes) for binary transmission. From this figure, it can be extracted that the row sampling time (which generally ranges from tens of nanos to microseconds) constitutes, essentially, the receiver's sampling period. In contrast, in GS cameras, the receiver's sampling period corresponds to the time elapsed between two consecutive frames [7]. Thus, RS cameras outperform GS cameras in terms of achievable data rate [10].

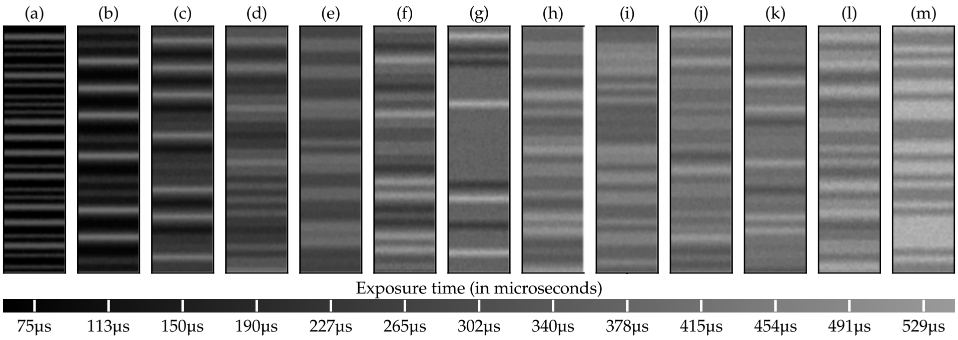


Fig. 2. Captures of the same transmission with increasing exposure time.

Fig. 1(a) illustrates an ideal sampling, in which the exposure time equals the row sampling time ($t_{exp} \leq t_{row}$). In this case, there is no overlap between the exposure of two consecutive rows. On the other hand, Fig. 1(b) illustrates a case where the exposure time is roughly 2.5 times greater than the symbol time. In this case, the pixels accumulate, over time, the incoming light irradiance of several consecutive symbols, corrupting the transmitted signal with a devastating intersymbol interference (ISI). As seen in this figure, the received signals are comparatively different. Their brightness increases with the exposure as the pixels accumulate light over a longer time, while their peak-to-peak amplitude decreases and varies more slowly. Figure 2 shows real captures under increasing exposure times for the same transmission (i.e., same frequency and data). It can be appreciated that these examples significantly differ either in shape or brightness, despite capturing the same transmitted signal, which reveals the complexity faced by the receiver at the decoding stage.

In conclusion, the exposure time has devastating effects, which are more detrimental depending on the ratio between the exposure and the symbol time, called the exposure to symbol ratio (ESR) [4]. Experimentally it can be demonstrated that the exposure effects are neglectable when the exposure time is less or equal to half the symbol time ($t_{\text{exp}} \leq t_{\text{sym}}/2$ or ESR $\leq 1/2$). This criterion remarkably restricts the number of different camera devices that can be used as receivers for a given transmitter. First, not all cameras can achieve the required short exposure. Second, they might rely on automatic algorithms for setting the exposure based on ambient light conditions. Finally, they might not allow access to their internal settings. Therefore, it is important to propose equalization stages to mitigate the inevitable exposure-related ISI.

To improve the reception's bandwidth, Liqiong et al. [11] propose a CNN-based neural architecture for demodulating on-off keying (OOK) signals captured with an RS camera. In addition, Yun-Shen et al. [12,13] propose two different AI-assisted approaches for demodulating 4-level pulse amplitude modulation (PAM4) signals: a pixel-per-symbol labelling (PPSL) and a long short-term memory neural networks (LSTM). Despite these architectures notably improve data decoding, their performance has been evaluated in relatively good exposure conditions. Based on the configured exposure time and the attained baud rates, in these works, the exposure time does not exceed the symbol time (ESR $\leq$ 1).

On the other hand, Younus et al. [14–16] propose the use of a one-dimensional artificial neural network (ANN) equalizer. This network can equalize OOK (or 4PAM) signals affected by exposure times that are up to 4 times the symbol time (ESR $\leq$ 4), improving the receiver's bandwidth by nine times compared to non-equalized receivers. However, this approach did not take advantage of the spatial redundancy of the data replicated in nearby columns [17], which can be exploited to increase the signal-to-noise ratio (SNR). Hence, the proposed equalizer is very sensitive to noise. In addition, the experimental setup used to train the system was very complex to conduct. It requires placing the receiver camera alongside a photodiode to obtain the training signals. In other words, the authors propose an ad-hoc solution that involved a highly time-consuming training phase for equalizing the images captured by a specific camera.

Alternatively, a two-dimensional convolutional autoencoder (CAE) equalizer trained using exclusively synthetic images was proposed in a previous work [4]. The validation (with real images) demonstrated that this network records bandwidth improvements of up to 14 times for OOK signals (compared to non-equalized receivers) under moderate SNR conditions (12 to 18dBs). This equalizer can decode signals with exposure times that are up to seven times longer than the symbol time (ESR $\leq$ 7), attaining bit error rates (BER) below the forward error correction (FEC) limit (i.e., 3.8×10^{-3}). Furthermore, since the training is conducted using synthetic images, it can be done offline and on-demand for a wide range of RS cameras with different characteristics.

For all the works above mentioned there is a common requirement. The receiver must be tuned, or adjusted, to the transmitter's clock frequency and the camera's exposure time. Meeting this requisite is exceptionally challenging when these parameters are unknown in advance or cannot be established. For example, when the transmitter dynamically adapts its transmission frequency, or, when the camera does not provide access to its internal settings [18], or dynamically adjusts its configuration based on ambient light estimations. In those cases, the receiver must recover those parameters directly from the received images, which is an exceptionally complex task because the images are severely distorted by the exposure.

In this work, a novel estimator block based on convolutional neural networks (CNN) is proposed to address this challenge. This estimator, pretrained with large synthetic datasets, ingests real images containing data packets and estimates the required signal parameters delivering them to the following equalization and the decoding stages: the signal clock and the camera's exposure time. Furthermore, the training dataset contains thousands of representative cases and noise conditions, enabling the estimator to operate effectively in many configurations, regardless of the camera used in the final deployment. The role of the proposed estimator becomes indispensable in any RS-OCC link that operates over moderate exposure conditions. It is responsible for recovering, under harsh conditions, the exposure time used for adjusting the equalizers and the signal clock required for data decoding. Consequently, it allows decoupling of the receiver's equalization and decoding routines from the hardware used (i.e. cameras), enabling cloud architectures that can handle many different image streams.

In addition, in this work, a dataset containing more than 7000 real-captured images for different exposure times and transmission frequencies was generated and released to the scientific community (Dataset 1 [19]).

The remainder of this paper is organized as follows. Section 1 introduces the proposed architecture and details the role of the estimator, which inputs it takes and which parameters deliver to the equalizer and the decoder in the reception chain. Section 2 describes the methods, procedures, metrics and the experimental setup used to evaluate the estimator's performance, including the synthetic network training and the validation using real captured images. Section 3 presents the results. Finally, the conclusions of this work are summarized in section 4.

2. Communications scheme

Figure 3 shows the proposed system architecture and its functional blocks. At the transmitter side, a uniform illuminated flat-panel LED sends non-return to zero (NRZ) Manchester encoded on-off keying (OOK) pulses. Pseudo-random data bits are grouped into packets with a fixed length and enclosed with a header consisting of five consecutive ones and a zero-bit trailer. In addition, after three consecutive bits, a stuffed bit is inserted, preventing the header sequence from appearing within the payload. This redundant bit is set to one if the preceding bit is zero or zero otherwise. For instance, considering the following payload bits, '11011', the transmitted data packet is '11111-110-1-11-0'. On the receiver side, two independent subsystems are interconnected through a shared interface: the image and data acquisition subsystems. The former consist of a generic RS-camera that continuously delivers a stream of compressed JPEG images. This camera is configured, if possible, to select its lowest exposure time required for human or machine-supervised applications. The data acquisition subsystem includes functional software components that can be either embedded in a hardware platform physically interconnected to the camera or deployed in a cloud infrastructure. The first block, the stream manager, controls different image streams and constitutes the input interface to the reception chain. Through this interface, it receives images and stores them in buffers until decoding resources are available.

The blocks that follow the stream controller are responsible for recovering the data embedded in the image: the equalizer, the decoder and the estimator blocks. In the proposed architecture, the equalizer consists of a bank of pre-trained CAEs (detailed in [4]) trained for different exposure times and signal clocks. Their function is to mitigate the exposure-related ISI and reduce the noise. On the other hand, the decoder uses 2D correlation techniques for packet detection, synchronization, and data decoding (as detailed in [20]). Figure 3 shows that these subsystems receive as inputs two important parameters: the number of pixels per symbol (NPPS) and the exposure to symbol ratio (ESR). The former, as introduced in [16], represents the theoretical number of samples per symbol. It is the ratio between the symbol time, t_{sym} and the RS-camera's row sampling period, T_s, (NPPS = t_{sym}/T_s). Hence, this parameter is directly related to the signal clock. As an example, in Fig. 1 the signal's NPPS is 3, i.e., theoretically, 3 pixels (or samples) is the expected height for each symbol. Emphasize that the NPPS is not necessarily an integer value. The latter is the ratio between the exposure, t_{exp} and the symbol times, t_{sym}. It indicates how deteriorated is the signal due to the exposure time. As an example, in Fig. 1(a) the ESR is lower than 0.5 (i.e., ≈ 0.33) and, hence, the receiver is operating in optimal sampling conditions ($t_{exp} \leq t_{sym}/2$). However, in Fig. 1(b), the ESR is 2.5, and the ISI is significant. In conclusion, these dimensionless parameters (i.e., NPPS and ESR) characterize the signal received and allow the equalizer to be adjusted for its optimal performance and the decoder to recover the signal clock and proceed with data acquisition.

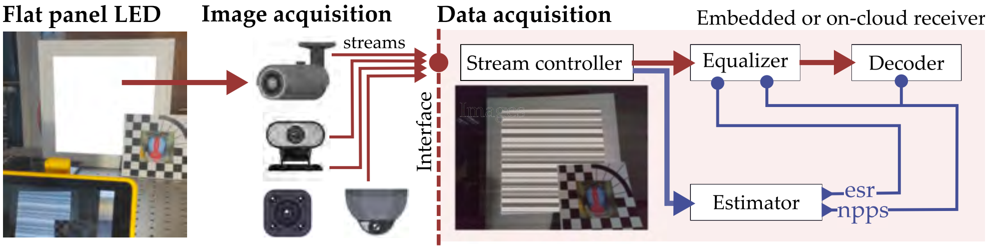


Fig. 3. Embedded or on-cloud receiver scheme.

Finally, the estimator block is responsible for predicting those parameters when they cannot be set or known in advance. To accomplish this, it ingests real images containing data packets and delivers the NPPS and ESR estimates to the corresponding blocks in the reception chain. In this way, the receiver communication algorithms are decoupled from the hardware used to stream the images, even from the camera's configuration.

It should be highlighted the complexity of estimating those parameters from the received signal when the exposure comes into play. Under ideal exposure conditions, the NPPS can be easily estimated by counting the number of sequences of ones and zeroes samples after thresholding the image with a fixed value, as detailed in [21,22]. In those works, the NPPS is known as the stripe width. Moreover, it is unnecessary to estimate the ESR as there is no need to perform the exposure-related equalization. This strategy for estimating the NPPS is accurate only under ideal sampling conditions. In any other case, the ISI severely distorts the signal, altering its shape, intensity and temporal evolution. This estimator is based on a convolutional neural network (CNN) (Fig. 4). The topology of this network consists of a sequence of feature extraction blocks (Fig. 4(b)) that obtains a set of features from the input images (Fig. 4(a)), and a regression artificial neural network (ANN) (Fig. 4(c)), that estimates the parameters as a function of those latent features. The feature extraction blocks are composed of depthwise separable convolutional layers (DSC) [23], replacing the 2D convolutional neural layers (Conv2D) used in traditional CNNs. Compared to the latter which performs the spatial and channel-wise convolution of the inputs using three-dimensional kernels (Fig. 4(e)), DSC layers split the computation into two steps (Fig. 4(f)): a depthwise convolution, followed by a pointwise convolution. The advantage of using DSC layers compared to Conv2D layers is that the number of training parameters is considerably reduced. Figure 4(e) compares the Conv2D and the DSC layers. The former uses a unique kernel, K, with the same number of channels (depth) as the input, I. The result of the convolution operation, Z, is a two-dimensional tensor, which is expressed in Eq. (1).

$$Z[i,j] = (I \circledast K)[i,j] = \sum_{m=0}^{k_w-1} \sum_{n=0}^{k_h-1} \sum_{l=0}^{k_d-1} I[m,n,l] \cdot K[i-m,j-n,l] \tag{1}$$

where i, j are the output's indexes, m, n, l, the horizontal, vertical and channel kernel's indexes, and k_w, k_h, k_d, the kernel's width, height and depth. On the other hand, a DSC separates this computation into two steps. First a depthwise convolution is applied to each independent channel of the image, I, using a two-dimensional kernel, $K_{\text{depthwise}}$. The outputs for each iteration are then stacked together, obtaining and output tensor that has the same depth as the input, expresed in Eq. (2). In this step, a kernel vector, $K_{\text{pointwise}}$ with a length, k_d equals to the number of input's channels is convolved with every spatial point of the stack, resulting in a two dimensional tensor, Z, which is expressed in Eq. (3).

$$Z'[i,j,k] = \sum_{\mathrm{n}=0}^{k_{\mathrm{w}}-1} \sum_{\mathrm{m}=0}^{k_{\mathrm{h}}} I[m,n,k] \cdot K_{\text{depthwise}}[i-m, j-n] \tag{2}$$

$$Z[i,j] = \sum_{\mathrm{l}=0}^{k_d} Z'[i,j,l] \cdot K_{\text{pointwise}}[l] \tag{3}$$

Following the convolution, the outputs are biased ($B[i,j]$) and transformed using a non-linear activation function, ϕ, which generates a feature map, $F[i,j]$ (Eq. (4)) for the next stage.

$$F[i,j] = \phi(Z[i,j] + B[i,j]) \tag{4}$$

The non-linear activation functions used in this work are the Leaky Rectified Linear Unit (Leaky ReLU) and the Sigmoid functions [24,25]. Afterwards, a sequence of pooling layers replaces the outputs in specific locations with a statistical summary of the outputs in the vicinity. This contributes to increase the non-linearity of the outputs and reduces the total number of network parameters. In this model, max-pooling and average pooling layers are used. The former returns the maximum value and the second the average value of a rectangular patch. This pooling stage starts with an average pooling layer (Fig. 4(g)) with kernel's dimensions 1xN where N is the number of columns considered. Then it follows a max-pooling layer (Fig. 4(h)) Mx1 where M is the number of rows.

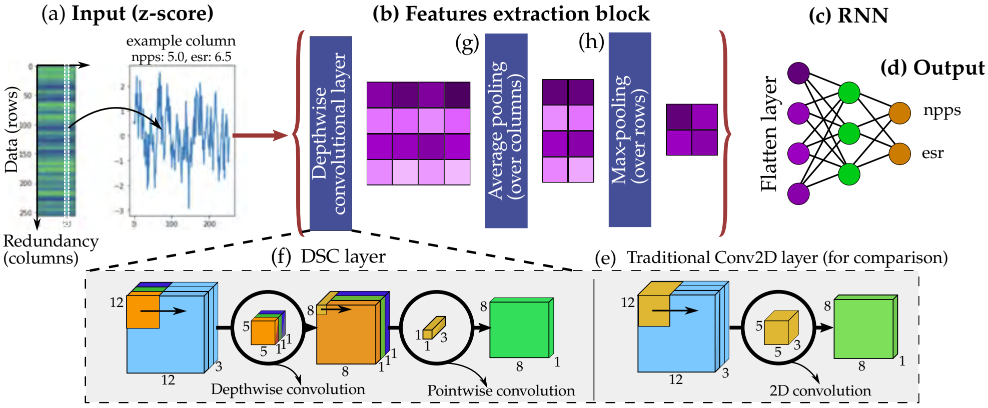


Fig. 4. Deep learning estimator model.

Finally, the RNN consists of a dense neural network with a single hidden layer and an output layer of 2 neurons, one per estimated parameter (i.e., NPPS and ESR). Also, as the outputs are normalized between zero and one, the activation function used in this case is the softmax.

3. Methodology

The evaluation of the proposed estimator's performance takes place in the following phases. In the first phase, both the synthetic and real datasets were generated. The algorithm used to create the synthetic images is detailed in [4]. For generating the real dataset (Dataset 1 [19]), the images were captured using the experimental setup shown in Fig. 5. It consists of a flat-panel LED pointing towards an RS-camera separated at a distance of 20 cm. At this distance, the transmitter occupies the image entirely. The signal is generated using an arbitrary wave generator, and a power supply is used to feed the light source. Table 1 summarizes the key parameters of the experiment necessary for its replicability.

The real dataset (Dataset 1 [19]) contains more than 7000 image samples for different NPPS and ESR. The NPPS values range from 4.0 to 7.0 in steps of 0.5 units and the ESR from 1.0 to 7.0 in steps of 0.5 units. After the generation of both datasets (synthetic and real), a rigorous analysis is carried out to verify that they are comparable, at least from the perspective of the neural network.

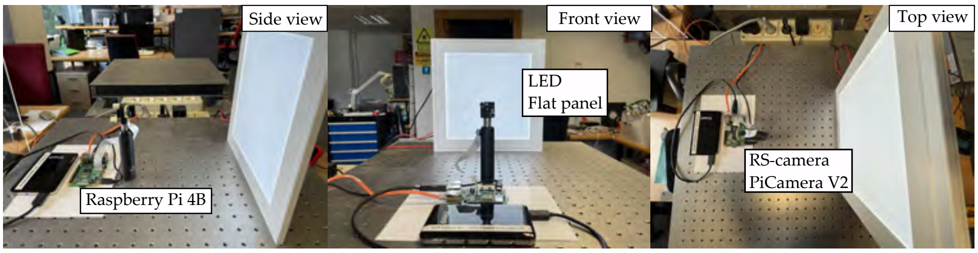


Fig. 5. Experimental setup.

Table 1. Experiment's key parameters.

Parameter	*Value*	Parameter	*Value*
Camera		*Transmitter*	
Hardware	Sony IMX219 [26]	LED color	Cold white
Aperture lens	f/2 |Focal length (3 mm)	Source voltage (V)	30 to 35
Image resolution	1920x1080 pixels	Symbol time, t_{sym} | NPPS	75 μs to 135 μs | 4 to 8
Exposure times, t_{exp}	20 μs to 1500 μs (step 19 μs)	Header, payload, stuffed, and trailer bits	5, 35, 12, 1
Sampling period, T_S	18.9035 μs	Random seed	31415

The following phase is the actual training of the network using the synthetic dataset. The training is conducted offline using a computer with local access to the datasets stored in memory. For training, the input images are z-score standardized, and the outputs are normalized between zero and one using a fixed value. The loss function used for training the network is the Mean Squared Error (MSE). Additionally, to improve the training results (i.e., minimizing the MSE), a network hyperparameter optimization is conducted, following the hyperband algorithm detailed in [27]. The layers of the optimized network model are summarized in Table 2. This table presents from top to bottom the subsequent layers, starting from the input layer, which ingests images, and finalizing with the output layer consisting of two neurons (one per estimate). In each row, it is detailed the layer type (such as DSC, average pooling, max pooling, dense,. . .), the total number of trainable parameters (such as the layer's weights and biases), specific parameters (such as the number of filters, the kernel size,..), and, finally, the layer's output shape. The latter is the shape of the output tensor delivered from a particular layer. For instance, the input layer returns a tensor with a shape of 32 images (i.e., corresponding to the training batch size) with a height and width of 256 and 64 pixels, respectively, and 1 colour channel (i.e., the image is in grayscale).

Table 2. Model summary.

Layer	Train params.	Description	Output shape
Input	0	-	(32, 256, 64, 1)
DSC	47	Filters=16, Kernel=(3,5)	(32, 256, 64, 16)
AveragePool	0	Kernel=(1,4), Strides=(1,2)	(32, 256, 32, 16)
MaxPooling	0	Kernel=(2,2), Strides=(1,1)	(32, 128, 16, 16)
DSC	688	Filters=32, Kernel=(3,3)	(32, 128, 16, 32)
AveragePool	0	Kernel=(1,2), Strides=(1,1)	(32, 128, 8, 32)
MaxPool	0	Kernel=(2,1), Strides=(1,1)	(32, 64, 8, 32)
DSC	2400	Filters=64, Kernel=(3,3)	(32, 64, 8, 64)
AveragePool	0	Kernel=(1,2), Strides=(1,1)	(32, 64, 4, 64)
MaxPool	0	Kernel=(2,1), Strides=(1,1)	(32, 32, 4, 64)
DSC	8768	Filters=128, Kernel=(3,3)	(32, 32, 4, 128)
AveragePool	0	Kernel=(1,2), Strides=(1,1)	(32, 32, 2, 128)
MaxPool	0	Kernel=(2,1), Strides=(1,1)	(32, 16, 2, 128)
DSC	17536	Filters=128, Kernel=(3,3)	(32, 16, 2, 128)
AveragePool	0	Kernel=(1,2), Strides(1,1)	(32, 16, 1, 128)
Flatten	0	-	(32, 2048)
Dense	524544	Units=256, bias=True	(32, 256)
Dense	16448	Units=64, bias=True	(32, 64)
Dense	128	Units=2, bias=False	(32, 2)
Total trainable params:		570,559	

The next phase consists of validating the model training using the real dataset (that has not been used during training). The validation metric is the root MSE (RMSE) of the estimates. In addition, based on the results obtained, the training is improved iteratively by refining the synthetic dataset to make it more similar to the real dataset. In this way, it is possible to progressively enhance the network's performance, reducing the validation error. In the last phase, the validation estimates are analyzed and dissected in the NPPS and ESR dimensions to examine possible influences between them. For instance, the ESR conditions may influence the estimation of NPPS and vice versa. Finally, the relative errors (RE) at estimating the transmitter clock and the camera's exposure time is computed.

4. Results

This section starts by introducing the preliminary examination and comparison between the synthetic and the real datasets. Then, the iterative process accomplished to refine the results obtained in the training and validation of the model is presented. At the end of this section, the final results obtained for the best training configuration are dissected and discussed, analyzing the model's precision and accuracy in estimating the NPPS and the ESR.

Figure 6 displays two example sets (left and right) with three different images to provide a visual comparison between the real and synthetic datasets. The first image (in those sets) belongs to the real dataset. It is a grayscale image that takes values between 0 and 255. The second image is a 2D tensor obtained from the z-score standardization of the previous image. It has values between -3 and 3 so that its mean and standard deviation are close to 0 and 1, respectively. The third image is the synthetic version (z-score standardized) created with the algorithm detailed in [4]. The left and right sets represent signals with NPPS of 4 and ESRs of 1.1 and 5.6, respectively. These sets are affected by different SNRs: 20 dB (left) and 7 dB (right). It should be remarked that only the z-score standardized tensors will be used as inputs in the training and validation. Therefore, they should be comparatively similar, at least from the perspective of the neural network. At first glance, they might look quite similar in this visualization. However, to give rigour to this analysis, the cross-correlation between both datasets was carried out. The resulting cross-correlation matrix is shown in Fig. 7. Each point of the correlation matrix represents the maximum correlation value obtained for a pair set (synthetic, real). Highlight that for this experiment, the transmitter is configured to send packets with the same payload, and the synthetic versions of these images also contain the same data. The results are distributed in the matrix as follows. The left axis represents the synthetic images, and the upper axis, the real images used for the correlation. The axis labels are sorted based on the pair sets (NPPS, ESR). From left to right and top to bottom, the NPPS value increases. Also, for each NPPS (e.g., 4.0), the ESR is increased until it reaches its maximum value, obtaining the corresponding pairs sets (e.g. ,(4.0, 1.0), (4.0, 1.5),. . . , (4.0, 7.0)).

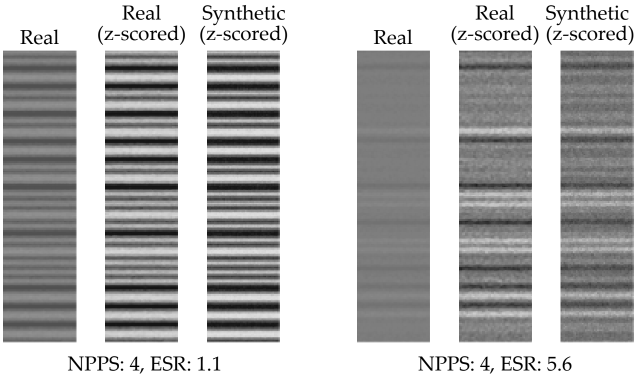


Fig. 6. Comparison between real and synthetic segments after applying the z-score standardization.

This matrix shows that the maximum values are in the diagonal, indicating that the synthetic generated versions match the camera images successfully. On the other hand, it can be appreciated that the correlation values decay softly from the diagonal towards the edges (i.e., the correlation values decrease softly as the ESR slightly increases or decreases). There is an explanation for this, and it is related to the inherent behaviour of the genuine acquisition regarding the exposure time. As the camera's exposure duration is increased (with respect to the original exposure conditions), the obtained images progressively lose their similarity with the initial image. The point at which increasing the exposure time makes two images significantly different can be determined from this matrix. This is the point where the correlation value decreases from 0.95 (high correlation) to 0.65 (low correlation). At this point, the ESR has increased (or decreased) by roughly one unit, as can be observed in Fig. 7(b). Finally, this matrix shows that the correlation output values are low for different NPPS (i.e., the images with different NPPS can be successfully distinguished). From these results, it can be preliminarily concluded that the model would perform worst at estimating the ESR compared to the NPPS.

Once the similarity of the images has been verified, it is important to highlight the benefits of using a synthetic dataset for training rather than the real dataset. The first advantage is reducing the generation time and memory resources. Capturing real images requires building and adapting

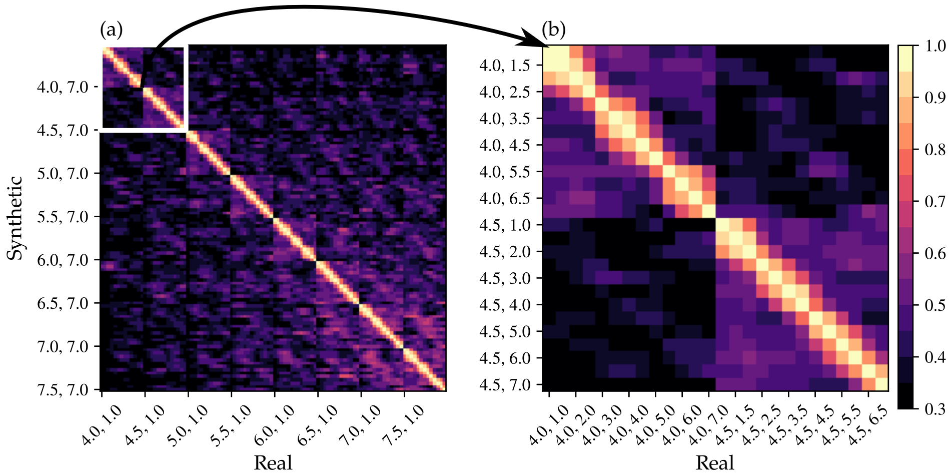


Fig. 7. Cross-correlation matrix of the real and synthetic datasets.

an experimental setup which consumes lots of resources and time. On the other hand, there is no need to rely on cameras to generate synthetic images; hence, they can be made in parallel at high speed. Moreover, specific parameters such as SNR and additional image compression effects are more controllable and rigorous in synthetic images. For example, in real images, to vary the SNR, it is required to precisely adjust the light power of the source, as it is considerably complex to control the noise contributions. In addition, the estimation of the SNR accurately in real images is a remarkably elaborate task. In contrast, different controlled noise contributions can be easily added to synthetic images. Finally, the use of synthetic images allows decoupling the network's training from the camera that will be used in the final deployment. This enables the training for a wide range of generic RS cameras.

Regarding the network's training, Fig. 8(a) shows the training and validation losses per training epoch. The solid blue line represents the validation loss achieved using the best training configuration. The blue dashed line represents, on the other hand, the training loss. Alternatively, the pink lines (solid and dashed) represent the losses (obtained in validation and training) for previous training configurations. These pink lines are used as references only to provide a visual comparison of the improvements obtained in each iteration. As shown in this graph, there is always an irreducible gap between validation and training losses. This is because the synthetic dataset (used for training) and the real dataset (used for validation) are not perfectly similar. However, despite this gap, the validation loss precisely follows the training loss, indicating that the model can optimally generalize the features of the training images without overfitting.

Highlight that the effectiveness of the training depends not only on the network's architecture but also on the design and selection of the appropriate training and validation datasets. Therefore, although the network hyperparameters were optimized using the original datasets, it is possible to further reduce the validation losses by adjusting the training dataset. This iterative dataset refinement procedure is analyzed in Fig. 8(b-f) based on the MSE obtained in the validation. Remark that this metric follows the same evolution as the loss, as the latter is derived from the MSE. However, the loss values are higher because of the regularizing parameters. In these graphs, the validation (solid line) and training (dashed line) MSE curves are coloured in red for the latest best training configuration (reference) and blue after introducing a new change. The pink lines are also maintained as references to other iterations. In the first iteration, the original synthetic generation algorithm [4] is modified by adding a binary quantization of the

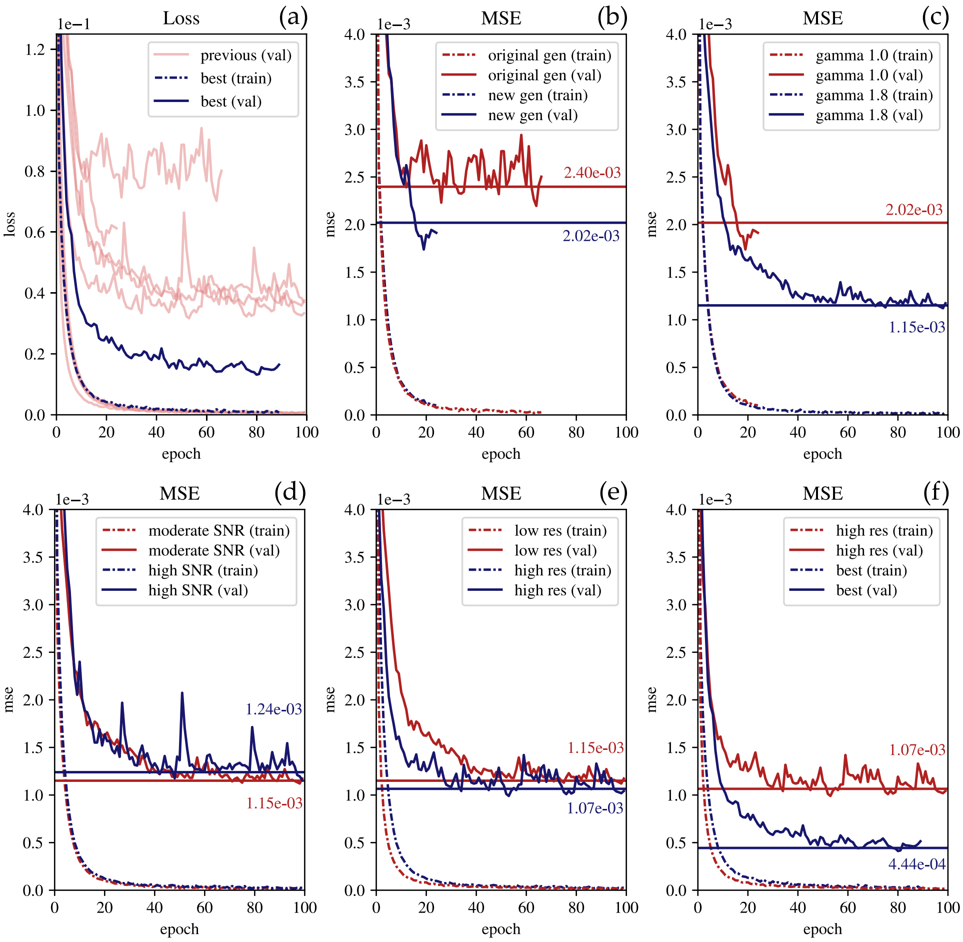


Fig. 8. Training and validation losses and MSE over epochs.

synthetic images, restringing its values to integers between 0 and 255. This step introduces a quantization error that the network can exploit efficiently. Furthermore, it makes synthetic images more representative by removing non-discrete values that can not happen in real images. In addition, a JPEG compression stage follows this binarization with different qualities (from 75 to 90). This introduces some artificial effects observed in real images. As shown in Fig. 8(b), these modifications turn out to be adequate and further reduce the gap between the validation and training MSE. In the second iteration (Fig. 8(c)), a gamma transformation of 1.8 is applied to the samples of the synthetic dataset, reducing the validation MSE considerably. In the third iteration (Fig. 8(d)), the SNR range used to generate the training images is increased. Instead of using an interval of 5 to 30 dB, it was extended from -2 to 40 dB. It can be seen that this approach does not reduce the validation MSE significantly. Furthermore, it has a drawback as it makes the validation unpredictable and unstable. For this reason, the original SNR range was preserved in the following iterations. In the fourth iteration (Fig. 8(e)), the number of images used from training is augmented. The NPPS and ESR's resolution step of 0.5 units is reduced to 0.25 units. The result is a slight improvement in the validation MSE. Based on the previous result, a combined strategy is accomplished in the last iteration (Fig. 8(f)). First, the gamma is increased to 2.2 (matching the gamma commonly used in JPEG-encoded images). Second, the training space was generated using random uniform distributed values for the NPPS and the ESR instead of picking them from a grid with a fixed resolution step. This procedure provides a significant improvement in the final validation MSE.

The model and the weights trained in this last iteration will be used in the final evaluation that is presented in the remainder of this section.

Regarding the estimator performance, Fig. 9(a) shows the estimates obtained from the validation dataset. The x-axis represents the NPPS, and the y-axis represents the ESR. The black star markers denote the target points, i.e., the ground truth. The black dots denote the mean of the estimates. Finally, the coloured dots represent the estimates obtained for different images. Dots with the same colour belong to the same target value. The colour does not hold any special meaning. It is only used to ease the identification of different clusters. At first glance, it can be seen that the dots in the lower part of the graph form clusters that are smaller than in the upper part (i.e., where the dots are more spread). This indicates that the model estimations are more precise under short to moderate exposure conditions (i.e., ESR from 1 to 3) than under high conditions (i.e., ESR from 5 to 7). In other words, the estimations under short to moderate exposure conditions have a lower error. Consequently, different errors are obtained depending on the location of the target values in the space domain.

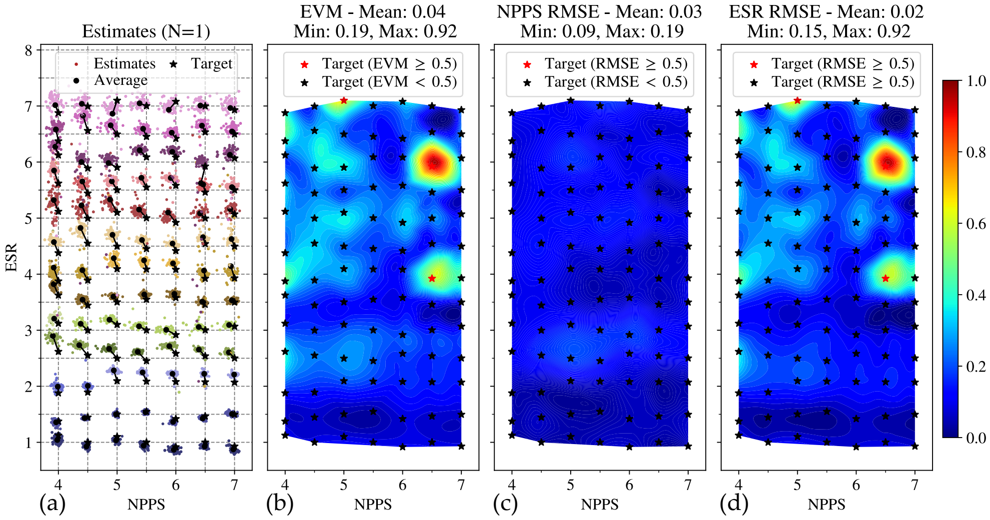


Fig. 9. NPPS and ESR estimates for the validation dataset.

The error vector magnitude (EVM) is used to quantify the errors in the estimations. Figure 9(b) shows the EVM obtained at different points in the space domain. The areas coloured in red represent areas where the estimates have the greatest observed error considering their target values (i.e., EVM equals 1). On the other hand, the areas coloured in blue represent the areas with the lowest estimation errors. As it can be seen, the point (NPPS=6.5, ESR=6) has the highest estimation error, while the point (NPPS=5, ESR=1.5) has the lowest.

However, although this metric quantifies well the estimation error, it does not provide an in-depth analysis of the nature of the error made. Ultimately the EVM metric combines the error obtained in the NPPS and ESR dimensions. Therefore, it is necessary to analyze the NPPS and the ESR errors independently. For this purpose, Fig. 9(c) shows RMSE obtained for the NPPS and Fig. 9(d) the RMSE for the ESR, respectively. In both cases, the RMSE values vary from 0 (blue) to 1 (red). These graphs reveal that the model can estimate the NPPS more accurately than the ESR. The ESR estimation errors contribute highly to the EVM.

Examining the EVM map in detail, it is observed that the errors are greater as the ESR increases, which is in line with the expected behaviour of the network. The longer the exposure time, the more severe distortion affects the received signal.

On the other hand, the errors also increase with the NPPS. This seems to contradict the expected behaviour, since increasing the redundancy per symbol (higher NPPS) should help the network at estimating the output parameters. However, the opposite seems to be happening. The reason behind this lies in the fact that the dimensions of the network's input image are fixed to 256x64x1 pixels. Hence, the samples of the transmitted signal correspond just to 256 pixel rows. Consequently, increasing the redundancy of symbol samples comes at the cost of reducing the number of different symbols that fit inside the image. In other words, increasing the redundancy reduces the variability of the signal samples. For instance, for NPPS equal to 7, the number of symbols within the image is approximately 36 (36 = 256 / 7), while for NPPS equal to 4, this number is 64. As a consequence, for higher NPPS, the network might deliver estimates that greatly deviate from their target value, estimation outliers. For example, in Fig. 9(a) can be seen a few purple-coloured dots around the coordinates (NPPS=6.5, ESR=3.5) when they should be located near their target point (NPPS=6.5, ESR=6). Nevertheless, these outliers represent a rare case. For this particular target point (NPPS=6.5, ESR=6), only four estimates of 85 fall in a region far distant from their target point. Furthermore, it is experimentally validated that these outliers are generated systematically and predictably. They appear when the transmitted payload bitstream has a considerable number of '01' or '10' bit sequences chained together (e.g. '01010101$\cdots$'). This causes the final transmitted signal, which is Manchester encoded, to be confused with another one generated with half the actual clock frequency. This is a classical problem in traditional clock recovery systems based on Manchester encoded systems, which is solved by inserting packet preambles or by sending pilot clock signals. Definitely, the estimator might be confused when facing those rare cases, which are more likely to appear when the NPPS is higher, due to the fixed size of the input image. These outliers can be avoided by changing the proposed bit stuffing technique to prevent the presence of '01' or '10' bit sequences or by inserting packet preambles.

Alternatively, outlier-resistant methods can be applied to a set of different estimates to effectively eliminate the presence of these outliers, revealing the actual trend of the errors that the network makes in the estimation of the output parameters.

The first strategy consists of computing the mean of a set of N estimates. This set is generated by collecting the estimates delivered from N random input images. The final ESR and NPPS estimates are obtained by computing the mean of the previous set.

Figure 10 shows the estimates obtained for N equal 5. In this figure, it can be observed that the impact of outliers is partially mitigated. The RMSE errors for the NPPS and the ESR are considerably reduced. However, Fig. 10(a) shows that there are still some estimates that spread over the ESR dimension. For example, the dots belonging to the target points (NPPS=6.5, ESR=4.0) and (NPPS=7, ESR=4) are spread within the range of ESR from 4.0 to 4.5. Consequently, with this approach, the outliers still significantly impact the ESR estimation.

On the other hand, given the low probability of the appearance of the outliers, the delivered estimations can be further improved by using the median instead of the mean. Figure 11 shows the estimates obtained for N equal 5. In contrast with the mean, the median produces much better results. The dispersion of the estimates in the ESR domain using the median is significantly reduced. This graph verifies that after eliminating the effect of the outliers, the behaviour of the network corresponds to what is expected. The errors are greater for low NPPS and high ESR.

Finally, Fig. 12 displays the relative error (RE) obtained for the NPPS and the ESR separately using the median approach, with N equals 10. To improve the visualization, the colour scale

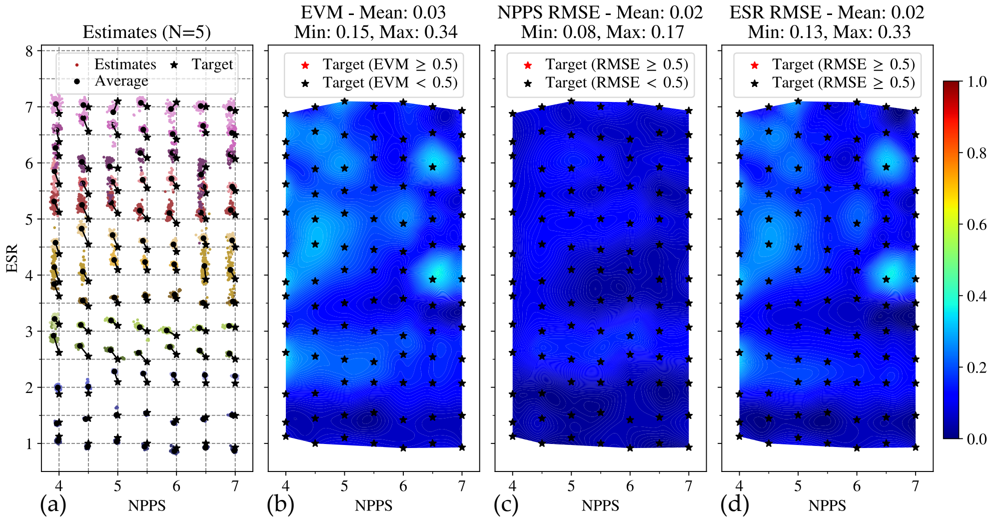


Fig. 10. NPPS and ESR estimates using the average of the outputs of 5 random images.

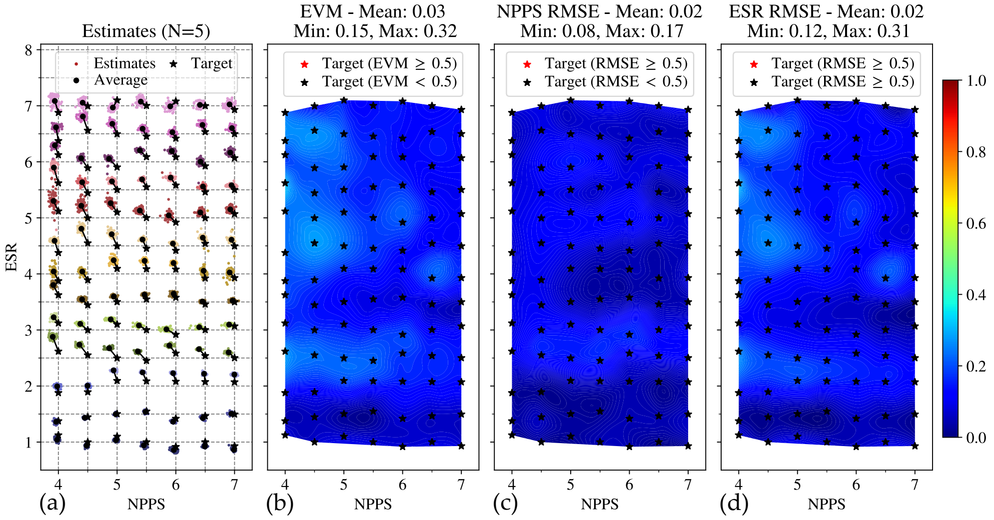


Fig. 11. NPPS and ESR estimates using the median of the outputs of 5 random images.

ranges from 0 to 0.14. These plots show that the maximum and minimum RE observed for the NPPS are approximately 3% and 0.2%, respectively. In the case of the ESR, they are 9% and 0.6%, respectively. On average, the error for both estimations is approximately 2%. From the communications perspective, these results imply, on the one hand, that the estimation of the signal clock, related to the NPPS, has a minimum and a maximum RE of 0.2% and 3% respectively. Furthermore, if the transmission clock frequency is known, the camera's row sampling time can be characterized by using the relation (NPPS = $t_{\text{sym}}/T_{\text{S}}$). The error made in the estimation of the row sampling time can be computed using the error propagation theory.

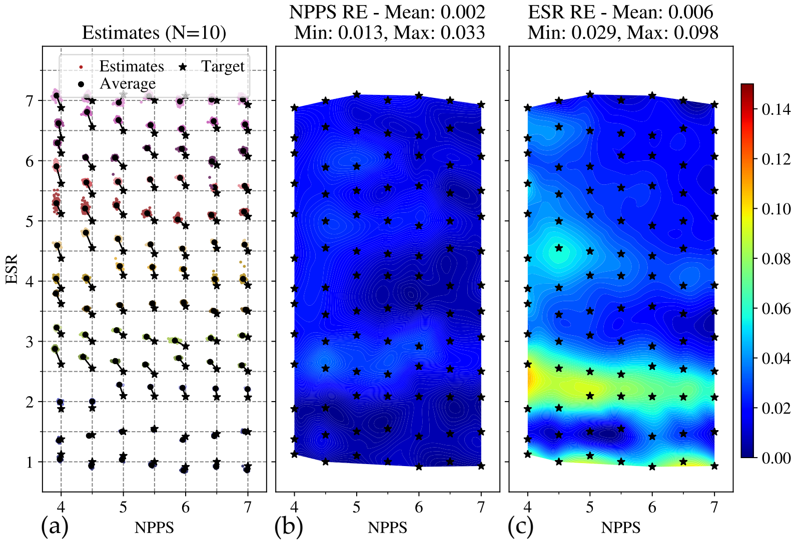


Fig. 12. NPPS and ESR RE using the median of the outputs of 10 random images.

On the other hand, the RE made at estimating the camera's exposure time is upper bounded to 9%. As discussed in the previous work [4], an exposure-equalizer trained for a given exposure time can successfully equalize images exposed with slightly higher or lower exposure times. Those equalizers allow a maximum deviation of up to 11% for the target exposure time while ensuring a Bit Error Rate (BER) lower than the Forward Error Correction (FEC) limit ($3.8x10^{-3}$). Therefore, as the estimator's ESR estimates have an average RE of 2% (upper bounded to 9%), it is ensured that the receiver will select the optimal equalizer based on the delivered ESR.

5. Conclusions

This work presents and evaluates a deep learning-based approach to accurately estimate two fundamental parameters of optical signals acquired with RS-cameras: the NPPS and the ESR. These dimensionless parameters relate the transmitted symbol duration with the camera's sampling frequency and exposure time, respectively. Hence, the NPPS is directly related to the signal clock and the ESR with the camera's exposure settings. These parameters are required during the equalization and decoding stages at reception. Its precise estimation will allow the receiver to select the optimal equalizer to mitigate the exposure-related ISI and recover the clock for synchronization and decoding, even when the transmission frequency and the camera's internal settings are unknown. Therefore it becomes an essential part of RS receivers operating over moderate exposure conditions. In addition, this estimator decouples the reception algorithms from the image stream providers, enabling cloud architectures that can practically handle multiple camera devices (or another type of image streamers). On the other hand, it can be used to characterize cameras in the time domain if needed. For example, if the camera does not reveal its internal settings due to operative system constraints. In addition, the network's training using synthetic images covers a vast training space, with thousands of representative cases considering multiple configurations for the NPPS and the ESR and different SNR conditions. Furthermore, the network's validation using a real dataset favoured the introduction and evaluation of incremental training improvements by refining the original datasets. The final evaluation of the model shows that the minimum, mean, and maximum relative errors for the NPPS estimates are 0.2%, 1.3%, and 3%, respectively. This implies an average RE of 1% at determining the transmission frequency of the source. Furthermore, these errors are significantly low when the exposure time is shorter. On the other hand, the minimum, mean and maximum relative errors for the ESR estimates are 0.6%, 3% and 9%, respectively. In this estimation, the neural network produced worse estimates, and still, the REs obtained are consistently lower (i.e. RE < 9%) than the 11% ESR deviation supported by pretrained equalizers. Therefore, it is ensured that the BERs after the equalization will remain below the FEC limit (3.8×10^{-3}) under the stated conditions. It should be highlighted that the RE does not exceed 3% in most cases, which indicates that the system is significantly robust in this estimation, favouring a better performance of the equalizers. In conclusion, this estimator is essential in all RS-based OCC links, in which both the signal clock and the exposure time must be retrieved from the images. Consequently, it enables the design of novel generic OCC links that do not require setting rigid requirements for the transmitter and the camera settings. Instead, those links will cover many different RS camera devices.

Funding. European Cooperation in Science and Technology (NEWFOCUS COST action (Ref: CA19111)); Agencia Estatal de Investigación (Project OCCAM, Ref. PID2020-114561RB-I00).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Dataset 1, [19].

Supplemental document. See Supplement 1 for supporting content.

References

1. "IEEE Standard for Local and metropolitan area networks–Part 15.7: Short-Range Optical Wireless Communications, *IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011)* pp. 1–407 (2019).
2. S. A. H. Mohsan, "Optical camera communications: practical constraints, applications, potential challenges, and future directions," J. Opt. Technol. **88**(12), 729–741 (2021).
3. G. Cossu, A. Sturniolo, and E. Ciaramella, "Modelization and characterization of a cmos camera as an optical real-time oscilloscope," IEEE Photonics J. **12**(6), 1–13 (2020).
4. C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, and R. Perez-Jimenez, "Convolutional autoencoder for exposure effects equalization and noise mitigation in optical camera communication," Opt. Express **29**(15), 22973–22991 (2021).
5. X. Li, N. B. Hassan, A. Burton, Z. Ghassemlooy, S. Zvanovec, and R. Perez-Jimenez, "A simplified model for the rolling shutter based camera in optical camera communications," in *2019 15th International Conference on Telecommunications (ConTEL)*, (IEEE, 2019), pp. 1–5.
6. H. Aoyama and M. Oshima, "Line scan sampling for visible light communication: Theory and practice," in *2015 IEEE International Conference on Communications (ICC)*, (2015), pp. 5060–5065.
7. W. A. Cahyadi, Y. H. Chung, Z. Ghassemlooy, and N. B. Hassan, "Optical camera communications: Principles, modulations, potential and challenges," Electronics **9**(9), 1339 (2020).
8. N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, "Optical camera communications: Survey, use cases, challenges, and future trends," Phys. Commun. **37**, 100900 (2019).
9. N. T. Le, M. Hossain, and Y. M. Jang, "A survey of design and implementation for optical camera communication," Signal Process. Image Commun. **53**, 95–109 (2017).
10. C.-W. Chow, C.-Y. Chen, and S.-H. Chen, "Visible light communication using mobile-phone camera with data rate higher than frame rate," Opt. Express **23**(20), 26080–26085 (2015).
11. L. Liu, R. Deng, and L.-K. Chen, "47-kbit/s rgb-led-based optical camera communication based on 2d-cnn and xor-based data loss compensation," Opt. Express **27**(23), 33840–33846 (2019).
12. Y.-S. Lin, C.-W. Chow, Y. Liu, Y.-H. Chang, K.-H. Lin, Y.-C. Wang, and Y.-Y. Chen, "Pam4 rolling-shutter demodulation using a pixel-per-symbol labeling neural network for optical camera communications," Opt. Express **29**(20), 31680–31688 (2021).

13. C.-W. Peng, D.-C. Tsai, Y.-S. Lin, C.-W. Chow, Y. Liu, and C.-H. Yeh, "Long short-term memory neural network to enhance the data rate and performance for rolling shutter camera based visible light communication (vlc)," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, (2022), pp. 1–3.
14. O. I. Younus, N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, L. N. Alves, P. A. Haigh, and H. Le Minh, "An artificial neural network equalizer for constant power 4-PAM in optical camera communications," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, (IEEE, 2020), pp. 1–6.
15. O. I. Younus, N. B. Hassan, Z. Ghassemlooy, P. A. Haigh, S. Zvanovec, L. N. Alves, and H. Le Minh, "Data rate enhancement in optical camera communications using an artificial neural network equaliser," IEEE Access **8**, 42656–42665 (2020).
16. O. I. Younus, N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, L. N. Alves, and H. Le-Minh, "The utilization of artificial neural network equalizer in optical camera communications," Sensors **21**(8), 2826 (2021).
17. P. Zhang, Q. Wang, Y. Yang, Y. Wang, Y. Sun, W. Xu, J. Luo, and L. Chen, "Enhancing the performance of optical camera communication via accumulative sampling," Opt. Express **29**(12), 19015–19023 (2021).
18. P. Nguyen, N. T. Le, and Y. M. Jang, "Challenges issues for occ based android camera 2 api," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, (2017), pp. 669–673.
19. C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, "Effects of the camera's exposure time on rolling shutter based optical camera communication links," figshare (2022), Accessed on: Feb. 10, 2022. [Online], doi: https://doi.org/10.6084/m9.figshare.19153166 (2022).
20. C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, "Correlation-based receiver for optical camera communications," Opt. Express **27**(14), 19150–19155 (2019).
21. Z. Huang, J. He, K. Yu, and W. Li, "Efficient demodulation scheme based on adaptive clock extraction andmapping-sampling for a mobile occ system," Appl. Opt. **60**(12), 3308–3313 (2021).
22. J. He, K. Yu, Z. Huang, and Z. Chen, "Multi-column matrices selection combined with k-means scheme for mobile occ system with multi-leds," IEEE Photonics Technol. Lett. **33**(12), 623–626 (2021).
23. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017).
24. A. Dubey and V. Jain, *Comparative Study of Convolution Neural Network's Relu and Leaky-Relu Activation Functions* (2019) pp. 873–880.
25. M. Khalid, J. Baber, M. K. Kasi, M. Bakhtyar, V. Devi, and N. Sheikh, "Empirical evaluation of activation functions in deep convolution neural network for facial expression recognition," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, (2020), pp. 204–207.
26. Sony Corporation, "IMX219PQH5-C Datasheet," Accessed on: Feb. 10, 2022. [Online]. Available: https://datasheetspdf.com/pdf/1404029/Sony/IMX219PQH5-C/1 (2014).
27. L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," The J. Mach. Learn. Res. **18**, 6765–6816 (2017).

5.3.1 Related contributions

Citation 1:
[27] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Effects of the camera’s exposure time on rolling shutter based optical camera communication links.,” *Figshare*, 3 2022

This dataset contains 7281 example images of a rolling shutter-based optical camera communication link. Those images are affected by increasing exposure times. In this dataset, the ESR ranges from 1 to 7 in steps of 0.5 units, and the NPPS ranges from 4 to 7 in steps of 0.5 units. The purpose of this dataset is to provide: (i) a tool for analyzing how the exposure time affects the received signal, (ii) a tool for validating novel AI-based approaches for the exposure equalization of the received signal, and (iii) a tool for validating novel estimator models of the intrinsic parameters of the received signal.

Citation 2:
[26] C. Jurado-Verdu, V. Guerra, C. Guerra, J. Rabadan, S. Zvánovec, and R. Perez-Jimenez, “On-demand training of deep learning equalizers for rolling shutter optical camera communications,” in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–5, 2022

This work addressed the same objectives as the previous work. Starting from a similar architecture, it adds structural changes to face the challenge of training a single exposure-related equalizer on demand. These modifications allow transferring the estimator’s knowledge acquired in an intensive offline training to the equalizers, using transfer learning techniques. To achieve this, both systems share the downward convolutional network used for extracting the image features. Transferring knowledge reduces the time and resources required for training the equalizer. With this approach, the training’s time is reduced by 435 times, and only 250 images are required rather than 35500. As a result, the equalizer training can be conducted on-demand during data acquisition without intensively occupying the receiver.

Chapter 6

Conclusions and future research

During this thesis, numerous contributions have been made regarding the mass adoption of RS systems. This mass adoption requires a solution that (i) is interoperable, i.e., that supports the great diversity of cameras available, and (ii) does not interfere with the primary functionality of cameras as imaging devices. For this purpose, it is necessary to properly understand the influence that specific camera parameters have on data reception. From the extensive analysis of the state of the art detailed in Chapter 2 it is clear that exposure time is a critical parameter that affects both image quality and communications. Short exposure settings reduce the sensitivity of the IS, resulting in dark images where objects are hardly discernible. In contrast, long exposures reduce the effective receiver bandwidth, causing the mixing of several consecutive symbols, thus, the appearance of ISI. In conclusion, the exposure time set a trade-off between the camera's sensitivity and the receiver bandwidth. However, equalizing the effects of this ISI will allow increasing the sensitivity of the sensor, leading to (i) images suitable for display, (ii) an increment of the receiver bandwidth, and (iii) an improvement in the SNR. On the other hand, the growing trend in the use of AI techniques and approaches for the detection, and the assisted demodulation and equalization stages in OCC links, positions AI as a suitable candidate for this purpose.

The three publications in the compendium of this thesis (Chapter 5), together with related contributions, illustrate the steps followed in the search for this solution.

In the first place, the requirement that cameras do not lose their functionality as imaging devices while operating as receivers is brought to light [20] (Section 5.1). This first paper analyzes the influence on the reception of (i) the deployment of the emitters in an industrial environment, (ii) the irradiance of the transmitter surface, and (iii) the camera's exposure time. From this analysis, a metric for the efficient distribution of nodes is derived based on: (i) the total aggregate data rate, (ii) the channel quality, and (iii) the space utilization ratio. In addition, an intelligent photobioreactor that simultaneously adapts its illumination to provide the best cultivation conditions for microalgae strains is proposed as the transmitter node.

In the second place, an AI-assisted equalization of the effects derived from long exposures is proposed [21] (Section 5.2) to enable effective reuse of RS cameras (i.e., simultaneous imaging and communication). The CAE equalizer proposed in this work increases the effective receiver bandwidth by a factor of up to 14 compared to non-equalized systems. It should be mentioned that the optical codes are encoded using

Manchester. In other words, this equalizer allows increasing the camera's exposure time up to 7 times the transmitted symbol period (i.e., ESR equals 7). Consequently, the camera's sensitivity can be significantly improved.

Compared to the results obtained by Younus et al. [88], who used a 1D NN for this task, the increase in effective bandwidth is 50% higher than they previously reported. Furthermore, these authors evaluated their model's performance in ideal SNR conditions (i.e., >30dBs). In contrast, the proposed CAE-based equalizer operates optimally at low to moderate SNR conditions (i.e., above 2dBs) while maintaining a constant BER below the FEC limit. This optimal performance is notably related to the coding-decoding strategy inherent in AE models. In addition, it can support an 11% deviation in the exposure time affecting the inputs with respect to the fixed exposure time value used for generating the training samples. A year after their first publication, the authors [56] used their previously proposed procedure for equalizing 4-PAM signals. In this case, given the complexity introduced by the multilevel modulation, the exposure time can only be increased up a factor of 1.3 the symbol time (i.e., ESR equals 1.3) before errors in the communication become significant. Finally, two more related works published after the date of publication of this proposal considers the exposure effects equalization of 4-PAM signals implicitly at the demodulation stage (Lin et al. [90], and He et al. [85]). The main difference between these two works lies mainly in signal generation and the network input size. The results obtained by Lin et al. show equalization capabilities for ESR equals 2.5. On the other hand, He et al. do not introduce a significant improvement in this aspect. Moreover, in their evaluation, the exposure time is half the symbol period (i.e., ESR equals 0.5).

Another main contribution of this proposal, which radically differentiates it from previous and subsequent work, is that it uses, for the first time, up to the authors' knowledge, exclusively synthetic samples in training the models. The procedure for the generation of synthetic samples, which is included in the publication and detailed in Chapter 4, only takes into account temporal parameters of the signal, such as the camera's exposure and sampling time, and the symbol period. It should be mentioned that the standardization step applied to both synthetic and real samples makes them comparable with greater independence of the signal strength. Furthermore, in later versions of this algorithm, the signal is randomly scaled (data augmentation) before applying non-linear transformations, taking into account the effects of several gamma factors and the distortions produced by JPEG compression. This allows for discarding the estimated expected signal strength at reception when generating the training samples. In other words, training and the normal operation of the equalizers are conducted without knowing the expected signal strength at the system deployment. Finally, the accuracy of the generation model is validated through the evaluation of the network using real-world samples.

The use of synthetic samples for training is significantly beneficial. First, it speeds up the acquisition time. For instance, other authors must develop a carefully characterized experimental setup to obtain the images. Moreover, idle times must be set between captures to avoid automation errors. Second, it also eliminates the acquisition's complexity. Because the transmitters and cameras are not perfectly synchronized, input and output samples must be correctly aligned before training. The previous authors rely on matched filtering techniques and well-known packet headers to solve this issue. This increases the complexity and introduces increasing errors as the exposure time grows. Third, it abstracts the receiver used. As samples can be generated for a wide range of transmitters, cameras, configurations, and noise conditions, the training is more robust and efficient, and the network can generalize well for different setups. In contrast, the previous authors evaluated their model with samples obtained from equivalent images found in the training dataset, using the same transmitter, camera, and configuration for training and testing the network. Therefore, the model's behavior is highly coupled with the experimental setup used to take the training samples and might not work well with slight changes in the operating conditions. Ultimately, their proposed system is an ad hoc solution that is difficult to implement given the great diversity of cameras. Fourth, the synthetic generation enables training on demand to perform the network's fine-tunings, avoiding interrupting the receiver's operation, which, in other cases, would require reserving time slots for the transmission of well-known signals for the receiver to obtain new samples.

Finally, the previously reported NN models use a 1D architecture with multiple inputs and one single output. This greatly conditions the efficiency of the equalization. First, it must be carried out iteratively by concatenating the network's outputs while traversing the entire column pixel by pixel. Hence, several iterations are required per image, depending on the column size and the network input dimensions. For example, if the expected packet size is 500 pixels, then at least 500 iterations must be accomplished per frame to decode the data, which makes this procedure completely inefficient. Second, it is required to reserve at least m-1 image pixels (where m is the input size) before (or after) the output sample, depending on whether its computation considers previous or subsequent pixel values. This affects the equalization/demodulation at the edges, reducing the effective ROI's area used for data recovery. In contrast, the output of a CAE has the exact dimensions as the input, allowing equalization to be carried by segments rather than by pixels. In each iteration, several pixels are equalized simultaneously (256 in this proposal). In addition, undesirable edge effects are effectively alleviated by the CAE, which conveniently evaluates both forward and backward pixels. Furthermore, the equalizer's input is two-dimensional, which has two advantages. On the one hand, it exploits the redundancy present in nearby columns, increasing the SNR. On the other hand, it might be robust to partial blocking, considering that at least one of the columns is not affected. This statement has not been validated yet, but preliminary experiments indicate this might be the case.

In third place, after demonstrating the feasibility of exposure equalization and validating the hypothesis H3, H2, and H1, a generic solution is conceptualized that abstracts (i.e., decouples) the reception routines from cameras and their exposure settings [22] (Section 5.3). This abstraction is achieved by estimating two important link parameters directly from the images: the exposure time required for equalizing the signal, and the signal clock, for decoding. The results reveal a very accurate estimation of these parameters with relative errors lower (on average) than 3% and 1.3%, respectively. Moreover, these errors remain within limits (with plenty of margin) supported by the proposed equalization and demodulation blocks, which guarantees the receiver's operation. Ultimately, these results directly validate hypothesis H4.

In conclusion, this architecture does not require prior knowledge of the camera used, its exposure setting, or the transmitter's baud rate. The estimator will deliver the estimates obtained directly from the images to the rest of the reception routines. It should be remarked that this proposal requires storing the parameters (i.e., a few megabytes) of several offline-trained equalizers adjusted for different exposure times.

The optimal equalizer will be selected based on the exposure estimate. The reason for not conducting an online training during the reception is that the training demands lots of resources, both in the number of samples (35500 synthetic images) and time. Thus, this training would require blocking the receiver intensively. To solve this problem, a related work, [26], introduces techniques for transferring the knowledge acquired by the estimator during an intensive (offline) training to the equalizer. In this way, the fine-tuned training of the equalizer is more efficient and can be performed on demand. The results reveal that transfer learning techniques in this context reduces the training time by a factor of 435, and the number of synthetic samples required is only 250 images. This allows the integration of this system in devices with few computational resources and, in addition, a more robust operation adapted to each scenario.

The results obtained in this thesis potentially impact the scientific community. First, the proposed and experimentally validated model for the generation of synthetic samples initiates, on the one hand, the path toward a deeper understanding of the influence of camera exposure on communications. On the other hand, it allows the development of RS link simulation tools for their preliminary evaluation, thus reducing the costs of their implementation. This, in turn, allows more resources to be devoted to developing new equalization stages and modulations more robust to long exposures. In addition, it will enable the massive generation of samples adapted to different scenarios and use cases. This impact can be recognized directly from the interest generated by the example dataset released in February 2022 [27]. At the time of writing this thesis, it has already exceeded 400 views and 150 downloads. Finally, it also enables the characterization of cameras in the time domain. This characterization can be achieved by correlating the images obtained with synthetic samples used as templates or using intelligent models trained for this purpose. For example, the CNN proposed for link parameter estimation enables alternatively to characterize the camera's exposure time and row sampling time. Second, the proposed AI-assisted equalization and estimation models exemplify the use of deep learning techniques in digital signal processing tasks. The methodology used to achieve the proposed solutions describes the problem and the exploratory analysis of the samples. It also discusses the choice of models that performs well in the sample domain and have been validated in similar tasks (such as image denoising). Finally, it details the iterative and supervised adjustment of the training samples to improve the model performance. Additionally, the proposed AI-assisted equalization is transferable to any communications system where the signal bandwidth is limited due to multiple factors: the transmitter's behavior, the channel response, or any element in the reception chain. Moreover, the exposure estimation is, ultimately, a direct estimation of the available bandwidth. Therefore, this strategy can be used to estimate the bandwidth cutoff frequency at reception. Third, there is the conceptualization of an RS-OCC system that decouples the reception routines from the cameras used, which invites progress in those research lines aimed at applying AI for estimating those parameters that might improve the link's performance, such as the relative movement of the emitters, the compression quality of the image, different atmospheric conditions, among others.

Regarding market impact, the results obtained accelerate the implementation of new OCC links, as reception routines are now compatible with many cameras. On the one hand, equalization stages alleviate the limitation of long exposure times. As a result, cameras that either (i) cannot set the exposure time to short values or (ii) cannot lose their functionality as imaging devices can still operate as receivers. This effective reuse of cameras makes this technology more attractive in scenarios and use cases that already use them for monitoring, surveillance, tracking, and other tasks. On the other hand, as already discussed, direct estimation allows reception to operate regardless of whether the camera's internal settings can be accessed or are automatically selected. For example, many manufacturers limit by firmware the access to the exposure time setting, and in many other cases, this parameter is automatically adjusted based on the ambient illumination.

Therefore, decoupling the reception routines from cameras enables the creation of a transferable communications software that can be easily integrated into developing multiple applications for smartphones, desktop and laptop computers, tablets, autonomous driving systems, ATMs, and more. Furthermore, it enables the development of applications hosted in cloud infrastructures that can process images supplied by different multimedia streams in real-time. In this way, it is also possible to technologically enable devices that do not have an open ecosystem to develop their applications. Also, these cloud applications can have additional functionality, such as coordinating and interoperating multiple streams.

Lastly, novel OCC transmitters have been developed which are directly compatible with current barcode scanners, Barcolits [25]. Hence, it is proved that this technology does not interfere with existing industrial processes, greatly facilitating its acceptance.

In short, the results of this thesis potentially accelerate the adoption of OCC technology, which translates into a potential impact on society. Furthermore, it could lead to new applications in multiple fields: interactive marketing, user identification and verification, guided assistance in museums, libraries, and entertainment venues, and key sharing.

In conclusion, this thesis balances the efforts to address (i) an industrial challenge in the particular field of RS-OCC links, which is achieving the effective reuse of cameras as receivers without losing their primary function as imaging devices, with (ii) a scientific-technical problem that is extensible to any communications field, which is the equalization of the ISI produced by the reduction of the available receiver bandwidth. Furthermore, it evolved and matured in line with the most recent advances in the field of AI and computer vision, registering results that exceed those obtained with classical algorithms and other preliminary work. Finally, it focuses on decoupling the reception on OCC links from the cameras used, with the aim of accelerating the mass adoption of this technology.

6.1 Future research

This thesis opens several future research lines in the field of OWCs, more specifically in RS-based OCC. However, as was aforementioned, the AI-assisted equalization techniques proposed in this thesis can be transferred to any communication system affected by a bandwidth restriction, either imposed by the transmitter, the channel, or the receiver's response. This extends the scope of the results of this thesis to the entire field of communications systems, regardless of whether they are radio-based, acoustic, or wired systems.

The following items summarise specific future research derived from this thesis.

- The analysis, modeling, and evaluation of the effect that the camera's exposure has on transmitters with a non-uniform illumination profile. The experiments carried out in this thesis used LED panels with a nearly uniform light intensity distribution profile; hence, the irradiance reaching the IS is uniformly distributed. Herefore, prior spatial equalization stages were not required or utilized in this thesis. This further reveals the robustness of the proposed exposure equalizers, which can minimize the impact due to possible deviations in the average intensity. However, when the irradiance reaching each pixel significantly varies, a spatial equalization stage might be required. This equalization task can be cascaded before or after the exposure equalizers or carried out directly by them.

- The analysis, modeling, and evaluation of the effect that the camera's exposure has on moving transmitters. From the literature review (Chapter 2), it is concluded that lateral motion produces skewness in the projection of the emitters, and vertical motion, the shortening or lengthening of the bands. Both motions are expected to produce appreciable distortions in the perceived emitter's intensity profile, especially if the transmitter's speed (in the image plane) is greater than the row sampling time. This effect is more detrimental as the exposure time increases and will negatively impact the exposure equalization.

- The analysis, modeling, and evaluation of the camera's exposure effect on sources subject to different adverse atmospheric conditions, such as rain, snow, and smoke. Another example might be assessing the impact of the presence of dust particles or in underwater environments. These phenomena might produce partial blocking or decrease the signal strength.

- The analysis, modeling, and evaluation of the camera's exposure effect on scenarios with the presence of interference agents. The experiments carried out in this thesis undergo controlled lighting conditions. Although preliminary tests indicate that the system will operate accurately under office lighting, its performance under different light interferers has yet to be evaluated.

- The analysis, modeling, and evaluation of different transformations done by the camera. The proposed synthetic generation only considers some of the transformations performed by the camera. For example, it only reflects the non-linear effect of gamma and the distortion produced by JPEG compression. Moreover, this thesis needs to analyze the possible noise sources for image generation in-depth. The synthetic samples only consider AWGN and quantization noise.

- The development of a simulation system that contemplates the sequential exposure of RS cameras. Further research might include this stage in the simulation pipelines of reference simulation software, such as Zemax.

- The development of a differentiable channel model including the RS acquisition. This differentiable model can be used to link transmitter and receivers models to compute the gradients of transmitter-based models based on a receiver-based loss function. This approach could help develop transmitter models to determine the best modulation schemes based on the channel conditions or get the best data transmission policies (using reinforcement learning).

- The assessment of the operation of the equalizers proposed in this thesis extended to other modulation and coding techniques, such as 4-PAM, QAM, among many others.

- The search for novel models that improve efficiency, reduce complexity, and perform better in the tasks of equalization and link parameter estimation. For example, further research might include using RNNs or transformer-based models.

- The development of cloud architectures. This includes the development of novel techniques for managing different image streams and priorities, classifying non-treatable image streams, and predicting communication interruptions, among others.

Bibliography

[1] H. Tataria, M. Shafi, M. Dohler, and S. Sun, "Six critical challenges for 6g wireless systems: A summary and some solutions," *IEEE Vehicular Technology Magazine*, vol. 17, no. 1, pp. 16–26, 2022.

[2] A. Trichili, M. A. Cox, B. S. Ooi, and M.-S. Alouini, "Roadmap to free space optics," *J. Opt. Soc. Am. B*, vol. 37, pp. A184–A201, Nov 2020.

[3] R. Marbel, R. Yozevitch, T. Grinshpoun, and B. Ben-Moshe, "Dynamic network formation for fso satellite communication," *Applied Sciences*, vol. 12, no. 2, 2022.

[4] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-based audio injection attacks on voice-controllable systems," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020.

[5] P. Chavez-Burbano, V. Guerra, J. Rabadan, and R. Perez-Jimenez, "Optical camera communication system for three-dimensional indoor localization," *Optik*, vol. 192, p. 162870, 2019.

[6] K. Aalimahmoodi, A. Gholami, and Z. Ghassemlooy, "Impact of camera parameters on the occ based indoor positioning system," in *2019 2nd West Asian Colloquium on Optical Wireless Communications (WACOWC)*, pp. 96–99, 2019.

[7] M. Rego and P. Fonseca, "Occ based indoor positioning system using a smartphone camera," in *2021 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pp. 31–36, 2021.

[8] "Purelifi - home page." Accessed on: Nov. 7, 2022. [Online]. Available: `https://purelifi.com/`.

[9] "Oledcomm - home page." Accessed on: Nov. 7, 2022. [Online]. Available: `https://www.oledcomm.net/`.

[10] "Signify - home page." Accessed on: Nov. 7, 2022. [Online]. Available: `https://www.signify.com/`.

[11] M. K. Hasan, M. O. Ali, M. H. Rahman, M. Z. Chowdhury, and Y. M. Jang, "Optical camera communication in vehicular applications: A review," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–22, 2021.

[12] M. Akram, R. Godaliyadda, and P. Ekanayake, "Design and analysis of an optical camera communication system for underwater applications," *IET Optoelectronics*, vol. 14, no. 1, pp. 10–21, 2020.

[13] Z. Zhou, W. Guan, S. Wen, W. Xu, and Y. Li, "Rse-based underwater optical camera communication impeded by bubble degradation," in *OSA Optical Sensors and Sensing Congress 2021 (AIS, FTS, HISE, SENSORS, ES)*, p. JTu5A.8, Optica Publishing Group, 2021.

[14] B. Majlesein, V. Matus, C. Jurado-Verdu, V. Guerra, J. Rabadan, and J. Rufo, "Experimental characterization of sub-pixel underwater optical camera communications," in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 150–155, 2022.

[15] B. Majlesein, V. Guerra, J. Rabadan, J. Rufo, and R. Perez-Jimenez, "Evaluation of communication link performance and charging speed in self-powered internet of underwater things devices," *IEEE Access*, vol. 10, pp. 100566–100575, 2022.

[16] S. Dwik and M. Lordwin Cecil Prabhaker, "Survey on energy harvesting cmos sensor based digital camera," *Optical Memory and Neural Networks*, vol. 31, pp. 97–106, Mar 2022.

[17] IEEE 820.15.7, "IEEE Standard for Local and metropolitan area networks–Part 15.7: Short-Range Optical Wireless Communications," *IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011)*, pp. 1–407, 2019.

[18] ISO 22738:2020, "Intelligent transport systems — Localized communications — Optical camera communication," *ISO 22738:2020*, p. 25, 2020.

[19] C. Jurado-Verdu, V. Guerra, and J. Rabadan, *Diseño, caracterización e implementación de un sistema de comunicaciones ópticas basado en cámara.* M.eng. thesis, Universidad de Las Palmas de Gran Canaria, 2019.

[20] C. Jurado-Verdu, V. Guerra, V. Matus, C. Almeida, and J. Rabadan, "Optical camera communication as an enabling technology for microalgae cultivation," *Sensors*, vol. 21, no. 5, 2021.

[21] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, and R. Perez-Jimenez, "Convolutional autoencoder for exposure effects equalization and noise mitigation in optical camera communication," *Opt. Express*, vol. 29, pp. 22973–22991, Jul 2021.

[22] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, "Deep learning for signal clock and exposure estimation in rolling shutter optical camera communication," *Opt. Express*, vol. 30, pp. 20261–20277, Jun 2022.

[23] C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, "Correlation based receiver for optical camera communications," *Opt. Express*, vol. 27, pp. 19150–19155, Jul 2019.

[24] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, R. Perez-Jimenez, J. Luis Gomez-Pinchetti, and C. Almeida, "Application of optical camera communication to microalgae production plants," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2020.

[25] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Barcolits: Barcodes using led tags and optical camera communications,” in *2022 IEEE 18th International Conference on Factory Communication Systems (WFCS)*, pp. 1–8, 2022.

[26] C. Jurado-Verdu, V. Guerra, C. Guerra, J. Rabadan, S. Zvánovec, and R. Perez-Jimenez, “On-demand training of deep learning equalizers for rolling shutter optical camera communications,” in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–5, 2022.

[27] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Effects of the camera’s exposure time on rolling shutter based optical camera communication links.,” *Figshare*, 3 2022.

[28] A. Mederos-Barrera, C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Discovering and tracking-based detection system for optical camera communication,” in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2020.

[29] V. Matus, S. R. Teli, V. Guerra, C. Jurado-Verdu, S. Zvanovec, and R. Perez-Jimenez, “Evaluation of fog effects on optical camera communications link,” in *2020 3rd West Asian Symposium on Optical and Millimeter-wave Wireless Communication (WASOWC)*, pp. 1–5, 2020.

[30] V. Matus, V. Guerra, C. Jurado-Verdu, S. R. Teli, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, “Experimental evaluation of an analog gain optimization algorithm in optical camera communications,” in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–5, 2020.

[31] A. Mederos-Barrera, C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, “Design and experimental characterization of a discovery and tracking system for optical camera communications,” *Sensors*, vol. 21, no. 9, 2021.

[32] V. Georlette, F. Piras, C. Jurado–Verdu, S. Bette, N. Point, and V. Moevaert, “Content triggering system using tricolor led strips and optical camera communication in rolling shutter mode,” in *2021 Third South American Colloquium on Visible Light Communications (SACVLC)*, pp. 01–06, 2021.

[33] V. Matus, V. Guerra, C. Jurado-Verdu, S. Zvanovec, and R. Perez-Jimenez, “Wireless sensor networks using sub-pixel optical camera communications: Advances in experimental channel evaluation,” *Sensors*, vol. 21, no. 8, 2021.

[34] V. Matus, V. Guerra, C. Jurado-Verdu, J. Rabadan, and R. Perez-Jimenez, “Demonstration of a sub-pixel outdoor optical camera communication link,” *IEEE Latin America Transactions*, vol. 19, no. 10, pp. 1798–1805, 2021.

[35] V. Matus, V. Guerra, C. Jurado-Verdu, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, “Design and implementation of an optical camera communication system for wireless sensor networking in farming fields,” in *2021 IEEE 32nd Annual*

International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1–6, 2021.

[36] C. Guerra-Yánez, V. Guerra, C. Jurado-Verdú, J. Rabadán, R. Pérez-Jiménez, Z. Ghassemlooy, and S. Zvánovec, "General framework for calculating irradiance distributions of symmetric surface sources," *Opt. Express*, vol. 30, pp. 43910–43924, Nov 2022.

[37] B. Majlesein, V. Matus, C. Jurado-Verdu, V. Guerra, J. Rabadan, and J. Rufo, "Experimental characterization of sub-pixel underwater optical camera communications," in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 150–155, 2022.

[38] M. Z. Chowdhury, M. T. Hossan, A. Islam, and Y. M. Jang, "A comparative survey of optical wireless technologies: Architectures and applications," *IEEE Access*, vol. 6, pp. 9819–9840, 2018.

[39] Y. Goto, I. Takai, T. Yamazato, H. Okada, T. Fujii, S. Kawahito, S. Arai, T. Yendo, and K. Kamakura, "A new automotive vlc system using optical communication image sensor," *IEEE Photonics Journal*, vol. 8, no. 3, pp. 1–17, 2016.

[40] N. T. Le, M. Hossain, and Y. M. Jang, "A survey of design and implementation for optical camera communication," *Signal Processing: Image Communication*, vol. 53, pp. 95–109, 2017.

[41] S. A. H. Mohsan, "Optical camera communications: practical constraints, applications, potential challenges, and future directions," *J. Opt. Technol.*, vol. 88, pp. 729–741, Dec 2021.

[42] N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, "Optical camera communications: Survey, use cases, challenges, and future trends," *Physical Communication*, vol. 37, p. 100900, 2019.

[43] W. LIU and Z. Xu, "Some practical constraints and solutions for optical camera communication," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, no. 2169, p. 20190191, 2020.

[44] T. Le, N.-T. Le, and Y. M. Jang, "Performance of rolling shutter and global shutter camera in optical camera communications," in *2015 INTERNATIONAL CONFERENCE ON ICT CONVERGENCE (ICTC)*, pp. 124–128, Korean Inst Commun & Informat Sci (KICS); IEEE Seoul Sect; IEICE Commun Soc, 2015. 2015 International Conference on Information and Communication Technology Convergence (ICTC), Minist Sci ICT & Future Planning (MSIP), Jeju Isl, SOUTH KOREA, OCT 28-30, 2015.

[45] W. A. Cahyadi, Y. H. Chung, Z. Ghassemlooy, and N. B. Hassan, "Optical camera communications: Principles, modulations, potential and challenges," *Electronics*, vol. 9, no. 9, 2020.

[46] P. Luo, M. Zhang, Z. Ghassemlooy, S. Zvanovec, S. Feng, and P. Zhang, "Undersampled-based modulation schemes for optical camera communications," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 204–212, 2018.

[47] C. Jurado-Verdu, V. Guerra, J. Rabadan, R. Perez-Jimenez, and P. Chavez-Burbano, "Rgb synchronous vlc modulation scheme for occ," in *2018 11th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, 2018.

[48] C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, "Correlation-based receiver for optical camera communications," *Opt. Express*, vol. 27, pp. 19150–19155, Jul 2019.

[49] H. Nguyen, T. L. Pham, H. Nguyen, and Y. M. Jang, "Trade-off communication distance and data rate of rolling shutter occ," in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 148–151, 2019.

[50] H. Aoyama and M. Oshima, "Line scan sampling for visible light communication: Theory and practice," in *2015 IEEE International Conference on Communications (ICC)*, pp. 5060–5065, 2015.

[51] Y. Liu, C.-W. Chow, K. Liang, H.-Y. Chen, C.-W. Hsu, C.-Y. Chen, and S.-H. Chen, "Comparison of thresholding schemes for visible light communication using mobile-phone image sensor," *Opt. Express*, vol. 24, pp. 1973–1978, Feb 2016.

[52] K. Liang, C.-W. Chow, Y. Liu, and C.-H. Yeh, "Thresholding schemes for visible light communications with cmos camera using entropy-based algorithms," *Opt. Express*, vol. 24, pp. 25641–25646, Oct 2016.

[53] Z. Zhang, T. Zhang, J. Zhou, Y. Lu, and Y. Qiao, "Thresholding scheme based on boundary pixels of stripes for visible light communication with mobile-phone camera," *IEEE Access*, vol. 6, pp. 53053–53061, 2018.

[54] V. Matus, V. Guerra, S. Zvanovec, J. Rabadan, and R. Perez-Jimenez, "Sandstorm effect on experimental optical camera communication," *Appl. Opt.*, vol. 60, pp. 75–82, Jan 2021.

[55] X. Li, N. B. Hassan, A. Burton, Z. Ghassemlooy, S. Zvanovec, and R. Perez-Jimenez, "A simplified model for the rolling shutter based camera in optical camera communications," in *2019 15th International Conference on Telecommunications (ConTEL)*, pp. 1–5, 2019.

[56] O. I. Younus, N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, L. N. Alves, and H. Le-Minh, "The utilization of artificial neural network equalizer in optical camera communications," *Sensors*, vol. 21, no. 8, 2021.

[57] P. Nguyen, N. T. Le, and Y. M. Jang, "Challenges issues for occ based android camera 2 api," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 669–673, 2017.

[58] A. Islam, L. Musavian, and N. Thomos, "Multi-agent deep reinforcement learning in vehicular occ," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pp. 1–6, 2022.

[59] T. Nguyen, A. Islam, and Y. M. Jang, "Region-of-interest signaling vehicular system using optical camera communications," *IEEE Photonics Journal*, vol. 9, no. 1, pp. 1–20, 2017.

[60] D. N. Choi, S. Y. Jin, J. Lee, and B. W. Kim, "Deep learning technique for improving data reception in optical camera communication-based v2i," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–2, 2019.

[61] T. L. Pham, M. Shahjalal, V. Bui, and Y. M. Jang, "Deep learning for optical vehicular communication," *IEEE Access*, vol. 8, pp. 102691–102708, 2020.

[62] M. F. Ahmed, M. Shahjalal, M. O. Ali, and Y. Min Jang, "Neural network-based led detection in vehicular system for high data rate in occ," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 662–665, 2020.

[63] M. F. Ahmed, M. K. Hasan, M. Shahjalal, M. M. Alam, and Y. M. Jang, "Design and implementation of an occ-based real-time heart rate and pulse-oxygen saturation monitoring system," *IEEE Access*, vol. 8, pp. 198740–198747, 2020.

[64] X. Sun, W. Shi, Q. Cheng, W. Liu, Z. Wang, and J. Zhang, "An led detection and recognition method based on deep learning in vehicle optical camera communication," *IEEE Access*, vol. 9, pp. 80897–80905, 2021.

[65] H. Takano, M. Nakahara, K. Suzuoki, Y. Nakayama, and D. Hisano, "300-meter long-range optical camera communication on rgb-led-equipped drone and object-detecting camera," *IEEE Access*, vol. 10, pp. 55073–55080, 2022.

[66] Q. Cheng, H. Ma, and X. Sun, "Vehicle led detection and segmentation recognition based on deep learning for optical camera communication," *Optoelectronics Letters*, vol. 18, pp. 508–512, Aug 2022.

[67] J. He and B. Zhou, "A deep learning-assisted visible light positioning scheme for vehicles with image sensor," *IEEE Photonics Journal*, vol. 14, no. 4, pp. 1–7, 2022.

[68] M. T. Kim and B. W. Kim, "Deepccb-occ: Deep learning-driven complementary color barcode-based optical camera communications," *Applied Sciences*, vol. 12, no. 21, 2022.

[69] M. H. Rahman, M. Shahjalal, M. K. Hasan, M. O. Ali, and Y. M. Jang, "Design of an svm classifier assisted intelligent receiver for reliable optical camera communication," *Sensors*, vol. 21, no. 13, 2021.

[70] M. F. Ahmed, M. K. Hasan, M. Z. Chowdhury, N. C. Hoan, and Y. M. Jang, "Continuous status monitoring of industrial valve using occ-enabled wireless sensor network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[71] H. Nguyen, V. L. Nguyen, D. H. Tran, and Y. M. Jang, "Rolling shutter ofdm scheme for optical camera communication considering mobility environment based on deep learning," *Applied Sciences*, vol. 12, no. 16, 2022.

[72] T. L. Pham, H. Nguyen, T. Nguyen, and Y. M. Jang, "A novel neural network-based method for decoding and detecting of the ds8-psk scheme in an occ system," *Applied Sciences*, vol. 9, no. 11, 2019.

[73] S. Zhong, Y. Zhu, X. Chi, H. Shi, H. Sun, and S. Wang, “Optical lensless-camera communications aided by neural network,” *Applied Sciences*, vol. 9, no. 16, 2019.

[74] T. L. Pham, T. Nguyen, M. D. Thieu, H. Nguyen, H. Nguyen, and Y. M. Jang, “An artificial intelligence-based error correction for optical camera communication,” in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 137–140, 2019.

[75] S. A. I. Alfarozi, K. Pasupa, H. Hashizume, K. Woraratpanya, and M. Sugimoto, “Robust and unified vlc decoding system for square wave quadrature amplitude modulation using deep learning approach,” *IEEE Access*, vol. 7, pp. 163262–163276, 2019.

[76] A. Pepe, S. D. Kumar, W. Zixian, and H. Y. Fu, “Data-aided color shift keying transmission for lcd-to-smartphone optical camera communication links,” in *Proceedings of the 2020 8th International Conference on Communications and Broadband Networking*, ICCBN ’20, (New York, NY, USA), p. 29–34, Association for Computing Machinery, 2020.

[77] S. Park and H. Lee, “Deep learning approach to optical camera communication receiver design,” in *2021 IEEE Region 10 Symposium (TENSYMP)*, pp. 1–5, 2021.

[78] X. Sun, Y. Yu, and Q. Cheng, “Led recognition method based on deep learning in uav optical camera communication,” *Appl. Opt.*, vol. 61, pp. 8688–8694, Oct 2022.

[79] J. Li and W. Guan, “The optical barcode detection and recognition method based on visible light communication using machine learning,” *Applied Sciences*, vol. 8, no. 12, 2018.

[80] L. Liu, R. Deng, and L.-K. Chen, “47-kbit/s rgb-led-based optical camera communication based on 2d-cnn and xor-based data loss compensation,” *Opt. Express*, vol. 27, pp. 33840–33846, Nov 2019.

[81] Z. Chen, R. Lin, H. Duan, Y. Chen, Y. Yang, R. Wu, and L. Chen, “Increasing the data rate for reflected optical camera communication using uniform led light,” in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 1274–1275, 2020.

[82] K. Yu, J. He, and Z. Huang, “Decoding scheme based on cnn for mobile optical camera communication,” *Appl. Opt.*, vol. 59, pp. 7109–7113, Aug 2020.

[83] Y.-S. Lin, Y. Liu, C.-W. Chow, Y.-H. Chang, D.-C. Lin, S.-H. Song, K.-L. Hsu, and C.-H. Yeh, “Z-score averaging neural network and background content removal for high performance rolling shutter based optical camera communication (occ),” in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2021.

[84] C.-W. Chow, Y. Liu, C.-H. Yeh, Y.-H. Chang, Y.-S. Lin, K.-L. Hsu, X.-L. Liao, and K.-H. Lin, “Display light panel and rolling shutter image sensor based optical camera communication (occ) using frame-averaging background removal and

neural network," *Journal of Lightwave Technology*, vol. 39, no. 13, pp. 4360–4366, 2021.

[85] J. He, Y. Yang, and J. He, "Artificial neural network-based scheme for 4-pwm occ system," *IEEE Photonics Technology Letters*, vol. 34, no. 6, pp. 333–336, 2022.

[86] P. Ling, M. Li, and W. Guan, "Channel-attention-enhanced lstm neural network decoder and equalizer for rse-based optical camera communications," *Electronics*, vol. 11, no. 8, 2022.

[87] V. L. Nguyen, D. H. Tran, H. Nguyen, and Y. M. Jang, "An experimental demonstration of mimo c-ook scheme based on deep learning for optical camera communication system," *Applied Sciences*, vol. 12, no. 14, 2022.

[88] O. I. Younus, N. Bani Hassan, Z. Ghassemlooy, P. A. Haigh, S. Zvanovec, L. N. Alves, and H. L. Minh, "Data rate enhancement in optical camera communications using an artificial neural network equaliser," *IEEE Access*, vol. 8, pp. 42656–42665, 2020.

[89] M. F. Ahmed, M. K. Hasan, M. Shahjalal, M. M. Alam, and Y. M. Jang, "Experimental demonstration of continuous sensor data monitoring using neural network-based optical camera communications," *IEEE Photonics Journal*, vol. 12, no. 5, pp. 1–11, 2020.

[90] Y.-S. Lin, C.-W. Chow, Y. Liu, Y.-H. Chang, K.-H. Lin, Y.-C. Wang, and Y.-Y. Chen, "Pam4 rolling-shutter demodulation using a pixel-per-symbol labeling neural network for optical camera communications," *Opt. Express*, vol. 29, pp. 31680–31688, Sep 2021.

[91] D.-C. Tsai, Y.-S. Lin, Y.-H. Chang, L.-S. Hsu, C.-W. Chow, Y. Liu, C.-H. Yeh, and K.-H. Lin, "Using pixel-per-bit neural network for two rolling shutter patterns decoding in optical camera communication (occ)," in *2021 30th Wireless and Optical Communications Conference (WOCC)*, pp. 102–105, 2021.

[92] D.-C. Tsai, Y.-H. Chang, Y. Liu, C.-W. Chow, Y.-S. Lin, and C.-H. Yeh, "Wide field-of-view (fov) light-diffusing fiber optical transmitter for rolling shutter based optical camera communication (occ)," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2022.

[93] D.-C. Tsai, Y.-H. Chang, C.-W. Chow, Y. Liu, C.-H. Yeh, C.-W. Peng, and L.-S. Hsu, "Optical camera communication (occ) using a laser-diode coupled optical-diffusing fiber (odf) and rolling shutter image sensor," *Opt. Express*, vol. 30, pp. 16069–16077, May 2022.

[94] B. Lin, Q. Guo, C. Lin, X. Tang, Z. Zhou, and Z. Ghassemlooy, "Experimental demonstration of an indoor positioning system based on artificial neural network," *Optical Engineering*, vol. 58, no. 1, p. 016104, 2019.

[95] S. R. Teli, S. Zvanovec, and Z. Ghassemlooy, "Performance evaluation of neural network assisted motion detection schemes implemented within indoor optical camera based communications," *Opt. Express*, vol. 27, pp. 24082–24092, Aug 2019.

[96] S. Jeong, J. Min, and Y. Park, "Indoor positioning using deep-learning-based pedestrian dead reckoning and optical camera communication," *IEEE Access*, vol. 9, pp. 133725–133734, 2021.

[97] G. Zhao, P. Du, D. Geng, A. Alphones, and C. Chen, "Enhancing localization accuracy of indoor occupancy tracking using optical camera communication and human pose estimation," in *2022 IEEE 14th International Conference on Advanced Infocomm Technology (ICAIT)*, pp. 42–48, 2022.

[98] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015.

[99] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018.

[100] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," *CoRR*, vol. abs/1612.08242, 2016.

[101] G. Jocher, A. Stoken, J. Borovec, NanoCode012, ChristopherSTAN, L. Changyu, Laughing, tkianai, A. Hogan, lorenzomammana, yxNONG, AlexWang1900, L. Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, F. Ingham, Frederik, Guilhen, Hatovix, J. Poznanski, J. Fang, L. Yu, changyu98, M. Wang, N. Gupta, O. Akhtar, PetrDvoracek, and P. Rai, *ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements*. Zenodo, Oct. 2020. `https://doi.org/10.5281/zenodo.4154370`.

[102] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," *CoRR*, vol. abs/1711.07064, 2017.

[103] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," *CoRR*, vol. abs/1908.03826, 2019.

[104] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," *CoRR*, vol. abs/1911.11907, 2019.

[105] A. Bochkovskiy, C. Wang, and H. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *CoRR*, vol. abs/2004.10934, 2020.

[106] L. Liu, R. Deng, and L.-K. Chen, "Spatial and time dispersions compensation with double-equalization for optical camera communications," *IEEE Photonics Technology Letters*, vol. 31, no. 21, pp. 1753–1756, 2019.

[107] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *CoRR*, vol. abs/1506.02025, 2015.

[108] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[109] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," *arXiv preprint arXiv:2106.11342*, 2021.

[110] Y. LeCun and Y. Bengio, *Convolutional Networks for Images, Speech, and Time Series*, p. 255–258. Cambridge, MA, USA: MIT Press, 1998.

[111] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[112] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014.

[113] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” 2014.

[114] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[115] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” 2016.

[116] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” 2016.

[117] B. Wu, A. Wan, X. Yue, P. Jin, S. Zhao, N. Golmant, A. Gholaminejad, J. Gonzalez, and K. Keutzer, “Shift: A zero flop, zero parameter alternative to spatial convolutions,” 2017.

[118] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for mobilenetv3,” 2019.

[119] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, “Designing network design spaces,” 2020.

[120] X. Mao, C. Shen, and Y. Yang, “Image restoration using convolutional autoencoders with symmetric skip connections,” *CoRR*, vol. abs/1606.08921, 2016.

[121] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *CoRR*, vol. abs/1501.00092, 2015.

[122] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2015.

[123] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “Deep convolutional autoencoder-based lossy image compression,” in *2018 Picture Coding Symposium (PCS)*, pp. 253–257, 2018.

[124] J. Williams, G. Carneiro, and D. Suter, “Region of interest autoencoders with an application to pedestrian detection,” in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, 2017.

[125] R. Siddalingappa and S. Kanagaraj, “Anomaly detection on medical images using autoencoder and convolutional neural network,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021.

[126] M. M. Badza and M. C. Barjaktarovic, “Segmentation of brain tumors from mri images using convolutional autoencoder,” *Applied Sciences*, vol. 11, no. 9, 2021.

[127] Sony Corporation, “IMX219PQH5-C Datasheet.” Accessed on: Feb. 10, 2022. [Online]. Available: `https://datasheetspdf.com/pdf/1404029/Sony/IMX219PQH5-C/1`, 2014.

[128] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, “Hyperband: A novel bandit-based approach to hyperparameter optimization,” *Journal of Machine Learning Research*, vol. 18, pp. 1–52, 04 2018.

Appendix A

Additional selected journal papers

Research Article
Vol. 27, No. 14 | 8 Jul 2019 | OPTICS EXPRESS 19150
Optics EXPRESS

Correlation-based receiver for optical camera communications

Cristo Jurado-Verdu,[*] **Vicente Matus, Jose Rabadan, Victor Guerra, and Rafael Perez-Jimenez**

IDeTIC, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[*]*cjurado@idetic.eu*

Abstract: In color-multiplexed optical camera communications (OCC) systems, data acquisition is restricted by the image processing algorithm capability for fast source recognition, region-of-interest (ROI) detection and tracking, packet synchronization within ROI, estimation of inter-channel interference and threshold computation. In this work, a novel modulation scheme for a practical RGB-LED-based OCC system is presented. The four above-described tasks are held simultaneously. Using confined spatial correlation of well-defined reference signals within the frame's color channels is possible to obtain a fully operating link with low computational complexity algorithms. Prior channel adaptation also grants a substantial increase in the attainable data rate, making the system more robust to interferences.

1. Introduction

Optical camera communications (OCC) have increased their relevance within the field of visible light communications (VLC) due to the widely-available cameras in a great number of devices (smartphones, tablets, surveillance systems, ...). Most of these cameras are based on rolling-shutter (RS) technique, which implements a row-by-row image scanning process [1]. Therefore, an LED light source, switching at a frequency higher than the shutter speed, will appear in the image as a series of dark and bright stripes, representing the binary data. Thus, this camera architecture provides to OCC a data rate much higher than its frame rate (fps) [2, 3]. OCC links need from several image processing stages to ensure a correct reception, such as detection and tracking of the light source (the region of interest, ROI), equalization of the optical power along the ROI, and threshold computation for demodulation. In ROI detection, blurring, dilating, and a variety of morphological operations have been proposed [4]. Other proposals [5] use anchor LEDs as spatial references for source detection, alignment, and spatial synchronization. However, this technique is not suitable for time synchronization in RS-based receivers.

For the non-uniformity of the received signal in the ROI, suggested solutions are based on searching the area in which this effect is partially mitigated [6], or by applying enhancing techniques to ameliorate it [7]. In the case of threshold computation, third order polynomial curve fitting (TOPF), iterative threshold (ITS), and quick adaptive scheme (QAS) have been evaluated in [8]. Furthermore, other techniques have been explored, such as entropy-based [9] and pixel-boundaries-based algorithms [10]. It is important to highlight that all the experiments above performed their threshold calculation frame by frame, with high computational resource consumption. Finally, the system has to spatially detect and synchronize the data packet within the ROI for correct data demodulation. In this case, two effects must be considered: the sampling frequency offset (SFO) of the camera, that varies slightly the sampling instants over time, and the blind period in which the signal cannot be acquired, either because the sensor is not being exposed (camera inter-frame processing), or because the signal does not fall within the ROI. Packet framing and repetition are used to address this issue [11]. Moreover, for RGB-based systems, the inter-channel cross-talk (ICCT), caused by the wavelength mismatch between the LEDs and the camera's Bayer filter, needs to be compensated as proposed in [12, 13], where a

#364046
Journal © 2019
https://doi.org/10.1364/OE.27.019150
Received 3 Apr 2019; revised 12 Jun 2019; accepted 12 Jun 2019; published 24 Jun 2019

training signal is employed.

In this paper, a correlation-based model is proposed for providing a unique solution to address all these issues. Using 2D-correlation processing over the images, the presented system is able to detect, track and identify the data source, even in case of spatial multiple access, with several sources included in the same image. Furthermore, using the correlation results, ICCT compensation and polynomial curve fitting for threshold computation can be obtained. Then, using the same correlation strategy, the system spatially synchronizes the data packet and provides the best-suited sampling spots for demodulation.

2. System description

The proposed OCC system is a broadcasting, RGB-based link with RS-camera as the receiver. The following subsections will describe in detail both transmitter and receiver structures.

2.1. Transmitter

Raw transmission data are partitioned on fixed-length packets. In order to keep data integrity, each packet is repeated during two frame times. Under this approach, data transmission is delimited in time within a super-frame structure. While the LED-lamp is sending data, a time window denoted as the Beacon-Only period, is reserved at the beginning of each super-frame for the transmission exclusively of N_b beacon packets. Additionally, when the emitter remains at an idle state, the beacons packets will be transmitted continuously. The functions of the beacon are: to ease the discovery stage, to uniquely identify each source (exploiting the spatial division multiplexing capability), and to help in the inter-channel cross-talk and threshold level computation. Figure 1(a) shows the beacon structure.

The length of the beacon, M, must be the same as the data packet. It consists of an N-length Hadamard sequence, corresponding to the number of wavelengths (R, G, and B in this case), followed by $M - N - 1$ single-color slots and a guard slot (depicted in black). This guard slot diminishes the effect of misidentifying two different sources. This configuration allows $N! \cdot (M - N - 1)^N$ different emitter identifiers within the image (where $N \geq (M - 1)$).

During the data transmission time, the packets being sent combine a synchronization signal that occupies exclusively one of the independent channels, and the OOK-modulated payload distributed over the other ones. This synchronization signal marks the start, end of the packet, and the optimal sampling instants (Fig. 1(b)).

2.2. Rolling shutter-based receiver

The receiving process can be split into three consecutive stages: discovery, training, and acquisition. The receiver will use two 2D-correlation templates: the beacon, and the synchronization templates, derived from the corresponding transmitter's signals (Fig. 1). They have a fixed width of w pixels, and a stripe height, $h = t_\text{chip}/t_\text{row}$. Where t_chip and t_row are the symbol and sensor's row delay times respectively. This h optimizes the correlation output, and increase the resilience to the broadening effect that the overlapping exposure of consecutive rows could generate in the stripe height. The total beacon height (in pixels), $H_\text{beacon} = h \cdot M$, must be less or equal to the half of the expected vertical height of the lamp's projection, H_lamp, within the frame, $H_\text{lamp}/2 \geq H_\text{beacon}$. The computation of H_lamp is discussed in [14]. This constraint ensures that a beacon could be recovered within the lamp's projection after the processing of two frames, and limits the final throughput, R_b (Eq. (1)).

$$Rb \leq \frac{H_\text{lamp}}{2} \cdot \frac{1}{h} \cdot \frac{fps}{2} \cdot N \tag{1}$$

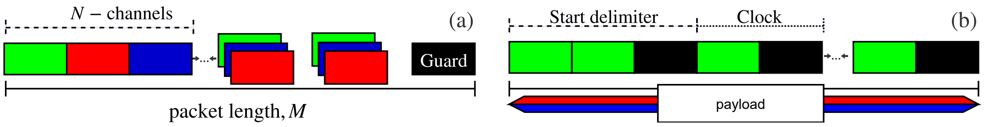


Fig. 1. Transmitter signals. (a) Beacon signal. (b) Data signal.

2.2.1. Discovery stage

In this stage, the receiver correlates a detection template that groups two beacon templates. If the maximum Pearson correlation coefficient, ρ, (Eq. (2)) exceeds the imposed detection threshold value, ρ_{th}, a source is considered as successfully detected.

$$\rho(x,y)=\frac{\sum_{x',y'}(T'(x',y')\cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'}T'(x',y')^2\cdot\sum_{x',y'}I'(x+x',y+y')^2}} \tag{2}$$

T and I are the template and image frames respectively, while x and y are the pixel coordinates. After detection, the receiver proceeds to the training stage.

2.2.2. Training stage

In this stage, the receiver locates the beacon template within the cropped detected ROI. Taking advantage of the beacon structure (independently switched channels), RGB to Bayer gains can be directly obtained from channel samples of N frame pictures.

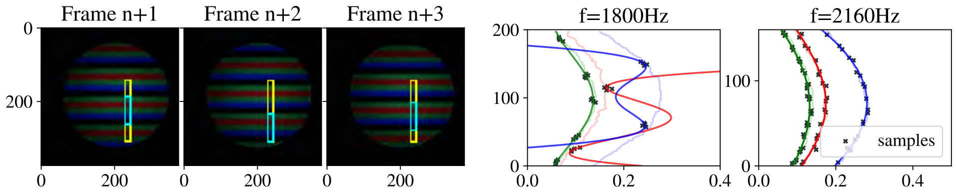


Fig. 2. (a) Inter-frame motion example. (b) Comparison of training samples (dark dots) and curve fitting for motionless $F_{\mathrm{tx}} = 1800$ Hz and $F_{\mathrm{tx}} = 2160$ Hz.

An example of the capture of 3 frames used for training is shown in Fig. 2(a). The yellow rectangle highlights the cropped ROI area, while the beacon is framed with cyan borders. As can be appreciated, the beacon is not fixed among frames but advances or falls behind its previous position. As has been stated in RS systems, there is an implicit relative motion of just a few rows per a large number of frames caused by the camera's SFO, v_{sfo}. This motion has been generally considered as an issue. Nevertheless, the proposed system benefits from this phenomenon since it allows to increase the spatial dispersion of the training samples between frames. Furthermore, to speed up this inter-frame motion a $v_{\mathrm{design}} = [N_{\mathrm{s}} \bmod M]\cdot h$ ($^{\mathrm{pixels}}/_{\mathrm{frame}}$), can be forced by selecting the length of the beacon template accordingly over the total number of available stripes in the frame, N_{s}. Taking this parameter into consideration, the total inter-frame motion is, $v_{\mathrm{inter}} = v_{\mathrm{design}} + v_{\mathrm{sfo}}$.

After processing L frames, the system performs a third order polynomial fitting to obtain the cross-talk matrix and the threshold for the entire ROI and proceeds to the next stage. Figure 2(b) depicts the collected samples from 5 consecutive frames and the output fitted curve when, v_{design} was set to zero (left graph). In that case, it can be observed that samples tend to form local groups (dark dots) instead of evenly distribute over the entire ROI, as it occurs when motion is forced (right graph). If a beacon template is not found, the receiver will restart.

2.2.3. Acquisition stage

In this stage, the receiver correlates both the beacon and the synchronization templates with the ROI. When a sync template is found, the receiver performs the enhancing exclusively over that area. This reduces the computational load needed for pre-processing the entire image. Finally, the binarization process and data assembly are performed. If any of the templates are not found during this stage, the receiver will restart the discovery stage.

3. Experiment and results

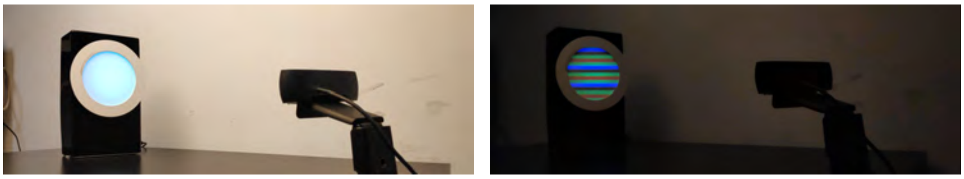


Fig. 3. Experiment setup.

As a proof of concept, a testbed was put together to measure the system performance as shown in Fig. 3. The transmitter is an RGB-LED lamp driven by an ST NUCLEO board. The signal was recorded using a Logitech C920 webcam. The videos were processed off-line using OpenCV libraries. In order to evaluate the operation of this proposal, it was carried out a series of experiments for different distances, d, and frequencies, f. The camera fps was set to 30 with full HD resolution (1080p). The exposure time was set to the minimum available, 300 µs. The measured row delay time was 31.4 µs. The camera's vertical Field of View (VFOV) was 43.3°. The ISO was set to 100, and the white balance correction to 6500K. The template's column width, w, was set to 15 pixels. Finally, the selected frequencies: 2160, 3240, 1800 and 2700 Hz (with stripes' height of 15, 10, 18 and 12 pixels respectively) can be grouped into a forced (first two) and non-forced motion sets. The physical height of the light source was 8 cm. For each pair (d, f), two different video recordings were made (10 minutes) while the source was transmitting beacons continuously and when it was sending pseudo-random packets of 10 bits. The last recording was performed three times, resulting in 1.25×10^5 received raw bits per experiment.

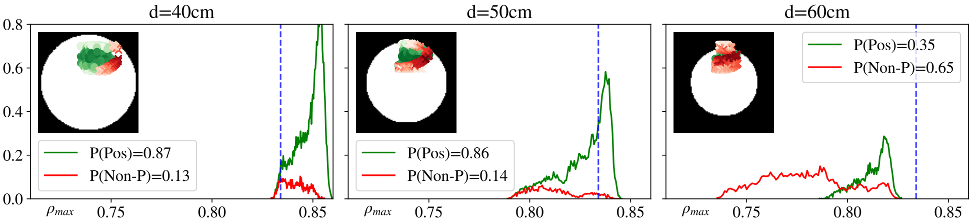


Fig. 4. Maximum Pearson correlation coefficient, ρ_{max}. $F_{\text{tx}} = 2160$ Hz.

The system's precision detecting legitimate sources with a certain degree of confidence was evaluated (positive case). To perform this evaluation, the maximum Pearson correlation coefficient between each frame and the template was obtained. Then, it was classified into a positive or non-positive sample collection. Samples are considered positive when the beacon template completely fits within the source's projection. Figure 4 depicts the detected position associated to the maximum correlation (green for the positive cases), and the histogram of the correlation coefficient for both positive and non-positive cases weighed by their corresponding *a priori* probabilities. It can be highlighted that as distance increases, both the source's projected area and the number of positive samples diminish. Moreover, the average correlation value also decreases. This occurs because the template is forced to be detected at the lamp's center, where it can fit entirely. Nonetheless, at this position, the pulse broadening effect is higher, leading to a lower correlation output value.

Based on these samples, a detection threshold is selected. Lowering the threshold level will increase the number of false positives (detecting a non-positive sample as a source). Thus, the receiver's precision will drop. Otherwise, if the threshold raises, the miss rate would rapidly grow. This has implications on the average source detection time that can be expressed in terms of the number of frames needed for source detection, $N_{\text{detection}}$. Figure 5(a) plots the precision/recall curve of the system. The dashed black line sets the minimum precision set as the design criteria of 0.9 (90 percent of the detected sources will belong to the true positive case). Figure 5(b) plots the average number of frames prior detection, $E[N_{\text{detection}}]$, against the detection threshold. In the extreme case in which the ROI height is comparable to the source projection over the image, there would be a higher probability of missing the detection. If the inter-frame motion were low, the probability of detecting the source on the next frame would also be small. Thus, there is an inverse relationship between the inter-frame motion and the variance of $N_{\text{detection}}$ for those cases. For instance, if the receiver captures a beacon halfway through its transmission, it will have to wait a long time for detection due to the scarce inter-frame motion.

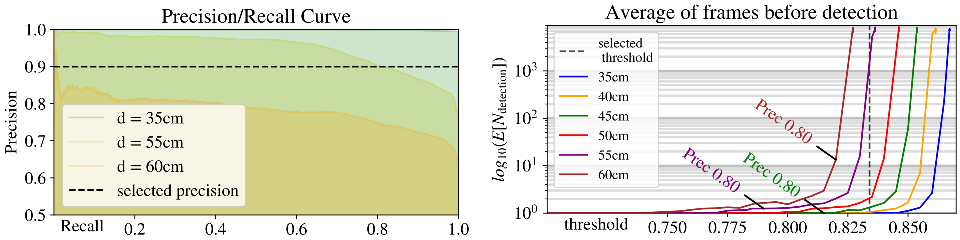


Fig. 5. (a) Precision-Recall curve. (b) Average of $N_{\text{detection}}$. $F_{\text{tx}} = 2160\,\text{Hz}$.

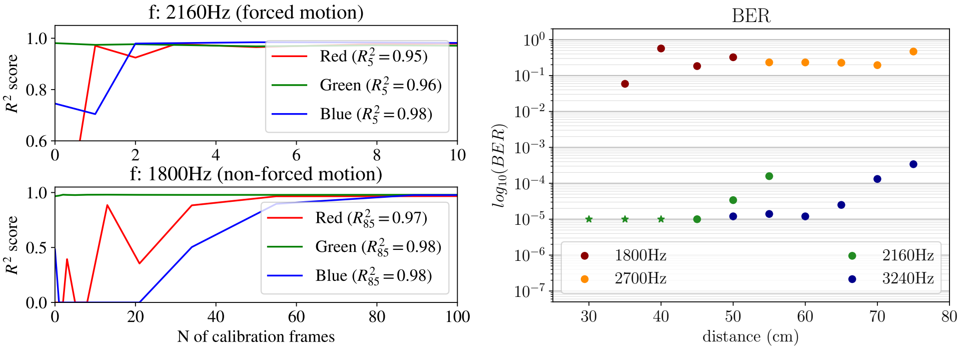


Fig. 6. (a) R^2 determination coefficient versus number of frames for training. (b) Bit error rate versus distance (star points indicate that no errors were detected under the experiment conditions).

Then, to evaluate the performance of the training, the R^2 determination coefficient is used. For both frequency sets, a set of samples were collected through N frames, fitted and compared with independently captured images from each channel. Figure 2(b) represents the third polynomial fit against the real curve (lighter lines), obtained with independent image captures. As can be seen in Fig. 6(a), the non-forced motion frequency set needs more frames for training due to sampling clustering, to obtain an optimal R^2 determination coefficient of the fitting.

Finally, Bit Error Rate (BER) performance is presented in Fig. 6(b). It was evaluated using 0.834 as the detection threshold (system precision of 0.9), and four calibration frames. It can be seen that the BER decrements with the distance. As has been mentioned, this is related to the rising of the stripe broadening effect, which has a harmful effect on calibration, since there is a higher probability to obtain color-mixed samples.

4. Conclusions

In this work, an experimental evaluation of an RGB LED-based OCC system is presented. It uses the green channel for data synchronization, while the red and blue ones carry OOK-modulated data. Furthermore, a beacon-based detection scheme is proposed and evaluated. The processing algorithms for ROI detection, source identification, training, and packet synchronization are combined into a single correlation-based procedure. This technique finds the best ROI in terms of the least pulse broadening (inter-symbol overlapping), improving the BER performance. It also carries out the ICCT mitigation and enhancing, only within the data region, reducing the computational load. Experimental results demonstrated that the proposed system is able to achieve 300 bps (over a transmission span up to 0.7 m, with a constant BER lower than 1×10^{-4}. However, higher data rates or higher distances could be achieved by increasing the physical size of the lamp and the framing structure or using spatially-multiplexed sources.

Funding

Spanish Research Administration (MINECO project: OSCAR, ref.:TEC 2017-84065-C3-1-R).

References

1. H. Aoyama and M. Oshima, “Line scan sampling for visible light communication: Theory and practice,” in *IEEE International Conference on Communications (ICC),* (IEEE, 2015), pp. 5060–5065.
2. N. T. Le, M. A. Hossain, and Y. M. Jang, “A survey of design and implementation for optical camera communication,” Signal Process. Image Commun. **53**, 95–109 (2017).
3. Z. Ghassemlooy, P. Luo, and S. Zvanovec, “Optical Camera Communications,” in *Optical Wireless Communications: An Emerging Technology,* M. Uysal, C. Capsoni, Z. Ghassemlooy, A. Boucouvalas, and E. Udvary, eds. (Springer, 2016), pp. 547–568.
4. J.-W. Lee, S.-J. Kim, and S.-K. Han, “Multi-Level Optical Signal Reception by Blur Curved Approximation for Optical Camera Communication,” in *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC),* (2017).
5. A. D. Griffiths, J. Herrnsdorf, M. J. Strain, and M. D. Dawson, “Scalable visible light communications with a micro-LED array projector and high-speed smartphone camera,” Opt. Express **27**, 15585 (2019).
6. J. He, Z. Jiang, J. Shi, Y. Zhou, and J. He, “A Novel Column Matrix Selection Scheme for VLC System With Mobile Phone Camera,” IEEE Photonics Technol. Lett. **31**, 149–152 (2019).
7. W. Guan, Y. Wu, C. Xie, L. Fang, X. Liu, and Y. Chen, “Performance analysis and enhancement for visible light communication using CMOS sensors,” Opt. Commun. **410**, 531–551 (2018).
8. K. Liang, C.-W. Hsu, C.-W. Chow, C.-Y. Chen, Y. Liu, S.-H. Chen, and H.-Y. Chen, “Comparison of thresholding schemes for visible light communication using mobile-phone image sensor,” Opt. Express **24**, 1973 (2016).
9. K. Liang, C.-W. Chow, Y. Liu, and C.-H. Yeh, “Thresholding schemes for visible light communications with CMOS camera using entropy-based algorithms,” Opt. Express **24**, 25641–25646 (2016).
10. Z. Zhang, T. Zhang, J. Zhou, Y. Lu, and Y. Qiao, “Thresholding Scheme Based on Boundary Pixels of Stripes for Visible Light Communication With Mobile-Phone Camera,” IEEE Access **6**, 53053–53061 (2018).
11. T. Nguyen, C. H. Hong, N. T. Le, and Y. M. Jang, “High-speed asynchronous Optical Camera Communication using LED and rolling shutter camera,” in *International Conference on Ubiquitous and Future Networks, ICUFN,* (2015).
12. K. Liang, C.-W. Chow, and Y. Liu, “RGB visible light communication using mobile-phone camera and multi-input multi-output,” Opt. Express **24**, 9383–9388 (2016).
13. P. Luo, M. Zhang, Z. Ghassemlooy, H. Le Minh, H. M. Tsai, X. Tang, L. C. Png, and D. Han, “Experimental Demonstration of RGB LED-Based Optical Camera Communications,” IEEE Photonics J. **7**, 1–12 (2015).
14. P. Chavez-Burbano, V. Guerra, J. Rabadan, D. Rodriguez-Esparragon, and R. Perez-Jimenez, “Experimental characterization of close-emitter interference in an optical camera communication system,” Sensors **17**, 1561 (2017).

Appendix B

Additional selected conference proceedings

Application of Optical Camera Communication to Microalgae Production Plants

Cristo Jurado-Verdu, Victor Guerra, Vicente Matus,
Jose Rabadan and Rafael Perez-Jimenez
*Institute for Technological Development
and Innovation in Communications, IDeTIC
Universidad de Las Palmas de Gran Canaria, ULPGC
Las Palmas de Gran Canaria,Spain
{cjurado, vguerra, vmatus, jrabadan, rperez }@idetic.eu*

Juan Luis Gomez-Pinchetti and Carlos Almeida
*Spanish Bank of Algae
Universidad de Las Palmas de Gran Canaria, ULPGC
Las Palmas de Gran Canaria, Spain
{juan.gomez, carlos.almeida}@ulpgc.es*

***Abstract*—Optical Camera Communication (OCC) Systems have a potential application for monitoring flat-panel photobioreactors in microalgae production plants. On the transmitter side, the design of pulsed optical signals in terms of light intensity, spectral composition, and temporal variation is focused on optimizing the photosynthetic efficiency of the culture. At the reception, cameras are utilized for simultaneous online monitoring and in-situ sensing of the bioreactors culture. In this work, the analysis of the node arrangement in the plant has been carried out considering node visibility, data rate, and effective space utilization. Finally, a real experiment was performed to validate the application's feasibility.**

***Index Terms*—Optical Camera Communications, Microalgae culture, Light flashing, Light Management, Camera sensing.**

I. INTRODUCTION

Microalgae culture has gained significant momentum during the last decade, triggered by the necessity of developing new and sustainable resources. They have become a promising alternative source for biofuels and biogas production, human and animal nutrition, cosmetics, and bioactive supply for nutraceutical and pharmaceutical applications.

Microalgae biomass production is currently carried out using both open ponds (raceways, e.g.) and closed photobioreactors [1]. The last ones are preferable at laboratory and pilot-plant scales since with the appropriate design, they can optimize growth conditions (nutrient levels, carbon concentration, temperature, acidity, etcetera). Among these parameters, light radiant energy is a capital factor that affects the photosynthetic efficiency, and therefore the overall productivity [2].

Generally, microalgae production plants are designed to take advantage of the sun as primary light source due to cost optimization. However, the sun's irradiance depends on several factors, such as weather conditions, latitude, and day time. Furthermore, these open-air plants need vast extensions to be profitable since effective biomass production depends on the direct-exposed surface. Nonetheless, artificial lighting can provide advantages in terms of photosynthetic efficiency (custom spectrum and intensity profiles) and the tight control they offer of microalgal biochemistry and growth, increasing the reliability of industrial processes. Besides, using this type of lighting, biomass production depends no longer on the plant's exposed surface, but on its equivalent volume. Although optimizing light quality could reduce the energy consumption considerably, another approach concerns the way light is delivered to the culture: flashing (pulsed) light instead of continuous illumination.

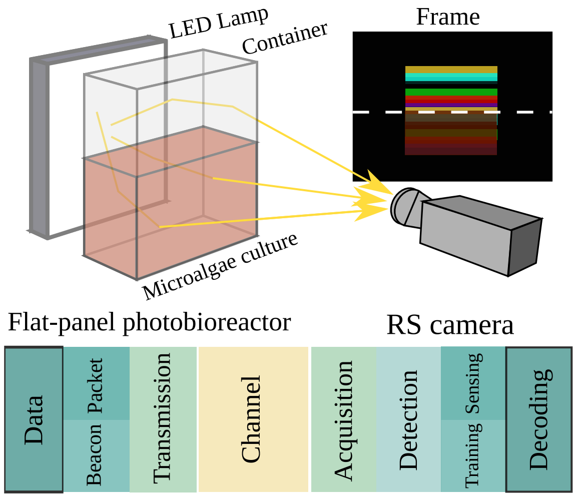


Fig. 1. Proposed architecture based on flat-panel photobioreactors.

In this context, pulsed illumination could serve as a communication link, enabling the re-utilization of light sources as effective Visible Light Communication (VLC) transmitters. The low bandwidth transmission channel they provide (attaining the stated frequency restrictions) could be used for online monitoring of the culture's state conditions via deployable sensors within the photobioreactor module.

Nevertheless, there are still a few parameters that either are complex to be measured *in situ* or imply invasive methods. These parameters are microalgae biomass concentration and its growth phase. In [3], [4] different approaches based on the digital processing analysis of red-green-blue (RGB) images for low cost, fast and accurate quantification biomass concentration are proposed and experimentally validated. Therefore, cameras could be used as sensing devices of these parameters.

On the other hand, the use of cameras as communication receivers for VLC links is currently an interest research area. Furthermore, there is a current revision of the standard specification by the working group IEEE 802.15.7r1 [5] planning to integrate this new strategy known as Optical Camera Communications (OCC) [6]. These devices are bandwidth limited

---

This work is funded in part by the Canary Islands Regional Government (Project ATICCuA ProID2017010053).

978-1-7281-6743-5/20/$31.00 ©2020 IEEE

compared to traditional photodetectors (PIN and Avalanche Photodiodes). However, due to the use of image-forming optics, cameras provide a inherent spatial multiplexing capabilities, which can be easily exploited for simultaneous monitoring within a microalgae production plant.

This work proposes the use of low-cost cameras for both remotely sensing microalgae culture parameters and for establishing a direct communication link with flat-panel photobioreactors. A comprehensive analysis based on geometrical constraints is carried out, providing results such as optimal camera positioning respect to a custom-defined metric that relates aggregate data rate and effective space exploitation. Furthermore, intrinsic algae-based restrictions are also analyzed and discussed.

This work presents part of the results obtained in the ATTICuA Project; a cooperation research carried out by the Spanish National Algae Bank (BEA) and the Photonics and Communications Division of the Institute for Technology Development and Innovation in Communications (IDeTIC). This project addresses the development of prototypes based on the Visible Light Communications (VLC), and more specifically those used in underwater communication systems (Underwater Wireless Optical Communications, UWOC), for application in microalgae culture systems. Precisely, its main objective consists of the design of an LED-based dual-use system, which provides configurable lighting for the culture and production of microalgae and cyanobacteria, and optical wireless communication capabilities for optical underwater channel characterization.

The remaining of the paper is structured as follows. Section II introduces the proposed OCC-based architecture in a bottom-up manner. Moreover, three different node arrangements are analyzed and compared in Section III. This section includes a real-world experiment that validates the feasibility of this proposal in terms of the received spectral signature. Finally, some conclusions are extracted in Section IV.

II. PROPOSED ARCHITECTURE

The proposed architecture consists of an LED flat-panel photobioreactor facing a rolling-shutter camera, as can be seen in Fig. 1. In this system, the transmitter sends two different signals: a well-know beacon signal used for detection and identification, and data packets composed of manchester encoded pulses. At the receiver side, the beacon signal is used to estimate the channel response to train the system before the decoding phase; and to estimate relative parameters of biomass concentration and growth phase of the microalgae species.

A. Transmitter

The competitiveness of artificial light-driven microalgal production requires improvements in photon harvesting and conversion efficiency of light sources [2]. Hence it is crucial to optimize light quality and delivery. In terms of light energy, the photosynthesis rate is directly related to the irradiance power, excessive or insufficient incident light constraints optimal performance and can induce the photo-inhibition and photo-oxidation of the cells, eventually attaining photo-damage and even leading to culture death [7]. In terms of light spectra, the radiant energy absorbed by microalgae highly depends on the chemical nature of their native pigments, which have specific absorption bands in the visible and near-infrared spectrum. Thus, better energy usage can be achieved by adjusting the light source's emission spectrum [8]. Finally, recent experimental studies have shown that combining short and intense light flashes with extended dark periods, instead of continuous illumination, increases the growth efficiency as discussed in [9], [10]. The frequency and duty cycle of the light pulses is particular to the microalgae species and also to the expected biomass product result (lipids, carotenoids, etcetera), and may vary between a few Hz up to tens of kHz. In short, the transmitting source's design must take into account all these restrictions concerning the selected strain.

B. Channel

Light attenuation in flat-panel photobioreactors with microalgae suspension is affected by different phenomena: the absorption capacity of the culture, the shading effect between cells, and light scattering. The Lambert-Beer law can be used to describe the attenuation of light due to absorption by the biomass concentration (Eq. 1). This equation states that the attenuation of light over distance is proportional to the light intensity, where is the volumetric absorption coefficient. The latter is the product of the specific light absorption coefficient, $\alpha_{x,}$, and the biomass concentration (ρ). The integration of Eq. 1 over the light path, and taking into account the wavelength dependency, results in Eq. 2.

$$\frac{dI_\lambda(x)}{d(x)} = -C_{x,\lambda} \cdot I = -\alpha_{x,\lambda} \cdot \rho \cdot I \tag{1}$$

$$I(x) = \sum_{\lambda=800}^{\lambda=300} I_\lambda(0) \cdot e^{-\alpha_{z,\lambda} \cdot \rho \cdot x} \tag{2}$$

However, the Lambert-Beer Law does not account for scattering. Instead, Schuster's law can be used in those cases where the light scattering must be considered [11]. In this architecture, the scattering effect is a valuable resource for communications, because it distributes the light within the reactor up to the frontal panel of the container.

C. OCC receiver

The image acquisition operation mode of the cameras inherently limits the available bandwidth for communications. In global-shutter cameras, the whole image sensor is exposed at the same instant; thus, the data rate is constrained to their frame rate, restraining the switching frequency of the light source considerably. On the other hand, rolling-shutter cameras scan the scene sequentially row by row of pixels (in the vertical dimension of the image), allowing to capture different light states (intensity, color variations) within the lamp's source projection in the frame [12]. In this case, the theoretical bandwidth limit is imposed by the row shift time, t_{row}, which is the fixed duration between the start of a row scan and its consecutive, also known as sampling time. Also, these rows are not disjointly exposed, but in an overlapped manner. This overlap depends on exposure time, t_{exp}, which is the span time that each row is integrating light. The effect of this overlap can be significantly simplified (for a uniform light source) as the product of a weighed moving-average filter, which further restricts the available bandwidth (cutoff frequency - Eq. 3) [13]. Therefore, it is desirable to

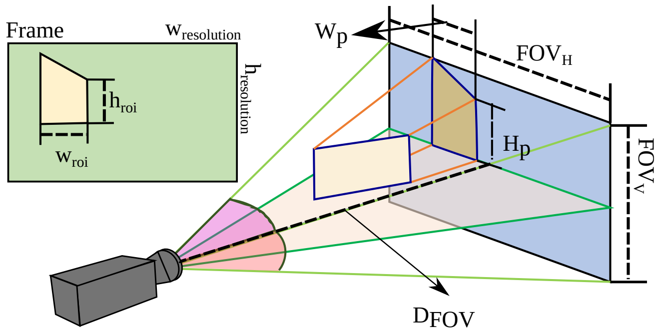


Fig. 2. Camera and geometrical parameters related to the node’s data rate.

minimize exposure time from a bandwidth viewpoint, but taking into account its negative impact on the Signal-to-Noise Ratio (SNR).

$$f_{\text{cutoff}}^{3dB} = \frac{1}{2 \cdot t_{\text{exp}}} \tag{3}$$

D. Node’s data rate

The achievable data rate of lamp source depends on the total number of samples (in the vertical dimension) of the lamp projection within the frame, in other words, the total number of rows of the image sensor that are affected by the projection of the light source [14]. Consequently, it is necessary to take into account the geometrical configuration of the scenario, which implies the relative position and facing direction between the node and the camera. Equations 4-5 relate the projection dimensions (in meters) of the lamp source over the scene plane, H_{pr}, W_{pr} with its corresponding pixels dimensions, h_{roi}, w_{roi}. Fig. 2 shows all the geometry and camera parameters involved in a generalized scenario.

$$h_{roi} = H_p \cdot \frac{h_{\text{resolution}}}{2 \cdot \tan(\text{AoV}_{\text{V}}/2) \cdot \text{D}_{\text{FOV}}} \quad (pixels) \tag{4}$$

$$w_{roi} = W_p \cdot \frac{w_{\text{resolution}}}{2 \cdot \tan(\text{AoV}_{\text{H}}/2) \cdot \text{D}_{\text{FOV}}} \quad (pixels) \tag{5}$$

Where $h_{resolution}$ and $w_{resolution}$ are the pixel dimensions of the image, AoV_h and AoV_w, the horizontal and vertical Angle Of View (in degrees) of the camera, and D_{FOV}, the distance (in meters) from the camera lens to the scene plane.

Relating Eq. 3 with the Nyquist criterion, ($f_{\text{transmitter}} \geq 2 \cdot f_{\text{cutoff}}$) the minimum number of vertical samples of the signal required per symbol can be computed using Eq. 6.

$$h_{symbol} = \frac{f_{\text{sampling}}}{f_{\text{transmitter}}} = \frac{t_{exp}}{t_{\text{shift}}} \quad \text{(samples)} \tag{6}$$

In addition, it is important to highlight that the communication takes place in a windowed manner. In other words, the camera can only detect the light source transitions within its projection in the image, being blind to changes that occur when the sensor is scanning another part of the image. To overcome this issue, the transmitter can send repeated data packets (during the acquisition time of two frames), to assure the correct reception. Also, at the receiver side, it is necessary to detect a minimum vertical source projection, h_p that can fit the number of samples of two data packets, as it is detailed in previous work [15]. Hence, the overall data transmission rate (in bauds) could be obtained from Eq. 7, attaining to the restriction imposed by Eq. 8.

$$R_b = P_{symbols} \cdot \frac{\text{fps}}{2} \tag{7}$$

$$h_{\text{packet}} \geq 2 \cdot P_{symbols} \cdot h_{symbol} \tag{8}$$

On the other hand, the node width in pixels (number of columns) plays an important role in communications. Pixels located in the same row integrates light at the same time, so, while being affected by the same source, they could be used to filter out signal noise, thus strengthening the signal to noise ratio. Besides, wider widths aid the source detection and tracking routines, and eases the decoding process considerably. As a consequence, minimum pixel width also becomes a constraint design parameter of the communication link.

E. Nodes’ distribution

In the proposed architecture is required to carefully analyze the distribution of the nodes across the scenario to establish a minimum data rate to each photobioreactor using the multi-spatial capabilities of a single camera receiver. This distributions must not just be optimized to communication specifications: number of photobioreactors being monitored, N, and minimum assured data rate, R_b^{min}; but also to a specific parameter of the production plant: the space utilization ratio, SUR (Eq. 10). This term relates the total equivalent volume of the containers bound to biomass harvesting, $\text{V}_{\text{container}}$, and the overall space used to store the photobioreactors while providing the monitoring link, V_{total}. It is decisive to minimize this term to design a viable and competitive solution that meets the associated production costs.

As a design criteria, a figure of merit, F, that takes into account this parameters is defined in Eq. 9.

$$F = N \cdot R_b^{min} \cdot \text{SUR} \tag{9}$$

$$\text{where}: \quad \text{SUR} = \frac{\text{V}_{\text{container}}}{\text{V}_{\text{total}}} \tag{10}$$

Therefore lower F imply better arrangements of the nodes within the scenario.

III. PRELIMINARY RESULTS

Two different analysis has been carried out to evaluate the feasibility of this proposal as preliminary results. The first discussion is focused on the analysis of F for three different node’s arrangements. The second experiment evaluates the signal reception of a transmitted beacon signal for low, medium, and high biomass concentrations of the microalgae’s species selected.

A. Node arrangement

Figure 3 shows three different cases analyzed in this discussion. All cases begin on a particular scenario, in which there are 12 photobioreactors placed on a shelf. The dimensions of the containers are 150x50x9 cm: the dimensions of the shelf, 300x300x10 cm.

Also, additional restrictions related to the communication link must be met. Firstly, it is established that the projection of the upper photobioreactors will fill the upper half of the

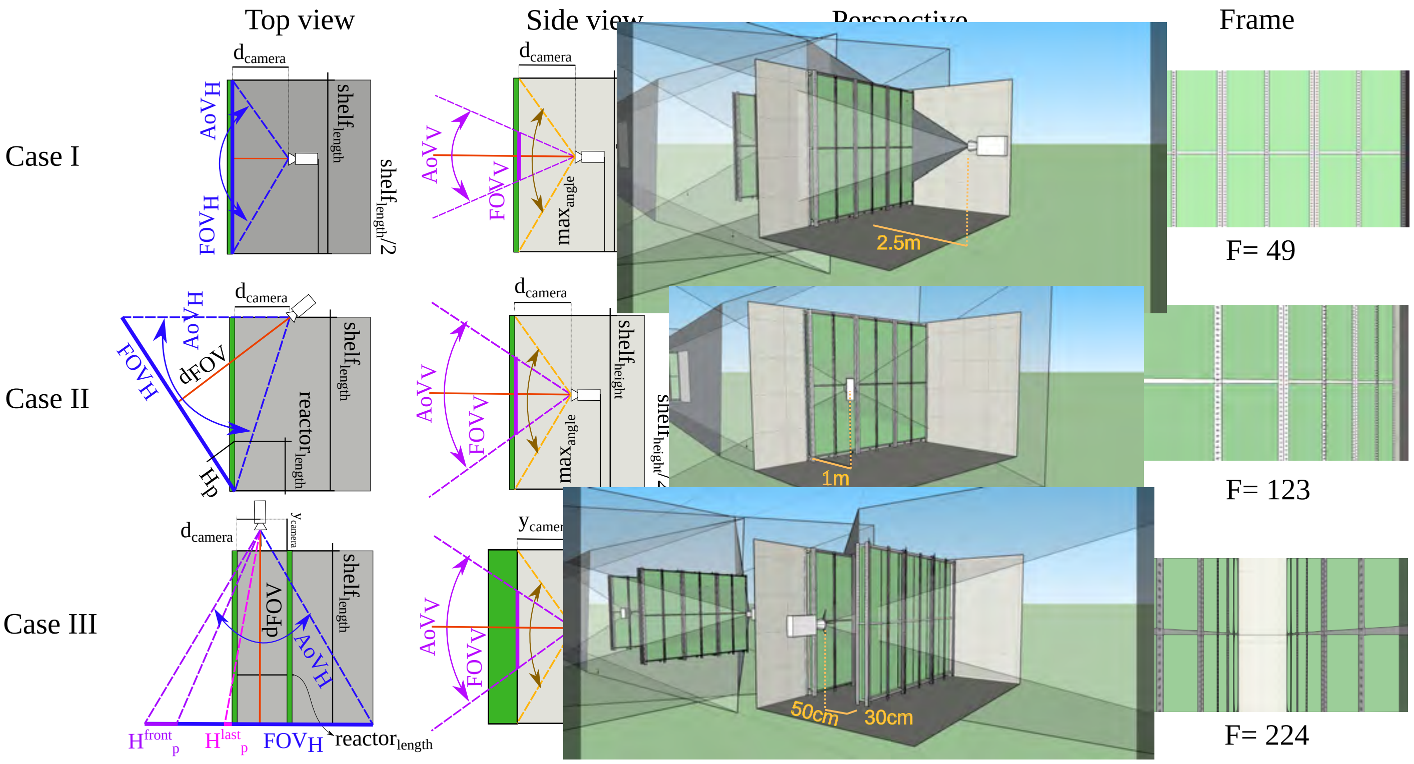


Fig. 3. Analysis of different node's distribution.

image and the lower ones, the lower half. Consequently, the camera should always point to the vertical center of the shelf. In this way, the transfer rate is assured to be equal to all bioreactors, and the number of vertical pixels is equal to the half of the vertical resolution of the image (Eq. 11). Secondly, the minimum pixel width of any reactor must be 30 pixels (Eq. 12). In addition, both vertical and horizontal Field of View are related by the sensor aspect ratio, $\text{aspect}_{\text{ratio}}$, thus altering one of them will influence the other. The image sensor Sony IMX219 specifications are used as a reference. The sensor resolution (in Megapixels) is fixed to 8 Mpx; the image format is 3280×2464 pixels, the exposure time of the camera is set to 1/10000 seconds (100 μs), its frame rate to 10 fps and the row shift time, experimentally measured, is 18.904 μs. Hence, the maximum baud rate achievable using Eq. 7 is 1108 (Bd/reactor) and the aggregate baud rate is 13296 Bd.

$$h_{roi} = H_p \cdot \frac{h_{\text{resolution}}}{\text{FOV}_V} \leq h_{\text{resolution}}/2 \quad \text{(pixels)} \tag{11}$$

$$w_{roi} = W_p \cdot \frac{w_{\text{resolution}}}{\text{FOV}_H} \geq 30 \quad \text{(pixels)} \tag{12}$$

$$\text{where :} \quad \text{AoV}_H = \text{AoV}_V \cdot \text{aspect}_{\text{ratio}}$$

a) Case I: The camera is located vertically and horizontally, pointing to the middle of the shelf that holds the containers. In this case, to attain to the previous restrictions, the horizontal Field of View, FOV_H, of the camera must cover at least the entire shelf length, $\text{shelf}_{\text{length}}$ (Eq. 13), to assure that all bioreactors are within the visibility range.

The other restriction is that the vertical Field of View of the camera, FOV_H, must be lower or equal than the shelf's height. Otherwise, the vertical projection of the photobioreactors will decrease, because the floor, and the ceiling will be visible within the image.

This condition is less restricted than the first one. The reason is that it is not necessary to view the entire photobioreactor in order to obtain a source projection that fills the frame, as it is shown in *Frame* section in Fig. 3.

$$FOV_H = 2 \cdot \tan\left(\frac{\text{AoV}_H}{2}\right) \cdot d_{\text{camera}} \geq \text{shelf}_{\text{length}} \tag{13}$$

$$FOV_V = 2 \cdot \tan\left(\frac{\text{AoV}_V}{2}\right) \cdot d_{\text{camera}} \leq \text{shelf}_{\text{height}} \tag{14}$$

In this case, the first strategy to resolve the equations is to fix the AoV_V to its maximum value (Eq. 14). Thus FoV_H will increase, reducing the camera distance considerably. However, that increment will intensify the distortion effects introduced by the lens, making the vertical lines bend in the image (barrel distortion). Besides, image processing techniques can not mitigate this effect because that will introduce a non-linear distorsion in the received signal. To avoid these unwanted effects, AoV_H was restricted to 70°. Using equations 13 and 14 the camera distance, d_{camera}, was obtained, and the resulting F for this configuration is shown in Fig. 3.

b) Case II: The camera is translated on the y-axis. This translation has the advantage that the camera distance is considerably reduced, so the resulting F is lower than the previous case. However, as can be seen in Fig. 3, the horizontal projection of the reactors is shrinking relative to its distance to the camera. Therefore, the Eq. 12 acquires and important part in the consideration restrictions. In this case, the FOV_H changes slightly, and it is defined in Eq. 15.

$$FOV_H = 2\tan\left(\frac{\mathrm{AoV}_H}{2}\right)\cdot d_{\mathrm{FOV}_H} \tag{15}$$

$$\text{where}: \quad d_{FOV_H} = \frac{d_{\mathrm{camera}}\cdot\cos(\mathrm{AoV}_H/2)}{\cos(\mathrm{AoV}_H)}$$

$$AoV_H = \left(\frac{\mathrm{shelf}_{\mathrm{length}}}{d_{\mathrm{camera}}}\right)$$

In this case, utilizing the maximum $AoV_H = 70^\circ$ restriction, w_{roi} considerably exceeds the restriction imposed in 12, thus this configuration is viable.

c) Case III: Another shelf has been added. Both shelves are facing each other, and the camera is now pointing at the center of the corridor. In this configuration, part of the image corresponds to the frontal wall, reducing the available left space for every photobioreactor. In this case, as in the previous case, it is especially important to analyze the horizontal projection of the reactors. However, two exceptional cases can be highlighted. If AoV_H is too wide, the horizontal projection of the last reactor can not meet the minimum constraint. If AoV_H is too narrow, the same will happens but with the first reactor. This is because the first reactor is partially visible in this configuration. Otherwise, AoV_H must be greater or equal to 180° degrees, leading to image distortions.

$$h^i_{roi} = H^i_p \cdot \frac{h_{\mathrm{resolution}}}{\mathrm{FoV}_H} \geq 30 \quad (\text{pixels}) \tag{16}$$

$$H^{last}_p = \frac{d_{\mathrm{camera}}\cdot \mathrm{reactor}_{\mathrm{length}}}{\mathrm{shelf}_{\mathrm{length}} + y_{camera} - \mathrm{reactor}_{\mathrm{length}}} \tag{17}$$

$$\begin{aligned} H^{first}_p &= (\mathrm{shelf}_{\mathrm{length}} + y_{camera}) \quad \cdot \quad \Rightarrow \\ &\Rightarrow \left[\tan\left(\frac{\mathrm{AoV}_H}{2}\right) - \frac{d_{\mathrm{camera}}}{(\mathrm{reactor}_{\mathrm{length}} + y_{camera})}\right] \end{aligned} \tag{18}$$

This case has the higher F number (224), and, althought further photobioreactors will have a shorter width projection, it is enough to stablish a communication link.

B. Transmission experiment

This experiment was carried out at the facilities of the National Spanish Bank of Algae (BEA, by its Spanish acronym). In this experiment, a beacon signal was transmitted through a medium with different biomass concentrations of the Rhodosorus marinus species: high, medium, and low. This signal consists of five pulses (one for each available LED that integrates the lamp), followed by a dark period. The key parameters of the experiment are shown in Table I.

TABLE I
Key parameters of the experiment.

Module	**Characteristics**
	Photobioreactor
LED Lamp	Eglo Tunable White - RGB connect - 1 White cold LED (6500K) - 1 White warm LED (2700K) - 1 RBG LED* *(Wavelenghts [nm]: 630 (Red) 530 (Green), 475 (Blue))
Container	Square flat panel (self made) - dimensions [cm] : 50x50x9
Signal	(Beacon) Pulsed signal T_{chip} [s] = 1/8400
	Channel
Microalgae	- species: Rhodosorus marinus - genus: Rhodosorus - concentration: HIGH, MEDIUM, LOW
Distance [m]	2
	Receiver
Camera	- sensor: Sony IMX586 - aperture lens: f/1.75 - focal length (equivalent) [mm]: 26mm - video resolution [px]: 1920x1080 - exposure time [s]: 1/10000 - ISO: 450 - frame rate [fps]: 30

The results are presented in Fig. 4. This Figure displays a snapshot taken for three concentration density: low (a), medium (b), and high (c). The white rectangle enclosed the detected beacon signal. The graph shown below the pictures represents the red, green, and blue pixel values from one column located within this rectangle. Also, picture (d) represents the reference case where the signal is extracted without being affected by the microalgae channel. The signal recovered corresponds to the sequentially pulsed LEDs: white cold, white warm, dark guard (no pulse), green, red, and blue. It can be observed that this species produces significant attenuation in the blue and green portions of the spectrum (attaining to the Bayer filter's spectrum response of the camera). Despite that the green, blue, and red pulses can not be distinguished in picture (c) (due to low optical transmitting power), it still possible to differentiate the white cold and white warm spectral signatures. Therefore, those LEDs are viable for communication purposes.

IV. CONCLUSIONS

In this work, the use of OCC has been proposed as a suitable communication technology for microalgae production plants based on artificial lighting. Concretely, flat-panel based plants have been considered as a preliminary topology.

Several configurations have been tested, and a custom merit figure relating the aggregate data rate and volume yield were evaluated. It was observed that the best configuration for flat-panel monitoring corresponded to the two-sided corridor monitoring position (F=224). This topology improves the number of simultaneously monitored photobioreactors significantly. Also, taking into account that these facilities produce high-latency low-rate data, it is more critical to maximize multi-user capability instead of the achievable data rate.

A real experiment involving a culture of Rhodosorus marinus was carried out for low, medium, and high concentrations. It was observed that blue and green light presented high absorption profiles, limiting the use of these wavelengths for data transmission. However, this characteristic provides significant information for remotely sensing algae concentration. Furthermore, these results suggest that the type of microalgae must be taken into account when designing this type of OCC link.

It has been demonstrated that microalgae production plants are a potential use case for OCC, in which a significant amount of devices could be easily monitored using a cost-effective deployment.

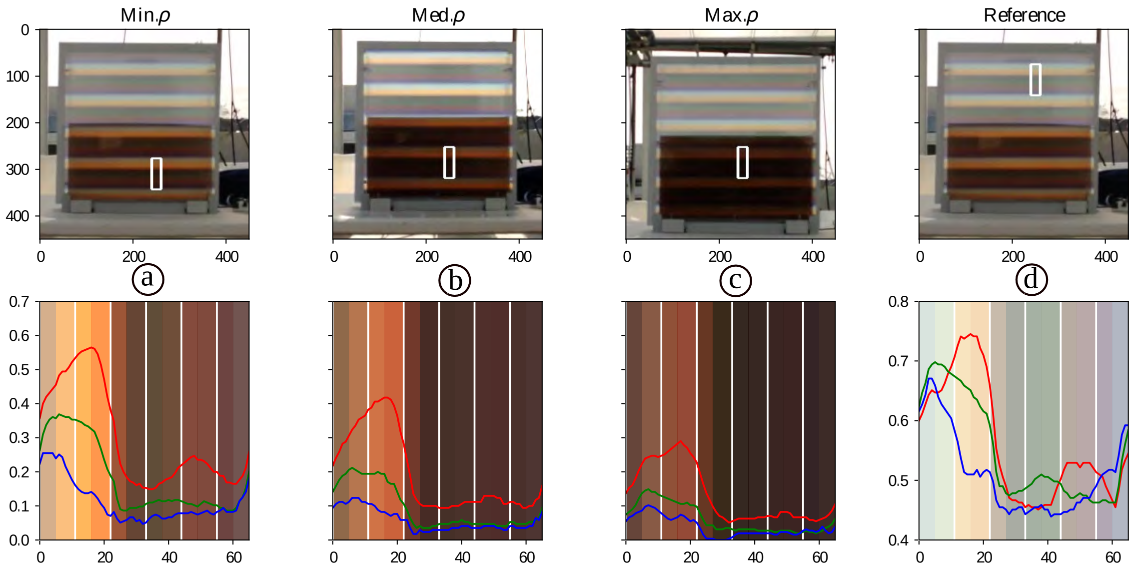


Fig. 4. Experimental transmission of a well-known beacon signal.

REFERENCES

[1] F. G. Acién Fernández, J. M. Fernández Sevilla, and E. Molina Grima, “Photobioreactors for the production of microalgae,” *Reviews in Environmental Science and Biotechnology*, vol. 12, no. 2, pp. 131–151, 6 2013.

[2] E. G. Nwoba, D. A. Parlevliet, D. W. Laird, K. Alameh, and N. R. Moheimani, “Light management technologies for increasing algal photobioreactor efficiency,” 2019.

[3] M. H. Sarrafzadeh, H. J. La, J. Y. Lee, D. H. Cho, S. Y. Shin, W. J. Kim, and H. M. Oh, “Microalgae biomass quantification by digital image processing and RGB color analysis,” *Journal of Applied Phycology*, vol. 27, no. 1, pp. 205–209, 2014.

[4] T. E. Murphy, K. Macon, and H. Berberoglu, “Multispectral image analysis for algal biomass quantification,” *Biotechnology Progress*, vol. 29, no. 3, pp. 808–816, 5 2013.

[5] “IEEE Standard for Local and metropolitan area networks–Part 15.7: Short-Range Optical Wireless Communications - Redline,” *IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011) - Redline*, pp. 1–670, 2019.

[6] N. Saeed, S. Guo, K. H. Park, T. Y. Al-Naffouri, and M. S. Alouini, “Optical camera communications: Survey, use cases, challenges, and future trends,” *Physical Communication*, vol. 37, p. 100900, 12 2019.

[7] A. P. Carvalho, S. O. Silva, J. M. Baptista, and F. X. Malcata, “Light requirements in microalgal photobioreactors: An overview of biophotonic aspects,” pp. 1275–1288, 3 2011.

[8] T. de Mooij, G. de Vries, C. Latsos, R. H. Wijffels, and M. Janssen, “Impact of light color on photobioreactor productivity,” *Algal Research*, 2016.

[9] S. Abu-Ghosh, D. Fixler, Z. Dubinsky, and D. Iluz, “Flashing light in microalgae biotechnology,” pp. 357–363, 3 2016.

[10] P. S. Schulze, R. Guerra, H. Pereira, L. M. Schüler, and J. C. Varela, “Flashing LEDs for Microalgal Production,” *Trends in Biotechnology*, vol. 35, no. 11, pp. 1088–1101, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.tibtech.2017.07.011

[11] A. P. Koller, H. Löwe, V. Schmid, S. Mundt, and D. Weuster-Botz, “Model-supported phototrophic growth studies with ¡i¿Scenedesmus obtusiusculus¡/i¿ in a flat-plate photobioreactor,” *Biotechnology and Bioengineering*, vol. 114, no. 2, pp. 308–320, 2 2017. [Online]. Available: http://doi.wiley.com/10.1002/bit.26072

[12] H. Aoyama and M. Oshima, “Line scan sampling for visible light communication: Theory and practice,” *IEEE International Conference on Communications*, vol. 2015-Septe, pp. 5060–5065, 2015.

[13] X. Li, N. B. Hassan, A. Burton, Z. Ghassemlooy, S. Zvanovec, and R. Perez-jimenez, “A Simplified Model for the Rolling Shutter Based Camera in Optical Camera Communications.”

[14] M. Shahjalal, M. K. Hasan, Y. M. Jang, V. T. Nguyen, and M. Z. Chowdhury, “Performance Analysis and Improvement of Optical Camera Communication,” *Applied Sciences*, vol. 8, no. 12, p. 2527, 2018.

[15] C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, “Correlation-based receiver for optical camera communications,” *Optics Express*, vol. 27, no. 14, p. 19150, 7 2019.

Barcolits: Barcodes using LED tags and optical camera communications

Cristo Jurado-Verdu,*[1] Victor Guerra,[2] Jose Rabadan[1], Rafael Perez-Jimenez[1]
[1] Institute for Technological Development and Innovation in Communications (IDeTIC)
Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain
[2] *Pi Lighting Sarl*, Sion, Switzerland
{*cjurado, vguerra, jrabadan, rperez*}*@idetic.eu* ***https://orcid.org/0000-0002-7371-5563*

2022 IEEE 18th International Conference on Factory Communication Systems (WFCS) | 978-1-6654-1086-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/WFCS53837.2022.9779172

***Abstract*—This work presents Barcolits, active LED tags that exploit the sequential acquisition of rolling-shutter cameras to materialize virtual barcodes in images. The generated barcodes have valuable properties. Compared to physically printed barcodes, its size (in pixels) does not change with the distance to the camera. Also, its stripes are always vertically aligned in the image independently of the rotation and position of the camera. Moreover, as Barcolits are active devices, they can be scanned under low light conditions or from reflections in walls or mirrors. Furthermore, this approach reuses conventional barcode detection and decoding algorithms, allowing interoperability and backwards compatibility with current barcode scanners (image or photodiode-based). These benefits have been experimentally validated using three different barcode encoding schemes: Code 128, EAN-13 and ITF. In the experiment conducted, these barcodes are compared in terms of their minimum required size to ensure high reliability, their data density, their signal level, and their likelihood to produce flickering in the LED source. The experiment results show that Code 128 are ideal for high throughput data sharing, with the highest achievable data density of 4.92 bits per pixel and the lowest power consumption. It outperforms standard manchester codes by encoding 42% more bits using the same payload size. On the other hand, EAN-13 barcodes require the smallest size for a successful scan but produce noticeable flickering if a long stream of different barcodes is transmitted. Finally, ITF codes mitigate flickering at the cost of reducing their density. This extensive analysis provides a comparison tool to help choose the optimal barcode scheme for different applications and environments.**

***Index Terms*—Optical camera communication, visible light communication, rolling shutter, barcode, industry 4.0**

I. INTRODUCTION

Barcodes have revolutionized business and industry since their invention in 1949. They are widely used worldwide, throughout the supply chain and in many contexts. They are applied to products in retail stores as part of the purchase process, in warehouses to track inventory [1], on invoices to assist accounting, in advertising to help reach additional content [2], in the healthcare and hospitals to identify patients and medicaments [3], among other applications. Barcodes are simple, flexible, cost-effective and provide high accuracy. However, given the extent of their use, the consequences associated with the 1% inaccurate barcode scans can have a significant detrimental impact in many sectors and with remarkably diverse consequences.

One of the most important factors influencing the reliability of barcodes is their image projection size (in pixels) for image-based scanners [4]. For a successful scan, It is necessary to guarantee a minimum width resolution for the narrowest segment. The attained size will depend on geometrical parameters such as the barcode's physical size, position, distance, and rotation relative to the camera and the camera's resolution, focal distance and field of view. Rotated and perspectively distorted barcode readings obtained from an improper camera position increase the complexity of the decoding algorithms. Other critical factors are the light scene conditions and light properties of the barcode. Since it is a passive reflector element, it relies on an external light source. Hence, the scanner will fail if the reflected light does not uniformly contrast white and black segments across the barcode. Printing barcodes on highly reflective, transparent, translucent (plastics bags), coloured, noisy or glossy surfaces; undergoing object shadows; or even configuring the camera with low exposure and gain settings are examples of possible scan fails. In addition, the camera's lenses can distort the barcode, increasing the difference in width of similar-sized segments, thus corrupting decoding. Last but not least, movement is also detrimental. Rolling shutter (RS) cameras deform moving images due to their sequential scanning in time, causing the barcodes to become noticeably corrupted under dynamic conditions.

Despite these numerous drawbacks, the industry has a steady interest in improving barcode detection and data recovery algorithms under harsh conditions, as demonstrated by the continuous update of the methods, including artificial intelligence models that outperform classical techniques [5]–[7]. However, a qualitative leap is still possible to overcome all these challenges without leaving behind all the technology developed, but by reusing it in better conditions.

This work presents Barcolits, active LED tags that exploit the sequential scanning of RS cameras to generate virtual barcodes in images. The term virtual refers to the fact that there is no physical printed barcode. The barcode materializes during the capture of the image as a result of the camera's acquisition mechanism. The Barcolit itself is a squared-shape light marker that appears uniformly illuminated to the naked eye. However, it switches between on and off states at high

---

This work has been funded in part by the spanish research administration (AEI, Project OCCAM, Ref. PID2020-114561RB-100) and under the frame of the NEWFOCUS COST action (Ref: CA19111), sponsored by the European Union.

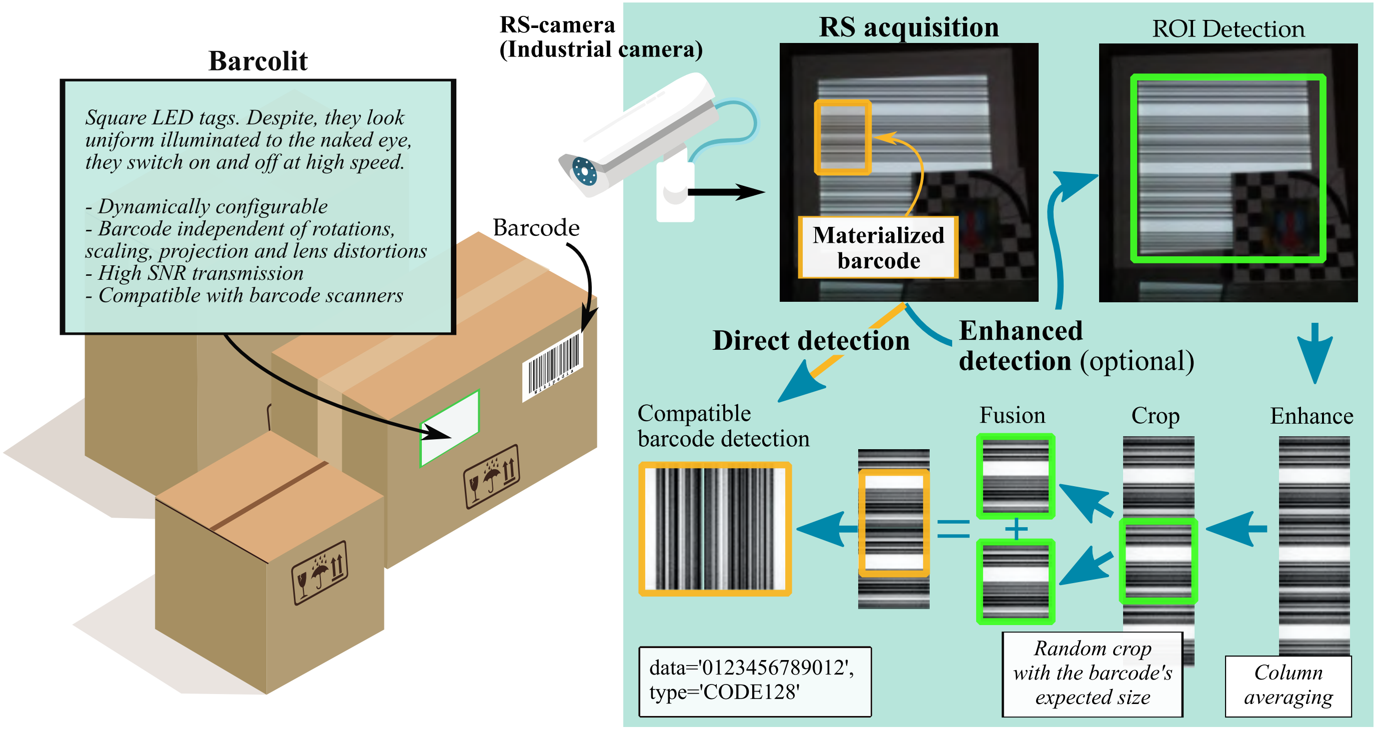


Fig. 1. Proposed communications scheme.

speed (10 to 50K changes per second). These transitions are detected by an RS camera. Because the camera exposes the image sensor sequentially row by row of pixels, it samples the light at successive instants, generating images that contain different illuminated stripes. The stripe intensity depends on the intensity level of the LED at the time of activation of each row [8]. The technology behind this proposal is known as RS-based optical camera communications (OCC) [9]–[12], and is considered a branch of visible light communication (VLC) [13] in the field of optical wireless communication (OWC). This technology, as detailed in the standard [14], aim to reuse light sources and camera-enabled devices to establish optical wireless communications links. Previous related works have proposed and experimentally validated different modulations schemes for RS-based OCC links such as amplitude shift-keying (ASK) [15], frequency-shift keying (FSK) [16], [17], quadrature amplitude modulation (QAM) [18], color-shift keying [19]; different multiplexing strategies such as orthogonal frequency division multiplexing (OFDM) [20], MIMO approaches [21]; and custom encoding schemes as presented in [22]. However, these approaches rely on custom decoding algorithms that have not yet been brought to a real application environment. On the other hand, Barcolits uses standard barcode encoding schemes, which allows the reuse of available ready detection and decoding barcode algorithms developed and tested by millions of users. This also offers backward compatibility and interoperability with conventional image and photodiode-based barcode scanners, which helps enable an OCC technology that is completely non-intrusive with the processes, use cases, and applications in which printed barcodes are currently used.

Furthermore, Barcolits present unique features and advantages that position them as promising candidates for replacing traditional barcodes. These advantages are the following. The barcode's size (in pixels) does not change with the camera distance, and its stripes will always be vertically aligned within the image, independently of the camera's rotation and position. In addition, they will not suffer from lens distortions and will always be straight. Moreover, the Barcolits can be scanned under low light conditions, even in dark scenarios, as they are active devices. Furthermore, they can be recovered from reflections in walls and mirrors, allowing non-line of sight detection [23]–[25]. All these benefits will be explained and justified in Section II supported with image examples that validate them experimentally.

The remainder of this paper is organized as follows. Section II details the communications scheme and presents preliminary results from the experimental evaluation of Barcolits. Section III describes the experimental setup and methodology used in the validation. Finally, Section IV discuss the results obtained, summarized in the conclusions section (Section V).

Communications scheme

II. COMMUNICATIONS SCHEME

Figure 1 shows the communications scheme proposed. In this picture, a Barcolit is attached on the side of a generic postal box, alongside a traditional barcode for comparison purposes, which has the exact dimensions.

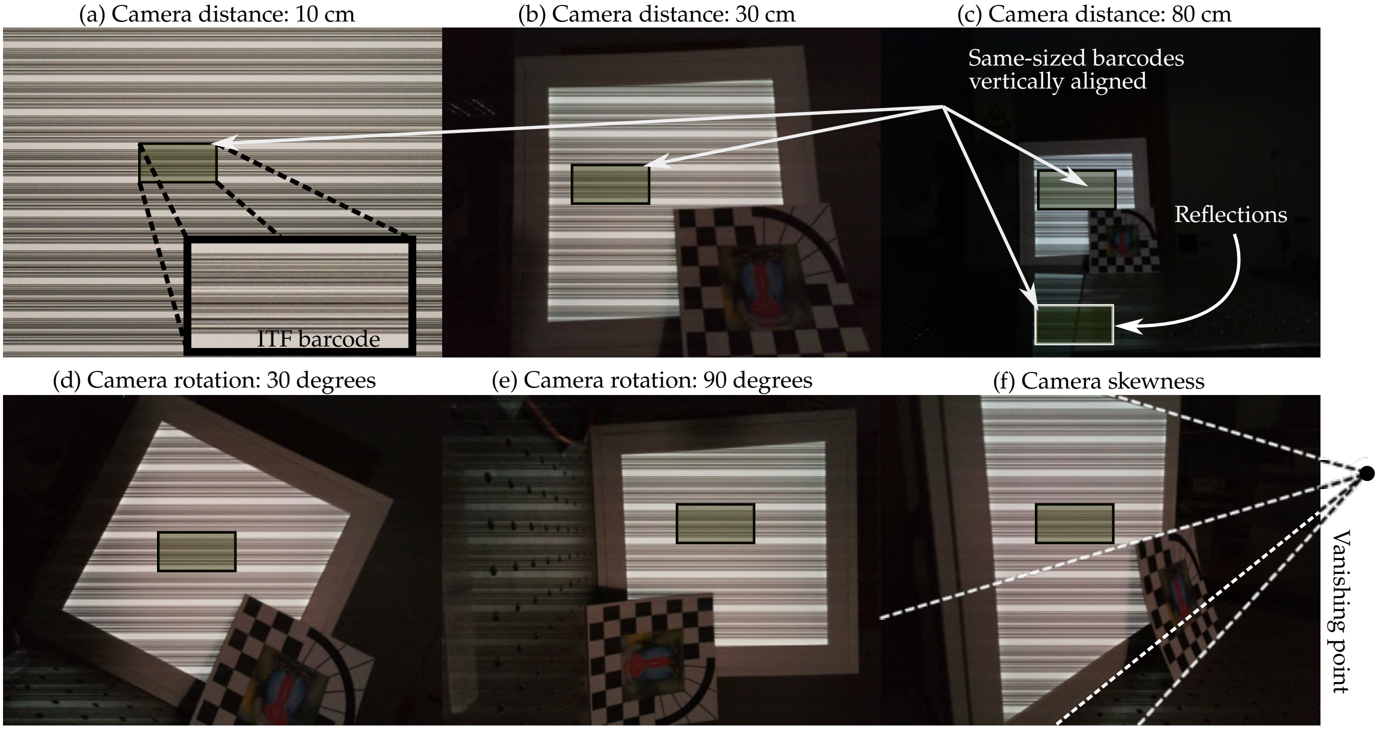


Fig. 2. Configuration examples.

Barcolits, as described in the introduction, are battery-powered LED square markers that switch between on and off illumination states to produce bright and dark stripes in the images captured by RS cameras. These different illuminated stripes are produced due to the RS acquisition mechanism, in which the image sensor is exposed sequentially row by row, sampling the light at different instants. As a result, virtual barcodes materialize within the Barcolit's image projection. Once the image is captured and retrieved by the camera, classic barcoding detection and data recovery algorithms can be utilized without exception, as both traditional and virtual barcodes follow the same encoding scheme.

The main benefit of this barcode generation is that the stripe height does not depend on the distance to the camera, unlike traditional barcodes. This size is directly defined by the relation between the time the LED panel maintains a particular illumination state and the interval between the activation of two consecutive rows. The former is known as the transmitted symbol time, t_{sym}, and the latter as the camera's row sampling period, which, in the end, corresponds to the receiver sampling period, T_{s} (using communications terminology). Eq. (1) relates both times to the stripe's height, known as the number of pixels per symbol (NPPS) [26], which is the theoretical number of samples per symbol in a communications system.

$$\mathrm{NPPS} = \frac{t_{sym}}{T_{\mathrm{s}}} \tag{1}$$

Therefore, by selecting the appropriate t_{sym} and T_{s}, the NPPS can be adjusted to the preferred barcode's size based on the application. Nevertheless, in most cases, this size is desired to be as small as possible (while ensuring a reliable detection) to exploit the vertical resolution of the image efficiently, i.e., to fit more subsequent barcodes within the image. Figures 2(a-c) shows examples of the barcodes captured using the RS-acquisition mechanism at different distances. The virtual barcode is highlighted in all pictures by surrounding it with a black rectangle. These figures show that the barcode size, h_{barcode} is independent of the distance. However, the Barcolit image projection, h_{Barcolit} does decrease with the distance. At 10 cm, it occupies the image entirely, and at 80cm, a third portion of the vertical resolution, h_{image}. In the same way, the number of materialized barcodes that fit inside the projection decreases. At 10 cm, nine barcodes materialize one after the other, while at 80 cm, only two barcodes fit within the Barcolit's image projection. The number of recovered barcolits, N, depends on the ratio between the image projection and the expected barcode sizes (Eq.2).

$$N = \frac{h_{\mathrm{barcode}}}{h_{\mathrm{Barcolit}}} \tag{2}$$

Note that the barcodes look similar in this example case. This is because the LED sends the same barcode repeatedly. However, it is possible to transmit different barcodes over time, increasing the data rate. On the other hand, Figures 2(d-f) show examples in which the camera have been rotated, 30 and 90 degrees (Fig. 2(d,e)), and translated (Fig. 2(f)). These figures show that the barcode stripes remain straight and perfectly aligned to the image's vertical dimension in all cases. This is exceptionally convenient for barcode detection and decoding, significantly reducing its complexity. This feature of the virtual barcodes is particularly beneficial in the latter example, in which traditional barcodes stripes will follow the image's vanishing point, ultimately acquiring a trapezoidal shape.

In conclusion, independently of the Barcolit distance and its projection over the image plane, the barcode will always have a fixed size and will be perfectly aligned to the scanning dimension of the camera, as opposed to traditional barcodes.

Regarding the detection and decoding methods, as barcodes and virtual barcodes are similar in the images, the image-based scanners underlying algorithms can be used interchangeably. Highlight that these algorithms are developed to detect barcodes by searching for the particular patterns that constitute them. Its encoding scheme is thus designed to facilitate its detection. Therefore, in contrast to conventional OCC links, it is not required to previously detect the region of interest (ROI) where the barcodes are expected to be located in the image. Instead, barcodes can be directly searched in the entire image.

In conclusion, because virtual barcodes are similar to barcodes, they can be correctly detected using conventional well-tested algorithms, as long as the following requisites are met: (i) the Barcolit must illuminate its surface uniformly and (ii) its image projection must be large enough to fit at least two barcodes (Eq.3).

$$h_{\mathrm{Barcolit}} \geq 2 \cdot h_{\mathrm{barcode}} \tag{3}$$

In other words, the size of the Barcolit's image projection must be at least twice (and not the same) the expected size of the barcode. This requisite prevents the consequences of the non-perfect synchronization between the camera and the Barcolit. Because of this non-perfect synchronization, the barcode will appear in different positions within the Barcolit image projection. Consequently, if its size is equal to the barcode's size, there is a high probability that the barcodes will appear trimmed. This requisite imposed on the size of the projection regarding the expected barcode size, which is expressed in literature as the ROI size required based on the data size, governs the design of OCC links [27].

This requirement is a clear disadvantage of the proposed Barcolit compared to traditional barcodes. For instance, the Barcolit size must double the physically printed barcode size to be detected at the same distance.

However, this work proposes a novel approach that enables barcode detection under conditions where the size of Barcolit's projection matches the barcode size. Thus the requisite expressed in Eq.3 is replaced for the new expressed in Eq.4.

$$h_{\mathrm{Barcolit}} \geq h_{\mathrm{barcode}} \tag{4}$$

Figure 1 shows the processing steps required to satisfy this condition. First, it is required to detect the ROI or select a region where it is expected to find a barcode. ROI detection can be accomplished using classical image processing techniques focused on finding the shape enclosing the light source. After determining the ROI, the columns can optionally be averaged to increase the signal to noise ratio (SNR). This averaging increases approximately the SNR by a factor of n, corresponding to the number of columns used (assuming that the image is affected by an additive white gaussian noise (AWGN)). This step is not strictly necessary to decode barcodes, but it can improve the system's performance.

Second, it follows a random crop from the enhanced segment with the barcode's expected size, h_{barcode}. Remark that the size of this crop must be precisely the expected size of the barcode. Finally, the extracted segment is cloned and concatenated at its very end to generate a new image that doubles its original size. In this image emerges a barcode that can be detected using conventional algorithms. This is because the concatenation of the cropped segment fuses the upper and lower part of a trimmed barcode. In conclusion, this novel implementation (one of the contributions of this work) solves one of the critical challenges highlighted in recent literature related to the required ROI size regarding the expected size of data to prevent data trimming.

A. Compatibility with barcode scanner devices

The proposed Barcolits are designed explicitly to reuse the underlying barcode detection algorithms (software elements) utilized in conventional image-based scanners. However, Barcolits are also compatible with actual scanner devices under certain conditions. This subsection discusses the preliminary requisites to ensure its compatibility with different types of scanner devices.

There are four types of barcode scanners available: camera-based readers, laser scanners, pen-type readers, and CCD readers. This list is ordered from highest to lowest penetration in the industrial and commercial sectors. In addition, the last two are being discontinued because of their worse performance compared to the first ones.

Camera-based readers include an RS camera that delivers a continuous stream of images and a computing unit that detects barcodes using conventional image processing algorithms. Therefore, these scanners are fully compatible with Barcolits. The only requirement is to configure the camera with a short exposure time. The exposure time is the interval in which pixels are exposed to light. It determines the amount of light received from the scene. For instance, long exposures produce bright images. Accordingly, the exposure time is increased under low light conditions to allow perceiving objects in the images. However, long exposures can negatively impact the Barcolit's transmitted barcode. Under prolonged exposures, the barcode's stripes bar appears mixed. This is because the pixels, exposed during a longer period, integrate the light of several consecutive symbols, producing a significant intersymbol interference (ISI). This ISI can be prevented by configuring the camera with short exposure times. In conclusion, the compatibility of the proposed Barcolits with conventional camera-based readers is guaranteed in cases where the camera is configured with short exposure times.

Laser scanners include a laser, an oscillating mirror (or rotating prims) and a photodiode. During the scan, the laser generates a light beam which is redirected by the mirror into the barcode. This mirror is incrementally rotated to redirect the laser beam across the barcode. At each rotation, the photodiode samples the reflected light from a particular barcode position. At first glance, it might seem that these scanners are not compatible with Barcolits. However, these readers follow a sequential scanning comparable to the sequential scanning of RS cameras. In the same way, as in RS cameras, the row sampling time corresponds to the receiver sampling period, in laser scanners, it corresponds to the time taken to increment the rotation angle of the mirror one unit. Therefore, if both times are comparable, the laser scanner can detect barcodes directly from the Barcolits' surface. Yet other requirements must also be met. First, the duration of one scanning cycle must be greater than the time required to transmit an entire barcode with the Barcolit, ensuring that at least one barcode is transmitted during scanning time. Second, the light spectrum of the Barcolit's must include the wavelength of the laser, which can be red or infrared.

Pen-type readers include a light source and a photodiode to read the barcodes. It uses the same approach as laser scanners, as it measures the reflected light of the different stripe bars. The difference is that the user must move its tip across the bars at a relatively uniform speed to operate this scanner. Highlight that, in this case, the pen is continuously sampling the incoming light at a fixed rate, independently of the user speed. Therefore, if this sampling time is comparable to the row sampling time of the camera, these pens are compatible with the Barcolits. Furthermore, it is not required that the user move the pen across the Barcolit, as the barcode is transmitted directly by varying the illumination.

Finally, CCD scanners include a charge-coupled device (CCD) consisting of a line of hundreds of pixel photodiodes. These photodiodes are simultaneously exposed to the light coming from the barcode's surface at the scanning time. This acquisition mechanism is known as global shutter acquisition, and it is incompatible with Barcolits. This is because, at scanning time, all the photodiodes are exposed to the same illumination state.

III. EXPERIMENTAL SETUP

Figure 3 shows the experimental setup. As proof of concept, a rectangular white LED panel is connected to an arbitrary wave generator (AWG), which generates the barcode signal.

At the receiver side, a camera (PiCamera V2) [28] attached to a Raspberry Pi captures images continuously and streams them to a processing unit deployed in a cloud infrastructure. A reference picture stands next to the Barcolit to account for the position and rotation of the camera. Finally, a specular reflector is placed on the ground. The experiment's key parameters are detailed in Table I.

Regarding the decoding algorithms, the core library is written on Python and uses the Pyzbar python wrappers of the open-source C/C++ barcode reading library. The use of this library demonstrates the compatibility of conventional detection and decoding algorithms with virtual barcolits.

The conducted experiment aims to evaluate and compare the performance of three different barcode encoding schemes using this Barcolits proof-of-concept: code 128, defined in ISO/IEC 15417:2007) [29]; European article number (EAN-13) and interleaved 2 of 5 (ITF), defined in GS1 [30], [31].

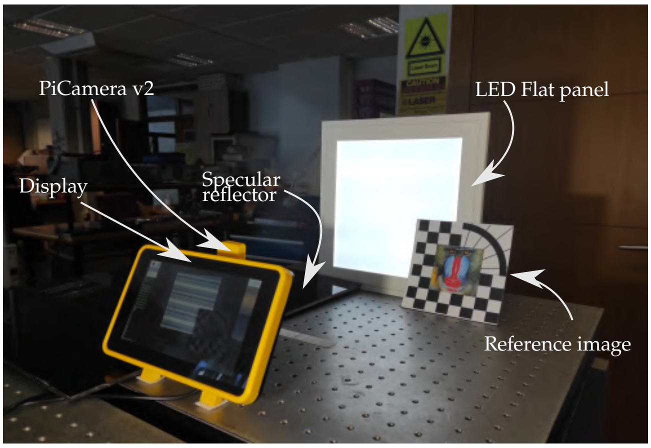


Fig. 3. Experimental setup.

On the one hand, it analyzes the maximum achievable data density, the amount of information the barcode encodes (in bits) and its required size (in pixels) to ensure detection and decoding. The data density has units of bits per pixel. It is important to clarify that the size of the barcode also considers the size of the quiet zone to be set between successive barcodes, expressed as a percentage of the original barcode's size. If this quiet zone is not large enough, the barcodes cannot be successfully detected or differentiated from each other. Therefore, the minimum size required for each code is determined by minimizing the size of the quiet zone for each encoding scheme until the system barcode recall is higher than 90%. In other words, until valid barcodes are detected in 90% of the images. More than 3000 images for each barcode type containing random digits and surrounded with increasing quiet zones are captured for this evaluation.

On the other hand, this evaluation analyzes the normalized signal level (i.e., the average of the normalized signal values) for each barcode type considering different payloads. The signal level is directly related to the LED's power consumption, and it is preferred to have the lowest possible value. In addition, barcodes carrying different payloads might have different signal levels, depending on the encoding scheme. Therefore, transmitting consecutive barcodes with different payloads might produce a noticeable flickering. This flickering is produced due to the slow variations in time of the signal level associated with each barcode. The standard deviation of the signal levels for different payloads can be assessed to account for the probability of producing a noticeable flickering. The higher the standard deviation, the higher the chances to produce flickering.

IV. RESULTS

Figure 4 depicts real examples of captures of different barcode types (Code 128, EAN-13 and ITF). It reveals that the required quiet zone for each barcode (highlighted with

TABLE I
EXPERIMENT'S KEY PARAMETERS

Parameter	*Value*
Barcolit	
Symbol span — npps	$37,8\mu s$ — 2 pixels per symbol
Barcode types	Code 128, EAN-13
Quiet zones	0 to 20 %
Receiver	
Camera model	PiCamera V2 [28]
Resolution	1920 x 1080 pixels
Aperture lens — Focal length (equivalent)	f/2 — 3mm
Row sampling period	$18.904\mu s$
Exposure time	$25\mu s$
Analog gain	12.0
Digital gain	1.0
Red and blue gains	1.85, 2.34
Link	
Relative distance	80cm

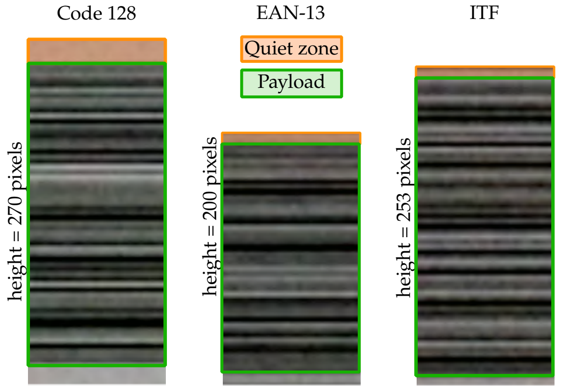


Fig. 4. Capture examples. Transmitted digits '0123456789012'.

TABLE II
COMPARISON TABLE OF DIFFERENT BARCODES.

Barcode type	Code 128	EAN-13	ITF
Barcode size (pixels)	270	**200**	253
Barcode payload (pixels)	**245**	190	241
Quiet zone (% — pixels)	10% (25)	**5%** (10)	**5%** (12)
Number of symbols	**103**	10	10
Barcode length (symbols)	**var** (13)	fixed (13)	fixed (12)
Barcode bits	**87**	43	40
Density (bits per pixel)	**4,96**	0,65	0,47
Reference bits	**61**	48	60
Reference ratio	**1.42**	0.90	0.67
Mean (μ) - Norm. energy level	**0.51**	0.54	**0.51**
Std (σ) - Norm. energy level	0.02	0.03	**0.00**

an orange rectangle) is different. Code 128 barcodes require the larger quiet zones to be successfully detected. In addition, this figure shows that the barcode size varies depending on the encoding scheme for the same payload consisting of the digits '0123456789012'. EAN13 barcodes have the smallest size.

Table II compares the barcode's evaluation metrics. In bold are selected the best values for each metric. Starting from the barcode's size, EAN-13 is the preferred type with a size of 200 pixels, having the minimum required quiet zone of only 10 pixels (equivalent to 5% of the barcode size). However, the number of different symbols it can transmit is only 10 (i.e., the digits 0,1,2,3,4,5,6,7,8 and 9), which is far from the number of symbols that can be transmitted using the barcodes Code 128 (i.e., 103).

Therefore, a fairer comparison requires analyzing the barcode's density. In these terms, Code 128 outperforms other barcode types by transmitting up to 4.96 bits per pixel. This value is derived from the fact that Code 128 barcodes have a fixed length of 13 slots that can accommodate one of 103 available different symbols. The theoretical amount of information each symbol transmits is 6,69 bits (obtained by applying the binary logarithm to 103). Therefore, the total amount of information transmitted using this barcode type is approximately 86 bits. Dividing the total size of the barcode (i.e., 270 pixels) by this number gives the corresponding barcode's density. The same approach is used to compute the barcode's density in the other cases.

In addition, this table adds another metric to compare the barcode encoding performance with the Manchester encoding scheme, which is conventionally used in traditional RS-based OCC links (as defined in the standard). This metric has the label of reference bits. It corresponds to the amount of information (in bits) a manchester encoded signal can carry using the same barcode size (in pixels). The ratio between the barcode and the reference bits is shown below. As it can be seen, Code 128 outperforms the manchester codes by encoding up to 42% more bits (i.e., 87 vs 61 bits) in the same space. Therefore, it is revealed that Code 128 is a more efficient encoding scheme than traditional Manchester encoding. It exploits better the available space to transmit the information.

Finally, in terms of the barcodes' normalized signal level, Code 128 has the lowest signal level (0.51) on average, similar to the ITF codes. Therefore, both code types are more suitable for low-power communications than the EAN-13 codes. Nevertheless, ITF codes have another important feature. The standard deviation of the signal level for different payloads is zero. This means that all the barcodes generated with this encoding scheme have the same signal level independently of the payload. This is because all its symbols are designed with the same number of ones (white stripes) and zeroes (black stripes) to ease the detection and decoding algorithms. In particular, each symbol has five bits: two ones and three zeroes. This explains why this code has the lowest data density. ITF barcodes favour detection at the expense of transmitting fewer bits. However, their intrinsic feature of always having the same

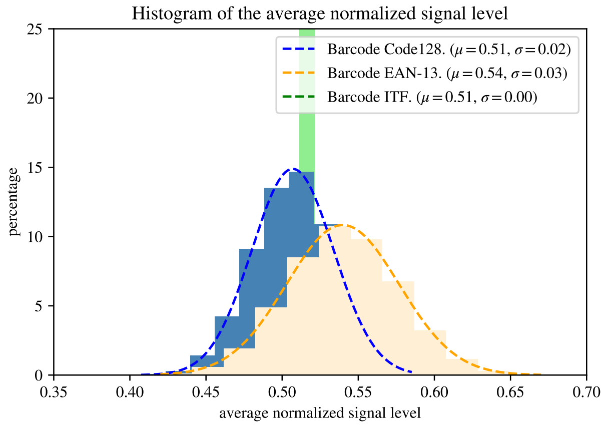


Fig. 5. Histogram of normalized energy level for random digits.

signal level makes them ideal candidates for transmissions where flickering can be an issue.

Alternatively, Code 128 with a standard deviation of 0.2 is preferred over EAN-13 codes if the application requirement is to reduce the probability of flickering perception. However, it is important to highlight that no flickering is perceived when the same payload is transmitted repeatedly, as the signal level of the barcode is perceived constant over time. The chance of perceiving flickering will depend on the sequential transmission of different barcodes.

Figure 5 shows the histogram of random payloads for each barcode type. As it can be seen, EAN-13 has a more significant standard deviation and average level, which makes them the least suitable in terms of energy, and flickering prevention. As mentioned, ITF codes have a standard deviation equal to zero, which means that all the barcodes share the same signal level independently of the payload. Hence, ITF codes are preferred in those cases where flickering can be an issue.

In summary, as a design criterion, Code 128 barcodes are preferred in cases where a significant amount of data needs to be transmitted, and the flickering requirements are permissive. In addition, the power consumption of the Barcolit is the lowest, with a generated signal level of around 0.51. In addition, Code 128 barcodes are variable in length and, therefore, can adapt their barcode size, in contrast to the other two. Still, its data density decreases abruptly with the number of symbols per barcode.

EAN-13 codes can be used when the guaranteed detectable ROI size is short since they require the smallest barcode size for detection and decoding. For instance, these barcodes are preferred when working with moderate distance links.

Finally, if the flickering requirements are very rigid, then ITF codes are ideal. In addition, as they are well-balanced codes, they produce less stress on the LED source and allow the continuous transmission of many different barcodes without perceptible flickering.

V. CONCLUSIONS

This work presented Barcolits, active LED tags that exploit the sequential acquisition of rolling-shutter cameras to materialize virtual barcodes in images, as a promising alternative to traditional printed barcodes. These Barcolits, in the end, establishes OCC links with RS cameras, in which the standard encoding schemes are replaced with barcode encoding schemes. This approach enables backwards compatibility with conventional image-based barcode scanners and benefits from reusing well-tested and user-ready barcode detection and decoding algorithms. Moreover, the virtual barcodes have valuable properties as their size does not change with the camera distance, and its stripes are always vertically aligned within the image independently of the rotation and position of the camera. The experimental validation of three different barcode schemes: Code 128, EAN-13 and ITF, offers practical comparison guidelines, design principles, and metrics to help choose the optimal barcode for different applications. On the one hand, it analyzes the communication constraints such as the minimum required barcode size, ensuring high scan reliability, and the maximum achievable data throughput. On the other hand, it compares the signal generation constraints in terms of the average signal level, which is preferably low to enable low-power transmissions, and the likelihood to produce perceptible flickering in the Barcolit source. The experiment results show that Code 128 barcodes are optimal in applications with high data throughput requirements and in which the Barcolit's image projection is guaranteed to be extense. With a data density of 4.96 bits per pixel, they outperform Manchester encoding schemes by increasing 40% the number of transmitted bits using the same pixels. On the other hand, EAN-13 are preferable in applications where the attainable Barcolit's projection is reduced, for example, in applications with moderate to high link distances. However, long streams of different payloads are prone to produce perceptible flickering. In those cases where the flickering requirements are not permissive, the solution is to send the same repeated barcode over time to avoid slow variations on the signal level. Alternatively, ITF barcodes can be used, which maintains a signal level that is perceived constant to the naked eye. These barcodes mitigate the flickering issues at the cost of reducing their data density. Moreover, its size is smaller than Code 128, making them flexible and adaptable candidates for different environments and applications.

REFERENCES

[1] S. Hong-ying, "The application of barcode technology in logistics and warehouse management," in *2009 1s International Workshop on Education Technology and Computer Science*, vol. 3, pp. 732–735, 2009.

[2] Y.-Y. Chen, K.-Y. Chi, and K.-L. Hua, "Design of image barcodes for future mobile advertising," *EURASIP Journal on image and video processing*, vol. 2017, no. 1, pp. 1–12, 2017.

[3] P. Lachman and E. van der Wilden, "Use of barcode technology can make a difference to patient safety in the post-COVID era," *IJQHC Communications*, vol. 1, 08 2021. lyab014.

[4] F. Duchon, P. Bucka, M. Dekan, and A. Babinec, "Reliability of barcode detection," *Journal of Control Engineering and Applied Informatics*, vol. 21, no. 3, pp. 80–87, 2019.

[5] Y. Xiao and Z. Ming, “1d barcode detection via integrated deep-learning and geometric approach,” *Applied Sciences*, vol. 9, no. 16, 2019.

[6] D. K. Hansen, K. Nasrollahi, C. B. Rasmussen, and T. B. Moeslund, “Real-time barcode detection and classification using deep learning,” in *International Joint Conference on Computational Intelligence*, pp. 321–327, SCITEPRESS Digital Library, 2017.

[7] A. Zamberletti, I. Gallo, and S. Albertini, “Robust angle invariant 1d barcode detection,” in *2013 2nd IAPR Asian Conference on Pattern Recognition*, pp. 160–164, 2013.

[8] H. Aoyama and M. Oshima, “Line scan sampling for visible light communication: Theory and practice,” in *2015 IEEE International Conference on Communications (ICC)*, pp. 5060–5065, IEEE, 2015.

[9] W. Liu and Z. Xu, “Some practical constraints and solutions for optical camera communication,” *Philosophical Transactions of the Royal Society A*, vol. 378, no. 2169, p. 20190191, 2020.

[10] N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, “Optical camera communications: Survey, use cases, challenges, and future trends,” *Physical Communication*, vol. 37, p. 100900, 2019.

[11] N. T. Le, M. Hossain, and Y. M. Jang, “A survey of design and implementation for optical camera communication,” *Signal Processing: Image Communication*, vol. 53, pp. 95–109, 2017.

[12] N. Saha, M. S. Ifthekhar, N. T. Le, and Y. M. Jang, “Survey on optical camera communications: challenges and opportunities,” *IET Optoelectronics*, vol. 9, pp. 172–183(11), October 2015.

[13] Z. Ghassemlooy, L. N. Alves, S. Zvanovec, and M.-A. Khalighi, *Visible light communications: theory and applications*. CRC press, 2017.

[14] “IEEE Standard for Local and metropolitan area networks–Part 15.7: Short-Range Optical Wireless Communications,” *IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011)*, pp. 1–407, 2019.

[15] V. P. Rachim and W.-Y. Chung, “Multilevel intensity-modulation for rolling shutter-based optical camera communication,” *IEEE Photonics Technology Letters*, vol. 30, no. 10, pp. 903–906, 2018.

[16] C.-W. Chow, R.-J. Shiu, Y.-C. Liu, X.-L. Liao, K.-H. Lin, Y.-C. Wang, and Y.-Y. Chen, “Using advertisement light-panel and cmos image sensor with frequency-shift-keying for visible light communication,” *Opt. Express*, vol. 26, pp. 12530–12535, May 2018.

[17] Y. Xiao, W. Guan, S. Wen, J. Li, Z. Li, and M. Liu, “The optical bar code detection method based on optical camera communication using discrete fourier transform,” *IEEE Access*, vol. 8, pp. 123238–123252, 2020.

[18] W. Chujo and M. Kinoshita, “Rolling-shutter-based 16-qam optical camera communication by spatial luminance distribution,” *IEICE Communications Express*, 2019.

[19] H.-W. Chen, S.-S. Wen, X.-L. Wang, M.-Z. Liang, M.-Y. Li, Q.-C. Li, and Y. Liu, “Color-shift keying for optical camera communication using a rolling shutter mode,” *IEEE Photonics Journal*, vol. 11, no. 2, pp. 1–8, 2019.

[20] H. Nguyen, T. L. Pham, H. Nguyen, V. H. Nguyen, and Y. M. Jang, “Rolling mimo-ofdm for optical camera communication system,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 103–106, 2020.

[21] S. R. Teli, V. Matus, S. Zvanovec, R. Perez-Jimenez, S. Vitek, and Z. Ghassemlooy, “Optical camera communications for iot–rolling-shutter based mimo scheme with grouped led array transmitter,” *Sensors*, vol. 20, p. 3361, Jun 2020.

[22] C.-W. Chen, C.-W. Chow, Y. Liu, and C.-H. Yeh, “Efficient demodulation scheme for rolling-shutter-patterning of cmos image sensor based visible light communications,” *Optics express*, vol. 25, no. 20, pp. 24362–24367, 2017.

[23] J.-K. Lain, F.-C. Jhan, and Z.-D. Yang, “Non-line-of-sight optical camera communication in a heterogeneous reflective background,” *IEEE Photonics Journal*, vol. 11, no. 1, pp. 1–8, 2019.

[24] N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, P. Luo, and H. Le-Minh, “Non-line-of-sight 2 × N indoor optical camera communications,” *Applied optics*, vol. 57, no. 7, pp. B144–B149, 2018.

[25] W.-C. Wang, C.-W. Chow, L.-Y. Wei, Y. Liu, and C.-H. Yeh, “Long distance non-line-of-sight (NLOS) visible light signal detection based on rolling-shutter-patterning of mobile-phone camera,” *Optics express*, vol. 25, no. 9, pp. 10103–10108, 2017.

[26] C. Jurado-Verdu, V. Matus, J. Rabadan, V. Guerra, and R. Perez-Jimenez, “Correlation-based receiver for optical camera communications,” *Optics Express*, vol. 27, p. 19150, 7 2019.

[27] Y. Yang, J. Hao, and J. Luo, “Ceilingtalk: Lightweight indoor broadcast through led-camera communication,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 12, pp. 3308–3319, 2017.

[28] Sony Corporation, *IMX219PQH5-C, Diagonal 4.60 mm (Type 1/4.0) 8 Mega-Pixel CMOS Image Sensor with Square Pixel for Color Cameras, Datasheet*. Sony Corporation, 2014.

[29] International Organization for Standardization, ISO, “ISO/IEC 15417:2017 Automatic identification and data capture techniques — Code 128 bar code symbology specification.”

[30] “GS1 General Specifications,” *GS1*, pp. 265–279, 2021.

[31] D. Chai and F. Hock, “Locating and decoding ean-13 barcodes from images captured by digital cameras,” in *2005 5th International Conference on Information Communications Signal Processing*, pp. 1595–1599, 2005.

On-demand training of deep learning equalizers for rolling shutter optical camera communications

Cristo Jurado-Verdu,[1] Victor Guerra,[2] Carlos Guerra,[3] Jose Rabadan[1], Stanislav Zvánovec,[3]
Rafael Perez-Jimenez[1]
[1] Institute for Technological Development and Innovation in Communications (IDeTIC)
Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain
[2] *Pi Lighting Sarl*, Sion, Switzerland
[3] Faculty of Electrical Engineering, *Czech Technical University in Prague*, Czech Republic

Abstract—**The camera's exposure time restricts the reception bandwidth in rolling shutter-based optical camera communication links. Short exposures are preferable for communications, but under these conditions, the camera produces dark images with impracticable light conditions for human or machine-supervised applications. Alternatively, deep learning equalization stages can mitigate the effects of increasing the exposure time. These equalizers are trained using synthetic images based on the camera's exposure time and row sampling frequency. If these parameters are unknown in advance, another artificial network is used to estimate them directly for the captured images, the estimator. This estimator is trained offline using a vast number (thousands) of representative cases. This work proposes to transfer the attained knowledge from the offline pretrained estimator to the equalizer by using transfer learning techniques. In this way, the equalizers' training time is significantly reduced (435 times compared to full training). Consequently, transfer learning enables equalizers' online and on-demand training at reception without interfering with the communications. Results reveal that the complete training requires using exclusively 250 synthetic images to guarantee a communication performance with a bit error rate below 10^{-4} after the equalization.**

Index Terms—**Rolling shutter, Optical Camera Communication, Visible Light Communication, Equalization, Transfer learning, Deep Learning, Artificial Intelligence**

I. INTRODUCTION

Optical camera communication (OCC) is a branch of visible light communication (VLC) [1] in which the optical receivers are the pixels of an image sensor. The interest in this technology lies in the reuse of embedded cameras in a wide range of end-user devices (e.g., smartphones, vehicle dashcams, laptops) [2], [3]. In this way, it is intended to break the market's entry barriers imposed on VLC due to the necessity of using dedicated reception hardware (i.e., photodiode-based receivers) [4]. As stated, cameras should be preferably reused, implying that they might not lose their functionality as image capture systems. They must serve simultaneously as communication endpoints and imaging devices. Nevertheless, when a rolling-shutter (RS) camera acts as a receiver, it must be configured to achieve the highest link throughput [5]. The camera should select the lowest possible exposure time; otherwise, it will restrict the available reception bandwidth (acting as a low pass filter [6], [7]). However, under these short exposure conditions, the camera delivers dark images with impracticable light conditions for human or machine-supervised applications (i.e., users cannot perceive objects in the scene). Consequently, a trade-off arises in selecting the optimal exposure time, constraining the OCC links (at the design stage) to work with RS-cameras configured with a fixed exposure time below a specified threshold, above which communication is unfeasible. This demanding prerequisite neglects that most cameras have automatic configuration algorithms for setting the exposure (among other parameters) based on estimated ambient light conditions. Furthermore, they might even not provide access to their internal settings. For example, cameras in firmware-protected Android [8], and iOS smartphones do not reveal their internal settings or do not allow them to be set. For this purpose (i.e., to allow increasing the exposure time), a previous work [6] proposed using an equalization stage based on artificial intelligence. Based on a convolutional autoencoder (CAE), this equalizer returns reconstructed denoised versions of the input images, mitigating the exposure-related intersymbol interference (ISI). These equalizers are adjusted, on training, to a fixed exposure time by using synthetically generated images. Furthermore, given the convolutional nature of these equalizers, they adapt pretty well to exposure conditions that slightly deviate (up to 11%) from the target exposure time. Its experimental validation reveals that this equalizer outperforms state of the art artificial neural network (ANN)-based equalizers' performance [9], [10] by increasing the reception bandwidth up to 14 times compared to non-equalized systems and under low to moderate signal-to-noise ratio (SNR) conditions. Despite advances in exposure equalization, there is still the challenge of retrieving the training parameters (to generate the synthetic images) when they are unknown beforehand. Recall that the camera can automatically and dynamically adjust the exposure time or even do not reveal it. Hence, it is necessary to estimate at reception the internal configuration of the camera directly from images to perform the equalizer training. For this purpose, a previous work [11] proposes an estimator block based on convolutional networks that accurately estimates the camera's exposure time. Moreover, this estimator retrieves parameters related to the transmitter clock frequency. The estimated parameters

---

This work has been funded in part by the Spanish research administration (AEI, Project OCCAM, Ref. PID2020-114561RB-100) and under the frame of the NEWFOCUS COST action (Ref: CA19111), sponsored by the European Union.

are delivered to the following equalizing and decoding blocks in the reception chain. Consequently, the accurate estimation of these parameters enables the receiver to decode the transmitter source while unaware of the camera or transmitter configuration. In this way, the image acquisition systems (i.e. cameras, video streams) are decoupled from the reception routines, allowing decentralized decoding platforms to be deployed in the cloud. This is an extended tendency in wireless networks where the access points' functionalities are decoupled from the local hardware and translated the heavy processing to the cloud (e.g., Open Radio Access Networks). Finally, to provide support for a wide range of cameras and configurations, in [11], was proposed the use of a pretrained bank of equalizers, each of them adjusted to different camera configurations. This strategy (i.e., bank of equalizers) implies storing each equalizer's weights and network parameters in memory, which is not inconvenient if the storage capabilities are extensive. Still, it represents a challenge for low-cost embedded platforms. In this paper, we propose to train the equalizer at reception on-demand (instead of using pretrained equalizers) based on the estimates delivered by the estimator block. However, to perform this on-demand training without disrupting the communications is necessary to reduce the training time significantly and the number of synthetic images used, avoiding using valuable receiver resources. To achieve more efficient training, it is proposed to transfer the attained knowledge from the estimator (pretrained offline) to the equalizer by utilizing transfer learning techniques. The remainder of this paper is organized as follows. Section II introduces the proposed architecture and details the online and on-demand training of the equalizer by using transfer learning from the estimator. Section III describes the methods, procedures, metrics and the experimental setup used to evaluate the equalizer performance in terms of the bit error rate (BER). Section IV presents the results. Finally, the conclusions of this work are summarized in section V.

II. COMMUNICATIONS SCHEME

Figure 1 depicts the proposed architecture and its functional blocks. The transmitter consists of an illuminated flat-panel LED. It sends non-return to zero (NRZ) Manchester encoded on-off keying (OOK) pulses. The transmitted packet includes a header of five one bits, a trailer of one zero bit, and the payload. In addition, bit stuffing is used to prevent the header from appearing within the payload. A stuffed bit is inserted for every three bits and set to one if the preceding bit is zero and zero otherwise.

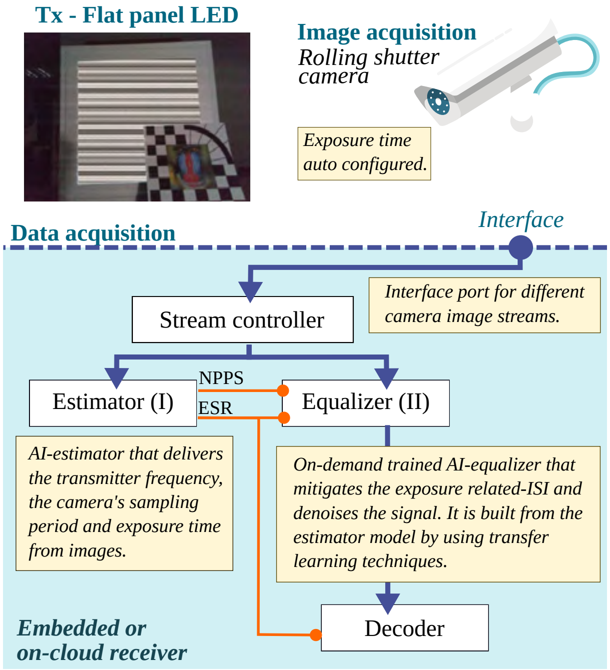


Fig. 1. Communications scheme.

The receiver is divided into two independent subsystems: the image and data acquisition. The former is a generic RS-camera that delivers a continuous image stream. The latter is a processing unit deployed in a hardware platform physically interconnected to the camera or in a cloud infrastructure. The stream controller provides the interface to this subsystem, handling different image streams and assigning priorities and resources. It is followed by the estimator, a deep learning block that ingests images (severely distorted by the camera's exposure time) and estimates important signal parameters used later in the reception chain: the number of pixels per symbol (NPPS) and the exposure to symbol ratio (ESR). These parameters relate the transmitter (Tx) symbol time, t_{sym}, with the receiver sampling period, T_{S}, which corresponds to the camera's row sampling time (Eq.1 [12]), and the exposure time, t_{exp} (Eq.2 [11]), respectively.

$$\mathrm{NPPS} = \frac{t_{\mathrm{sym}}}{T_{\mathrm{S}}} \tag{1}$$

$$\mathrm{ESR} = \frac{t_{\mathrm{exp}}}{t_{\mathrm{sym}}} \tag{2}$$

Estimating these parameters directly from the images implies that it is no longer necessary to restrict the Tx frequency at the design stage or establish very restrictive constraints for the camera devices and configurations. Consequently, the reception algorithm operates for a wider variety of camera devices, configurations, and transmitter speeds.

The retrieved parameters are used by the equalizer (NPPS and ESR) to generate synthetic image samples for training the network and by the decoder (NPPS) to recover the signal clock.

The equalizer is based on a Convolutional Autoencoder (CAE). Its responsibility is to ingest the severely distorted images by the camera's exposure time and deliver an equalized and denoised version. This CAE is adjusted on training by feeding the network with representative synthetically-generated images for particular NPPS and ESR conditions. The CAEs require large datasets ($>$ 30000 images) and many epochs for training to achieve acceptable results. Consequently, its training must be performed offline to avoid occupying the receiver intensively. Hence, a bank of offline pretrained CAEs must be stored in memory to be chosen for different scenarios. However, the number of CAEs that

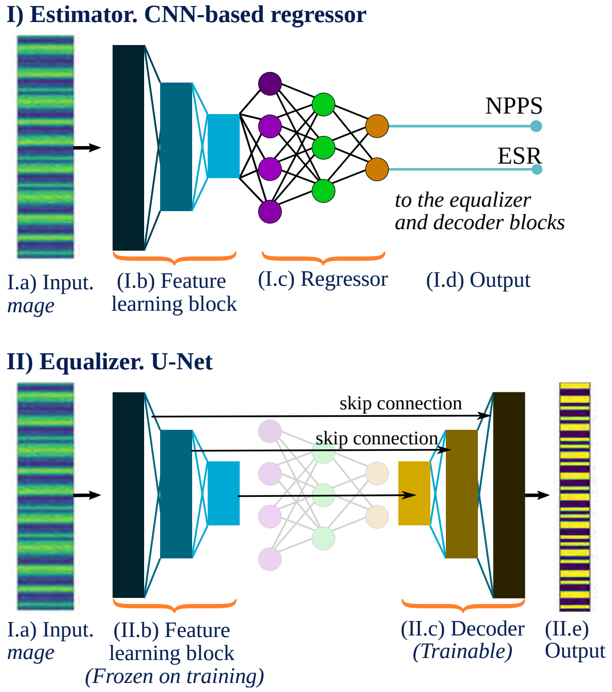


Fig. 2. Estimator and equalizer blocks.

the receiver can hold in memory depends on its storage capabilities. This can be challenging when using a low-cost embedded platform (e.g. smartphone). As an alternative, the training time and resources should be optimized to enable the receiver to train the equalizer on-demand. To accomplish this, this work proposes to transfer the attained knowledge of the estimator, which has been trained offline with a huge synthetic dataset containing all the possible configurations, to the equalizer by using transfer learning techniques [13]. In this way, the training time and the dataset size are considerably reduced. The following subsection describes how this transfer learning is performed and the conditions both the estimator and the equalizer must meet to accomplish it.

A. *transfer learning from the estimator to the equalizer*

The estimator is a convolutional neural network (CNN). Figure 2(I) shows the topology of this network, which consists of: a feature learning block that extracts a set of features from the input images using depthwise separable convolutional layers (DSC) [14]; and a regression ANN that estimates the NPPS and the ESR as a function of those latent features. Recall that this estimator is pretrained (before deployment) with a vast dataset consisting of synthetic images for a wide variety of NPPSs and ESRs. On the other hand, the equalizer is a CAE (as detailed in [6]). It consists of: an encoder block, based on convolutional layers (Conv2D), that extracts relevant features from the images; and a decoder block, primarily based on transpose convolutional layers (TConv), that tries to reconstruct the original image from this latent representation, reducing the noise and alleviating the exposure-related ISI. At this point, it is interesting to highlight that the feature learning block of the estimator and the encoder of the equalizer are practically similar. Ultimately, both networks use convolutional layers to extract relevant features from images (either DSC or Conv2D layers, respectively). Therefore, the feature learning block can be repurposed as the equalizer encoder block, as shown in Figure 2(II). The construction of the new equalizer model is accomplished by removing the estimator's regression network and connecting its feature learning block to the equalizer's decoder. In addition, a series of skip connections optimize even further the performance of this model. These skip connections feed the output of one encoder layer as the input to a same-level decoder layer. The resulting topology is known as U-Net [15] because it has the shape of a U. This topology outperforms traditional CAE in image segmentation tasks, as the skip connections preserve the spatial integrity of the input image in the output domain. In this work, these skip connections help to precisely locate the output data. Transfer learning is then achieved by transferring the learned network parameters of the estimator's feature learning block to the equalizer's encoder. Finally, during the equalizer training, those parameters are frozen (i.e. not trainable), and just the decoder part is trained.

III. METHODOLOGY

This work aims to evaluate the performance of the equalizer block, which is trained online (during frame acquisition) using the least number of synthetic images possible, ensuring a bit error rate (BER) below the forward error correction (FEC) limit (i.e., 3.8×10^{-3}). The on-demand synthetic image generation depends on the estimator block's previous NPPS and ESR estimates. These estimates are not ideal and may differ from the target values. Consequently, this experiment also validates whether the equalizer performance is robust to these estimation errors. To assess this behaviour, the worst estimates for the NPPS and ESR (with the most significant error) are selected from the results obtained in [11].

The procedures used in this evaluation are detailed below. The worst estimates for a particular target's values are specified in the first stage. The selected target values for the NPPS and the ESR are 4.0 and 2.62, respectively. In that case, the estimated values for the NPPS and the ESR are 3.9 and 2.89. Therefore, there is a relative error in the estimations of 2%, and 9.3%, respectively.

Once these values are specified, 20,000 synthetic images based on the estimated values are generated. The algorithm for generating those images is presented in [6] and improved in [11]. On the other hand, 3000 real photos are obtained by configuring the transmitter and the receiver with the corresponding target values used for the validation.

Figure 3 shows the experimental setup used to obtain the images used in the validation. Table I includes the experiment's key parameters. The network model structure is included in Annex I [16] for length reasons.

In the following stage, the estimator is trained using transfer learning from the previous equalizer block as detailed

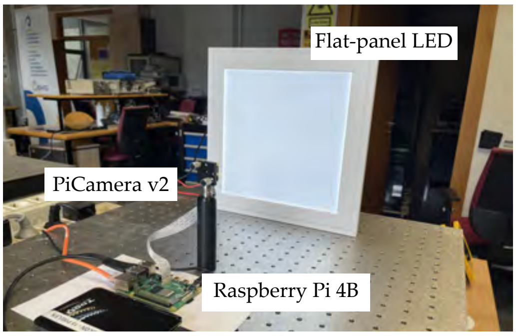


Fig. 3. Experimental setup.

TABLE I
EXPERIMENT'S KEY PARAMETERS.

Parameter	*Value*
Camera	
Hardware	Sony IMX219 [17]
Aperture lens (Focal length)	f/2 (3 mm)
Image resolution	1920x1080 pixels
Row sampling period, T_s	18.9035e-6
Exposure time (Target ESR)	198 μs ($\approx$ 2.62)
Transmitter	
Flat-panel LED	Eglo Tunable White - RGB connect. Cold white (6500K)
Tx symbol time (Target NPPS)	74.614 μs (4.0)
Header, data, stuffed, and trailer bits	5 (ones), 36, 12, 1 (zero)
Dataset size	3000 images
Proccesing plaftorm	
Computer CPU	Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz 2.21 GHz
RAM	16,0 GB
GPU	Geforce GTX 2070. Code not optimized for CUDA acceleration (operation at 10%)

in Section II. For this training, different subsets of the dataset are used. By varying the size of the training dataset (i.e. randomly selecting a fixed amount of images from the original dataset), it is expected to determine the least amount of images required without affecting the reception performance. Then, the network weights for each training epoch are stored. The network parameters trained in this step correspond only to the decoder part of the equalizer since the encoder network parameters are transferred from the estimator model and frozen. The training loss cost function is the mean squared error cost function. Finally, for each training epoch, the BER is computed with the validation images (i.e. the real images). In addition, the epoch time is measured to compare the transfer learning-based training and the full training of the equalizer.

IV. RESULTS

Figure 4 depicts the equalizer training losses obtained for different dataset sizes, 250, 500 and 1000, labelled as D250, D500 and D1000. Highlight that each dataset is immutable (i.e., the images do not change during the training). This figure demonstrates that the training losses consistently decrease independently of dataset size, which means that the training loss reaches, in all cases, a comparable value after the same number of iterations (not to be confused with the epochs). An iteration is defined as the number of times a batch of images is feedforwarded through the architecture. Therefore, as the batch size is 32, D250 is split into seven batches of 32 images (224 different images per epoch), D500 in 14 (448 images per epoch), and D1000 in 28 (896 images per epoch). Hence, in epoch 40, the network trained with D250 has seen 40 times 224 images (8.960 images), the same number of images as the network trained with D500, has seen in epoch 20 (20 × 448 = 8.960 images). In both cases, the network has reached approximately the same training loss (0.1). In other words, the training reaches practically the same loss after the same number of iterations. This is a promising result since, in advance, reducing the number of samples does not significantly deteriorate the training performance. Whether this behaviour extends to validation remains to be verified with the following figures. To account for a fairer comparison, Fig. 5 provides the BER evolution for each iteration (instead of each epoch). It shows that reducing the dataset size produces a higher variance in the BER obtained after equalization (i.e. BER successfully converges but oscillates between high and low values). Despite these variations, this graph demonstrates that the model successfully converges to the minimum error

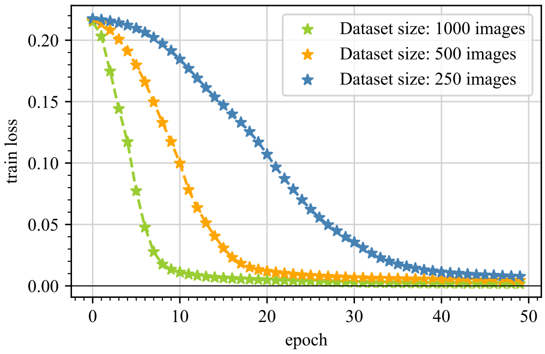


Fig. 4. Equalizer training losses for different dataset sizes.

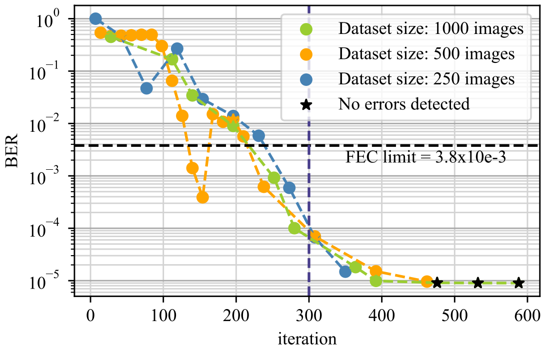


Fig. 5. Equalizer BER vs iterations for different dataset sizes.

independently of the dataset's size (constrained to these limits), approximately after 300 iterations. This implies that the time spent to complete the training (using transfer learning) is practically the same for all the dataset sizes, with the difference of a small overhead almost negligible. The total training time (measured experimentally) is approximately 75 seconds, which compared to a full training (without transfer learning) represents a speedup of roughly 435 times. This value is obtained by dividing the epoch time measured for the training using the dataset D250 (2 seconds) and the epoch time measured for fully training the network (without transfer learning) with a dataset containing 35000 images (870 seconds). Furthermore, the training time also benefits from reducing the time for generating the dataset samples (only 250 images), also reducing the memory requirements to store this temporary synthetic dataset.

V. CONCLUSIONS

This work proposes using transfer learning techniques to enable deep learning equalizers' online and on-demand training. The in-depth knowledge transfer from the estimator to the equalizer block considerably reduces the number of synthetic samples and the time required to train the equalizer while guaranteeing a reception performance with a BER lower than the FEC limit (i.e., 3.8×10^{-3}). The training time (with the proposed transfer learning approach) is roughly reduced 435 times compared to full training. In addition, it has demonstrated that using exclusively 250 synthetic images for training the equalizer, compared to the 35500 images required in [6] (i.e. without transfer learning), the receiver can decode data with a BER lower than 10^{-4}. Using fewer images reduces even further the training as the receiver only has to generate on-demand 250 samples. Furthermore, the receiver achieves this performance under non-ideal conditions, as the training images were adjusted to a sampling frequency and an exposure time with relative errors of 2% and 9%, respectively. Accordingly, it is demonstrated that the receiver can train the equalizer successfully, with just a few images, being robust to the errors obtained at the estimation of the received signal parameters. Consequently, the online and on-demand training of the equalizers can be achieved for multiple configurations, exceptionally relaxing the restrictions imposed on the design of current OCC links, which require cameras with a specified sampling frequency and configured with an extremely short exposure time. In addition, it enables the practical reuse of cameras that simultaneously work as receivers and imaging devices. Ultimately, this work pursues the total decoupling of the reception algorithms from the image capture devices, enabling OCC for its integration into mass-market applications.

REFERENCES

[1] "IEEE Standard for Local and metropolitan area networks–Part 15.7: Short-Range Optical Wireless Communications," *IEEE Std 802.15.7-2018 (Revision of IEEE Std 802.15.7-2011)*, pp. 1–407, 2019.

[2] W. A. Cahyadi, Y. H. Chung, Z. Ghassemlooy, and N. B. Hassan, "Optical camera communications: Principles, modulations, potential and challenges," *Electronics*, vol. 9, no. 9, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/9/1339

[3] N. T. Le, M. Hossain, and Y. M. Jang, "A survey of design and implementation for optical camera communication," *Signal Processing: Image Communication*, vol. 53, pp. 95–109, 2017.

[4] S. A. H. Mohsan, "Optical camera communications: practical constraints, applications, potential challenges, and future directions," *J. Opt. Technol.*, vol. 88, no. 12, pp. 729–741, Dec 2021. [Online]. Available: http://opg.optica.org/jot/abstract.cfm?URI=jot-88-12-729

[5] G. Cossu, A. Sturniolo, and E. Ciaramella, "Modelization and characterization of a cmos camera as an optical real-time oscilloscope," *IEEE Photonics Journal*, vol. 12, no. 6, pp. 1–13, 2020.

[6] C. Jurado-Verdu, V. Guerra, V. Matus, J. Rabadan, and R. Perez-Jimenez, "Convolutional autoencoder for exposure effects equalization and noise mitigation in optical camera communication," *Opt. Express*, vol. 29, no. 15, pp. 22 973–22 991, Jul 2021. [Online]. Available: http://opg.optica.org/oe/abstract.cfm?URI=oe-29-15-22973

[7] X. Li, N. B. Hassan, A. Burton, Z. Ghassemlooy, S. Zvanovec, and R. Perez-Jimenez, "A simplified model for the rolling shutter based camera in optical camera communications," in *2019 15th International Conference on Telecommunications (ConTEL)*. IEEE, 2019, pp. 1–5.

[8] P. Nguyen, N. T. Le, and Y. M. Jang, "Challenges issues for occ based android camera 2 api," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2017, pp. 669–673.

[9] O. I. Younus, N. B. Hassan, Z. Ghassemlooy, P. A. Haigh, S. Zvanovec, L. N. Alves, and H. Le Minh, "Data rate enhancement in optical camera communications using an artificial neural network equaliser," *IEEE Access*, vol. 8, pp. 42 656–42 665, 2020.

[10] J. He, Y. Yang, and J. He, "Artificial neural network-based scheme for 4-pwm occ system," *IEEE Photonics Technology Letters*, pp. 1–1, 2022.

[11] C. Jurado-Verdu, V. Guerra, J. Rabadan, and R. Perez-Jimenez, "Deep learning for signal clock and exposure estimation in rolling shutter optical camera communication," *Opt. Express*, vol. 30, no. 12, pp. 20 261–20 277, Jun 2022. [Online]. Available: http://opg.optica.org/oe/abstract.cfm?URI=oe-30-12-20261

[12] O. I. Younus, N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, L. N. Alves, and H. Le-Minh, "The utilization of artificial neural network equalizer in optical camera communications," *Sensors*, vol. 21, no. 8, p. 2826, 2021.

[13] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. PP, pp. 1–34, 07 2020. [Online]. Available: https://arxiv.org/abs/1911.02685

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241. [Online]. Available: https://arxiv.org/abs/1505.04597

[16] C. J. Verdu, C. Guerra, V. Guerra, J. Rabadan, S. Zvanovec, and R. Perez-Jimenez, "On-demand training of deep learning equalizers for rolling shutter optical camera communications: supplemental material," 3 2022. [Online]. Available: https://doi.org/10.6084/m9.figshare.19287452

[17] Sony Corporation, "IMX219PQH5-C Datasheet," Accessed on: Feb. 10, 2022. [Online]. Available: https://datasheetspdf.com/pdf/1404029/Sony/IMX219PQH5-C/1, 2014.

Appendix C

Repositories

The Python and C++ codes developed in this thesis are compiled in several repositories that can be found in https://gitlab.com/cristojv.