

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y

ELECTRÓNICA



TRABAJO FIN DE GRADO

SISTEMA DE CODIFICACIÓN ESTEGANOGRÁFICO PARA COMUNICACIONES OCC

Titulación: Grado en Ingeniería en Tecnologías de la Telecomunicación

- Autor: Alberto Ramos Monagas
- Tutores: Dr. Rafael Pérez Jiménez

Dr. Víctor Guerra Yánez

Fecha: Julio de 2020



ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y

ELECTRÓNICA

TRABAJO FIN DE GRADO

SISTEMA DE CODIFICACIÓN ESTEGANOGRÁFICO PARA COMUNICACIONES OCC

HOJA DE EVALUACIÓN

Calificación: _____

Presidente

Vocal

Secretario/a

Fecha: Julio de 2020

RESUMEN

Los recientes esfuerzos en comunicaciones por luz visible a través de enlaces screen-to-camera (S2C) han explotado los recursos de la pantalla para la comunicación de datos. Estas prácticas, aunque convenientes, han dado lugar a controversias entre el espacio asignado a los usuarios y el contenido reservado a la comunicación con dispositivos, además de su falta de estética visual y de distracción. En este documento se propone un sistema que permite la comunicación simultánea, de modo dual (entre la cámara y la pantalla), tanto para los usuarios como para los dispositivos sin que sea perceptible ninguna imagen codificada (por ejemplo, códigos de barras). El sistema propuesto codifica los datos en píxeles translúcidos sobre cualquier tipo de contenido de la pantalla utilizando el canal alfa, de modo que cualquier dispositivo equipado con una cámara puede obtener los datos volteando su cámara hacia la pantalla. Además, el aprovechamiento del canal alfa permite codificar los bits en el cambio de translucidez de los píxeles, utilizando un esquema diferencial. Al eliminar la necesidad de modificar directamente los valores RGB de los píxeles, se permite una comunicación discreta en tiempo real, a la vez que se soporta cualquier tipo de contenido en la pantalla. En definitiva, el sistema diseñado aprovecha la propiedad de fusión de parpadeo espaciotemporal de la visión humana y la rápida velocidad de frame de la pantalla moderna, multiplexando los datos en el contenido del vídeo, fotograma a fotograma, a través de una composición de fotogramas complementarios y una modulación similar a la de la CDMA (del inglés Code Division Multiple Access), basada en técnicas de espectro ensanchado. De este modo se garantiza una comunicación de datos discreta entre la cámara y la pantalla, sin que ello afecte a la calidad de la experiencia de usuario.

PALABRAS CLAVE

Comunicación pantalla-cámara; Comunicación oculta visible; Comunicación con luz visible; Canal alfa; Comunicación visible de modo dual; Vídeo de fotograma completo; Comunicación por cámara óptica.

ABSTRACT

Recent efforts in Visible Light Communications (VLC) through screen-to-camera (S2C) links have exploited the resources of the screen for data communication. These practices, while convenient, have led to controversies between the space assigned to users and the content reserved for the devices, in addition to their poor aesthetics. In this document, a system that allows simultaneous, dual-mode communication (between the camera and the screen) for both users and devices without displaying any encoded images (e.g., bar codes) is proposed. The system encodes the data in translucent pixels on any type of content using the alpha channel, making any device equipped with a camera capable of obtaining the data by turning its camera towards the screen. In addition, the use of the alpha channel is used to encode the bits in the change of translucency of the pixels, using a differential scheme. Eliminating the need to directly modify the RGB values of the pixels allows for discreet real-time communication, while supporting any type of content on the screen. In short, the designed system takes advantage of the space-time flicker fusion property of the human vision and the fast frame rate of the modern screens. It multiplexes the data in the video content, frame by frame, through a composition of complementary frames and a modulation similar to CDMA (Code Division Multiple Access), based on spread spectrum techniques. This ensures covert data communication between the camera and the screen without affecting the Quality of Experience (QoE).

KEY WORDS

Screen-camera communication; Hidden visible communication; Visible ligth communication; Alpha cannel; Dual-mode visible communication; Full-frame video; Optical Camera Communication.

Índice

CAP	CAPÍTULO 1: INTRODUCCIÓN						
1.1.	Antecedentes	16					
1.2.	Justificación e importancia	17					
1.3.	Objetivos y alcance del proyecto	18					
1.4.	Medios utilizados	19					
1.5.	Organización de la memoria						
CAP	ÍTULO 2: MODELO DE LA PERCEPCIÓN VISUAL HUMANA	22					
2.1.	Introducción	22					
2.2.	Espectro visible	23					
2.3.	El sistema visual humano (SVH)	24					
2.4.	Factores que determinan la sensibilidad al contraste	27					
2.5.	Sensibilidad temporal	29					
CAP	ÍTULO 3: EMISORES Y RECEPTORES PARA OCC	31					
3.1.	Introducción	31					
3.2.	Pantallas	31					
3.3.	Cámaras	48					
CAP	ÍTULO 4: SISTEMAS ESTEGANOGRÁFICOS	59					
4.1.	Introducción	59					
4.2.	Esteganografía en la era digital	60					
4.3.	Estegosistema: descripción y características	61					
		6					

4.4.	Introducción a los métodos y técnicas esteganográficas empleadas	64					
4.5.	Técnicas de espectro ensanchado para la esteganografía con imágenes (SSIS)						
CAPÍ	TULO 5: SISTEMAS DE TRANSMISIÓN SCREEN2CAMERA	69					
5.1.	Introducción	69					
5.2.	Antecedentes de las OCC	70					
5.3.	Trabajos previos en sistemas Screen-to-Camera	74					
5.4.	Idea básica del modelo de estudio						
CAPÍ	ΓULO 6: DISEÑO DEL SISTEMA DE TRANSMISIÓN	79					
6.1.	Justificación y objetivos del modelo	79					
6.2.	Limitaciones y soluciones del modelo	81					
6.3.	Descripción del algoritmo propuesto	84					
6.4.	Arquitectura del transmisor	88					
6.5.	Arquitectura del receptor	94					
6.6.	Implementación						
CAPÍ	ΓULO 7: VALIDACIÓN EXPERIMENTAL Y RESULTADOS	97					
7.1.	Objetivos de la validación experimental	97					
7.2.	Configuración experimental	97					
7.3.	Descripción de los experimentos	98					
7.3.1.	Sostenibilidad de la QoE	99					
7.3.2.	Mejora de la QoS	100					
7.4.	Análisis de datos	100					
7.4.1.	Test de hipótesis sobre la perceptibilidad de los códigos	100					

7.4.2.	Análisis de detectabilidad	101
7.5.	Resultados	101
7.5.1.	Resultados de perceptibilidad	101
7.5.2.	Resultados de QoS	105
CAPÍ	FULO 8: CONCLUSIONES Y RECOMENDACIONES	122
CAPÍ	TULO 9: PRESUPUESTO	124
9.1.	Recursos Materiales	124
9.2.	Trabajo tarifado por tiempo empleado	125
9.3.	Derechos del visado del COITT	126
9.4.	Gastos de tramitación y envío	127
9.5.	Aplicación de impuestos y presupuesto final	127
BIBLI	OGRAFÍA REFERENCIADA	129
ANEX	iOS	138
A.1.	Métricas obtenidas del vídeo 1 ("sea.mp4").	138
A.2.	Métricas obtenidas del vídeo 2 ("flower.mp4").	152
A.3.	Métricas obtenidas del vídeo 3 ("walk.mp4").	166
A.4.	Métricas obtenidas del vídeo 4 ("birds.mp4").	180

Índice de figuras

Figura 1. Espectro de ondas electromagnéticas.	23
Figura 2. Función de sensibilidad al contraste	29
Figura 3. (a) CSFs para velocidades de 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64 y 128 grados (de	recha
a izquierda). (b) La frecuencia espacial detectable más alta posible en función	de la
velocidad	30
Figura 4. Estructura interna de una pantalla de cristal líquido	33
Figura 5. Formación de la imagen en una pantalla LCD.	34
Figura 6. Disposición de las tiras conductoras en una matriz pasiva	35
Figura 7. Disposición de las tiras conductoras en una matriz activa.	35
Figura 8. Diagrama de una pantalla de plasma.	36
Figura 9. Funcionamiento de una pantalla de plasma	37
Figura 10. Diagrama de bloques de una pantalla LED.	38
Figura 11. Esquemas de Edge y Full LED	39
Figura 12. Base de funcionamiento de la tecnología OLED	40
Figura 13. Esquema de las principales pantallas en el mercado	42
Figura 14. Formación de una imagen mediante técnicas de barrido entrelazado	45
Figura 15. Funcionamiento del barrido progresivo y del entrelazado.	46
Figura 16. Diferencia entre el uso de barrido progresivo y entrelazado	47
Figura 17. Diagrama de bloques del sistema óptico.	48
Figura 18. Diagrama de un sensor de imagen CCD	49
Figura 22. Funcionamiento de filtro Bayer en procesado de imágenes	50
Figura 23. Filtro de Bayer	51
Figura 24. Estructura de los sensores CMOS y CCD.	54
Figura 25. Consideraciones de los sensores CMOS Y CCD.	54
Figura 26. Aplicaciones de los sensores CMOS y CCD	55
Figura 27. Luminosidad en los sensores CMOS y CCD	55
Figura 28. Intensidad media (Strobe flash) para sensores CMOS y CCD.	56
Figura 29. Cambios de formatos para sensores CMOS Y CCD.	57
Figura 30. Defectos de los píxeles en sensores CMOS y CCD.	57
Figura 31. Pasos de inserción y extracción del mensaje en un sistema esteganográfico	62
Figura 32. Proceso de codificación SSIS.	67
Figura 33. Proceso de decodificación SSIS.	68

Figura 34. Concepto de multiplexación de fotograma completo: Screen-to-eye para videos y
Screen-to-camera para la comunicación de datos a través del mismo canal visible
simultáneamente70
Figura 35. Categorías de OCC72
Figura 36. Estructura del receptor de la cámara óptica, procedimiento de comunicación y
aplicaciones73
Figura 37. Las prácticas actuales de comprometer las comunicaciones de Screen-to-eye y de
Screen-to-Camera. a) Un pequeño código QR en la esquina (en el caso de un partido de
hockey sobre hielo); b) Dos frames consecutivos, uno para cada canal79
Figura 38. Ilustración de la distorsión de la perspectiva en el canal screen-camera. Los píxeles de
la pantalla visualizados son borrosos y se reduce el tamaño en la vista frontal completa y
también en forma en la vista angular
Figura 39. Diseño de nuestro sistema. Usamos una pantalla de 60 Hz y vídeos de 50 FPS. Vi y Di
representan el i-ésimo frame del vídeo original y de datos84
Figura 40. Modelamiento de la forma de onda del suavizado temporal
Figura 41. El impacto de cambiar los valores de los píxeles de α . (a) Muestra una imagen con
diferentes valores de a: el valor predeterminado 1 (la izquierda), 0,99 (el centro) y 0,5 (la
derecha), utilizando un Samsung Note 3; (b) Muestra la intensidad de color del cuadro
capturado por el Samsung S5 al variar los valores de α en la pantalla del transmisor87
Figura 42. Diagrama de bloques del sistema transmisor
Figura 43. Tipos de localizadores usados en las pruebas94
Figura 44. Diagrama de bloques del sistema receptor94
Figura 45. Diagrama del flujo de trabajo de la algoritmia utilizada
Figura 46. Vídeos utilizados para codificar
Figura 47. SIR obtenida para una longitud de símbolo 7 y valor de alpha 3106
Figura 48. SIR obtenida para una longitud de símbolo 27 y valor de alpha 10106
Figura 49. Vídeo "flower" codificado con valores de alpha a) 3 y b) 10111
Figura 50. Vídeo "sea" codificado con valores de alpha a) 3 y b) 10111
Figura 51. Vídeo "birds" codificado con valores de alpha a) 3 y b) 10
Figura 52. Vídeo "flower" codificado con valores de alpha a) 3 y b) 10111
Figura 53. Métricas obtenidas en función de la SIR del vídeo "birds" con alpha 3 y longitud de
código 7113
Figura 54. Métricas obtenidas en función de la SIR del vídeo "sea" con alpha 3 y longitud de
código 7113
Figura 55. Métricas obtenidas en función de alpha del vídeo "birds" con SIR al 10% y longitud
de código 7114

Figura 56. Métricas obtenidas en función de alpha del vídeo "sea" con SIR al 10% y longitud de
código 7115
Figura 57. Métricas obtenidas en función de la longitud del símbolo del vídeo "birds" con SIR al
10% y alpha 3116
Figura 58. Métricas obtenidas en función de la longitud del símbolo del vídeo "birds" con SIR al
10% y alpha 3116
Figura 59. Métricas obtenidas en función de la SIR del vídeo "walk" con alpha 3 y longitud de
código 7117
Figura 60. Métricas obtenidas en función de la SIR del vídeo "flower" con alpha 3 y longitud de
código 7
Figura 61. Métricas obtenidas en función de alpha del vídeo "walk" con SIR al 10% y longitud
de código 7119
Figura 62. Métricas obtenidas en función de alpha del vídeo "flower" con SIR al 10% y longitud
de código 7119
Figura 63. Métricas obtenidas en función de la longitud del símbolo del vídeo "walk" con SIR al
10% y alpha 3120
Figura 64. Métricas obtenidas en función de la longitud del símbolo del vídeo "walk" con SIR al
10% y alpha 3121
Figura 65. Métrica con longitud del símbolo 7, alpha 3 y SIR al 10%138
Figura 66. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%139
Figura 67. Métrica con longitud del símbolo 7, alpha 3 y SIR al 90%139
Figura 68. Métrica con longitud del símbolo 7, alpha 5 y SIR al 10%140
Figura 69. Métrica con longitud del símbolo 7, alpha 5 y SIR al 50%141
Figura 70. Métrica con longitud del símbolo 7, alpha 5 y SIR al 90%141
Figura 71. Métrica con longitud del símbolo 7, alpha 10 y SIR al 50%142
Figura 72. Métrica con longitud del símbolo 7, alpha 10 y SIR al 10%142
Figura 73. Métrica con longitud del símbolo 7, alpha 10 y SIR al 90%143
Figura 74. Métrica con longitud del símbolo 14, alpha 3 y SIR al 10%143
Figura 75. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%144
Figura 76. Métrica con longitud del símbolo 14, alpha 3 y SIR al 50%144
Figura 77. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%145
Figura 78. Métrica con longitud del símbolo 14, alpha 5 y SIR al 10%145
Figura 79. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%146
Figura 80. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%146
Figura 81. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%147
Figura 82. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%147
Figura 83. Métrica con longitud del símbolo 27, alpha 3 y SIR al 10%148

Figura 84. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%	
Figura 85. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%	149
Figura 86. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%	149
Figura 87. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%	150
Figura 88. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%	150
Figura 89. Métrica con longitud del símbolo 27, alpha 10 y SIR al 50%	151
Figura 90.Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%	151
Figura 91. Métrica con longitud del símbolo 27, alpha 10 y SIR al 90%	
Figura 92. Métrica con longitud del símbolo 7, alpha 3 y SIR al 10%	
Figura 93. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%	
Figura 94. Métrica con longitud del símbolo 7, alpha 3 y SIR al 90%	
Figura 95. Métrica con longitud del símbolo 7, alpha 5 y SIR al 10%	154
Figura 96. Métrica con longitud del símbolo 7, alpha 5 y SIR al 50%	154
Figura 97. Métrica con longitud del símbolo 7, alpha 10 y SIR al 10%	
Figura 98. Métrica con longitud del símbolo 7, alpha 5 y SIR al 90%	155
Figura 99. Métrica con longitud del símbolo 7, alpha 10 y SIR al 90%	156
Figura 100. Métrica con longitud del símbolo 7, alpha 10 y SIR al 50%	156
Figura 101. Métrica con longitud del símbolo 14, alpha 3 y SIR al 50%	157
Figura 102. Métrica con longitud del símbolo 14, alpha 3 y SIR al 10%	157
Figura 103. Métrica con longitud del símbolo 14, alpha 5 y SIR al 10%	158
Figura 104. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%	
Figura 105. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%	
Figura 106. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%	
Figura 107. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%	
Figura 108. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%	
Figura 109. Métrica con longitud del símbolo 27, alpha 3 y SIR al 10%	
Figura 110. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%	
Figura 111. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%	
Figura 112. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%	
Figura 113. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%	
Figura 114. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%	
Figura 115. Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%	164
Figura 116. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%	
Figura 117. Métrica con longitud del símbolo 27, alpha 10 y SIR al 90%	
Figura 118. Métrica con longitud del símbolo 27, alpha 10 y SIR al 50%	
Figura 119. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%	
Figura 120. Métrica con longitud del símbolo 7, alpha 3 y SIR al 10%	

Figura	121.	Métrica	con	longitud d	el símbolo	7, alpha 5	5 y SIR al	10%	 167
Figura	122.	Métrica	con	longitud d	el símbolo	7, alpha 3	y SIR al	90%	 167
Figura	123.	Métrica	con	longitud d	el símbolo	7, alpha 5	5 y SIR al	90%	 168
Figura	124.	Métrica	con	longitud d	el símbolo	7, alpha 5	5 y SIR al	50%	 168
Figura	125.	Métrica	con	longitud d	el símbolo	7, alpha 1	0 y SIR a	1 50%	 169
Figura	126.	Métrica	con	longitud d	el símbolo	7, alpha 1	0 y SIR a	1 10%	 169
Figura	127.	Métrica	con	longitud d	el símbolo	14, alpha	3 y SIR a	1 10%	 170
Figura	128.	Métrica	con	longitud d	el símbolo	7, alpha 1	0 y SIR a	1 90%	 170
Figura	129.	Métrica	con	longitud d	el símbolo	14, alpha	3 y SIR a	1 90%	 171
Figura	130.	Métrica	con	longitud d	el símbolo	14, alpha	3 y SIR a	1 50%	 171
Figura	131.	Métrica	con	longitud d	el símbolo	14, alpha	5 y SIR a	1 50%	 172
Figura	132.	Métrica	con	longitud d	el símbolo	14, alpha	5 y SIR a	1 10%	 172
Figura	133.	Métrica	con	longitud d	el símbolo	14, alpha	10 y SIR	al 10%	 173
Figura	134.	Métrica	con	longitud d	el símbolo	14, alpha	5 y SIR a	1 90%	 173
Figura	135.	Métrica	con	longitud d	el símbolo	14, alpha	10 y SIR	al 90%	 174
Figura	136.	Métrica	con	longitud d	el símbolo	14, alpha	10 y SIR	al 50%	 174
Figura	137.	Métrica	con	longitud d	el símbolo	27, alpha	3 y SIR a	1 50%	 175
Figura	138.	Métrica	con	longitud d	el símbolo	27, alpha	3 y SIR a	1 10%	 175
Figura	139.	Métrica	con	longitud d	el símbolo	27, alpha	5 y SIR a	1 10%	 176
Figura	140.	Métrica	con	longitud d	el símbolo	27, alpha	3 y SIR a	1 90%	 176
Figura	141.	Métrica	con	longitud d	el símbolo	27, alpha	5 y SIR a	1 90%	 177
Figura	142.	Métrica	con	longitud d	el símbolo	27, alpha	5 y SIR a	1 50%	 177
Figura	143.	Métrica	con	longitud d	el símbolo	27, alpha	10 y SIR	al 50%	 178
Figura	144.	Métrica	con	longitud d	el símbolo	27, alpha	10 y SIR	al 10%	 178
Figura	145.	Métrica	con	longitud d	el símbolo	27, alpha	10 y SIR	al 90%	 179
Figura	146.	Métrica	con	longitud d	el símbolo	7, alpha 3	y SIR al	10%	 180
Figura	147.	Métrica	con	longitud d	el símbolo	7, alpha 3	s y SIR al	90%	 181
Figura	148.	Métrica	con	longitud d	el símbolo	7, alpha 3	3 y SIR al	50%	 181
Figura	149.	Métrica	con	longitud d	el símbolo	7, alpha 5	5 y SIR al	50%	 182
Figura	150.	Métrica	con	longitud d	el símbolo	7, alpha 5	5 y SIR al	10%	 182
Figura	151.	Métrica	con	longitud d	el símbolo	7, alpha 1	0 y SIR a	1 10%	 183
Figura	152.	Métrica	con	longitud d	el símbolo	7, alpha 5	5 y SIR al	90%	 183
Figura	153.	Métrica	con	longitud d	el símbolo	7, alpha 1	0 y SIR a	1 90%	 184
Figura	154.	Métrica	con	longitud d	el símbolo	7, alpha 1	0 y SIR a	1 50%	 184
Figura	155.	Métrica	con	longitud d	el símbolo	14, alpha	3 y SIR a	1 50%	 185
Figura	156.	Métrica	con	longitud d	el símbolo	14, alpha	3 y SIR a	1 10%	 185
Figura	157.	Métrica	con	longitud d	el símbolo	14, alpha	5 y SIR a	1 10%	 186

Figura	158. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%	186
Figura	159. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%	187
Figura	160. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%	187
Figura	161. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%	188
Figura	162. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%	188
Figura	163. Métrica con longitud del símbolo 27, alpha 3 y SIR al 10%	189
Figura	164. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%	189
Figura	165. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%	190
Figura	166. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%	190
Figura	167. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%	191
Figura	168. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%	191
Figura	169. Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%	192
Figura	170. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%	192

Índice de tablas

Tabla 1. Ventajas e inconvenientes de las pantallas LCD.	42
Tabla 2. Ventajas y desventajas de los televisores de plasma	43
Tabla 3. Comparación de OCC y VLC	72
Tabla 4. Datos recogidos del estudio realizado	
Tabla 5. Resultados obtenidos para los vídeos de "sea"	
Tabla 6. Resultados obtenidos para los vídeos de "walk"	
Tabla 7. Resultados obtenidos para los vídeos de "birds"	104
Tabla 8. Resultados obtenidos para los vídeos de "flower"	105
Tabla 9. Valores de SIR obtenidos para el vídeo "sea"	
Tabla 10. Valores de SIR obtenidos para el vídeo "flower"	
Tabla 11. Valores de SIR obtenidos para el vídeo "birds"	
Tabla 12. Valores de SIR obtenidos para el vídeo "walk"	
Tabla 13. Recursos hardware	
Tabla 14. Factor de corrección en función del número de horas invertidas	
Tabla 15. Presupuesto de ejecución material	
Tabla 16. Costes totales del proyecto	

1.1. Antecedentes

En la actualidad, el aumento de los dispositivos conectados a Internet, promovido por la creciente expansión del Internet de las Cosas (IoT, del inglés *Internet of Things*), ha generado una gran demanda de comunicaciones inalámbricas, así como la progresiva saturación del espectro radioeléctrico, lo que a día de hoy constituye un problema real [1]. Este hecho ha obligado a la búsqueda de soluciones alternativas que alivien la carga de los canales de radio convencionales al tiempo que se aseguren velocidades de transmisión altas. Aparecen así las Comunicaciones Ópticas Inalámbricas (OWC), que se posicionan como una tecnología alternativa en el sector de las comunicaciones. En concreto, dentro del campo de las Comunicaciones Ópticas, se sitúa la rama de las Comunicaciones por Luz Visible (VLC) [2], que utiliza el amplio espectro no regulado de la luz visible. El ánimo principal de esta tecnología consiste en utilizar la actual la infraestructura de luminarias LED para la transmisión de información, a la par que preservan su función principal como sistema de iluminación. En lo que respecta a los sistemas receptores, se utilizan comúnmente uno o varios fotorreceptores independientes.

El interés por esta tecnología es tal que el *Institute of Electrical and Electronics Engineers* (IEEE) posee un grupo asignado, el IEEE 802.15.7, para la regulación y normalización de estos sistemas, así como sus aplicaciones [3]. A esta rama tecnológica se le conoce como *Optical Camera Communications* (OCC) [4]. Los sistemas OCC comparten con cualquier sistema de comunicación los bloques generales de transmisor, canal y receptor.

En definitiva, las OCC han aumentado su relevancia dentro del campo de las VLC debido a la amplia disponibilidad de cámaras CMOS en un gran número de dispositivos (teléfonos inteligentes, tabletas, sistemas de vigilancia, ...). La mayoría de estas cámaras se basan en la técnica de persiana o *Rolling Shutter* (RS), que implementa un proceso de escaneo del sensor fila por fila. Por lo tanto, una fuente de luz LED, que cambie a una frecuencia superior a la velocidad de obturación, aparecerá en la imagen como una serie de rayas oscuras y brillantes, representando los datos binarios. Por lo tanto, esta arquitectura de la cámara proporciona a los sistemas OCC una velocidad de datos muy superior a la velocidad de *frame* (fps) [5,6].

Con la creciente popularidad de los dispositivos inteligentes en nuestra vida diaria, la gente depende cada vez más de ellos para recuperar y distribuir información en el mundo del IoT. Mientras tanto, muchos medios de difusión equipados con pantallas y cámaras, por ejemplo,

pantallas de publicidad, televisores, ordenadores de escritorio o portátiles, teléfonos inteligentes y tabletas, se han convertido en una fuente de información fácilmente disponible para los usuarios. En el contexto de esta tendencia, es deseable una comunicación de doble canal, en la que las pantallas y las cámaras transmitan información primaria a través de vídeos a los usuarios, mientras la información es captada por la cámara de un *smartphone* o tableta. Por ejemplo, un aficionado a los deportes que ve un partido en directo de la NBA en la pantalla del estadio podría recibir las estadísticas de los jugadores y equipos en su dispositivo inteligente sin recurrir a la Internet. Otro ejemplo podría ser el de una persona que ve anuncios en la televisión mientras recibe notificaciones instantáneas, ofertas y promociones en su dispositivo.

1.2. Justificación e importancia

Las técnicas de comunicación *Screen-to-Camera* están cobrando gran importancia dada la ubicuidad de los *displays* LED. Dichas técnicas surgen como un nuevo paradigma donde se incrustan los datos de manera imperceptible en vídeos. El Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones (IDeTIC) y en concreto la División Tecnología Fotónica y Comunicaciones ha incluido estos sistemas dentro de sus áreas de interés. Actualmente se está llevando a cabo una tesis en este campo y un proyecto con financiación privada.

A lo largo de este proyecto se pretenderá evaluar la capacidad de un sistema de comunicación OCC, donde la señal transmitida se encuentra incrustada en las imágenes de vídeo emitidas por un televisor, mientras que la recepción será realizada mediante una cámara integrada en un dispositivo móvil.

Al tratarse de una tecnología altamente novedosa, no hay muchas referencias sobre las técnicas a utilizar y las prestaciones y capacidades de este tipo de sistemas. En esta propuesta se pretenden analizar y buscar diferentes opciones y estudiar cuáles son sus características y parámetros más importantes.

Diversos autores propusieron [7] un sistema en tiempo real basado en el canal alfa (transparencia) para introducir información espacialmente distribuida sobre cualquier imagen. El algoritmo sugerido por los autores dividía la pantalla en una retícula uniforme y decidía, en función de la secuencia de *frame*, qué cantidad de información incluir o, incluso si descartar incrustar información (*frames* negros, por ejemplo) [8]. Otro ejemplo de esteganografía en vídeo es el uso de la transformada discreta del coseno (DCT del inglés *Discrete Fourier Transform*). Sin embargo, esta técnica ampliamente utilizada parte de la base de que tanto el emisor del mensaje esteganográfico como el receptor, disponen del fichero sin pérdidas [9]. En el caso que nos ocupa,

el canal de transmisión modificaría la información al tratarse de una captura *in situ* del vídeo con información imperceptible en reproducción.

En esta propuesta se pretenden analizar y proponer diferentes opciones y estudiar cuáles son sus características y parámetros más importantes. El modelo que se va a evaluar consta de los siguientes componentes:

- Sistema de transmisión OCC basado en vídeo. Consiste en una pantalla de televisión convencional que emite imágenes de un vídeo pre-procesado al que se le ha aplicado técnicas esteganográficas para embeber la señal de datos. Dicha integración debe permitir incorporar la señal de información de tal manera que no se altere de forma apreciable por el usuario la calidad de imagen en el televisor. En este sentido se tendrá en cuenta, en primer lugar, las posibles propiedades de la imagen (iluminación, contraste...) susceptibles de ser modificadas y los límites de dichas modificaciones que aseguren el mantenimiento de la calidad de las imágenes. A continuación, se procederá al estudio de las posibles codificaciones o modulaciones a realizar sobre dichas propiedades.
- Sistema de recepción OCC. Aunque el objeto de este proyecto no es el receptor, es conveniente conocerlo para trabajar con él. Consiste en una cámara de un móvil convencional que realizará la captación y procesado de la imagen de vídeo. Este tratamiento digital permitirá extraer la información embebida en la imagen. En este punto habrá que tener en cuenta aspectos tales como la zona de imagen a analizar o la complejidad computacional del procesado de señales a implementar.

Como resultado de este proyecto, se pretenden analizar las posibles soluciones de transmisión clasificándolas en función de su complejidad y prestaciones. Asimismo, las técnicas de detección asociadas y sus parámetros más relevantes como velocidad de transmisión, distancia o tiempo de adquisición de datos serán igualmente estudiadas.

1.3. Objetivos y alcance del proyecto

Realizando una investigación e implementación del presente proyecto con la herramienta de desarrollo MATLAB® se llegará a determinar el diseño de un aplicativo que permitirá efectuar la codificación esteganográfica del vídeo para llevar a cabo una comunicación síncrona entre transmisor y receptor.

Se realizará el análisis de los factores que influyen en la percepción visual humana, que serán de vital importancia en el próximo punto a tratar, el sistema de transmisión *Screen-to-Camera* (también conocido como S2C, nombre que recibirá a lo largo de este documento). En el siguiente punto anteriormente mencionado se efectuará un análisis de los S2C utilizados en la actualidad que serán de gran ayuda a la hora de diseñar nuestro sistema desde el punto de vista técnico. Además, se hará un especial hincapié en la forma en la que los datos irán codificados, por lo que se deberá realizar un estudio de las técnicas de codificación empleadas en dichos sistemas para obtener un esquema de codificación robusto frente a errores e interferencias. Se incluirá además una breve explicación de las características de los estegosistemas que nos permitirán conocer qué conjunto de métodos pueden ser utilizados para ocultar información.

En relación a los objetivos, el presente proyecto tiene como meta el estudio de los sistemas de codificación esteganográficos para las comunicaciones OCC.

Los objetivos fundamentales que se han fijado para este Trabajo Fin de Grado han sido:

- O1. Estudio de los sistemas S2C.
- **O2.** Evaluación sistemas de codificación que permitan utilizar una televisión/pantalla como fuente de datos sin afectar a la percepción visual.
- O3. Diseño y evaluación de un sistema de test.

1.4. Medios utilizados

Para la realización del código de este proyecto únicamente fue necesario disponer de un ordenador portátil con conexión a internet y del entorno MATLAB. Por el contrario, desde el punto de vista de las comunicaciones ópticas, ha sido necesaria la utilización de una pantalla que visualice el vídeo con el código incrustado (puede ser la misma pantalla del ordenador, tableta, televisor, ...) y un móvil con un trípode que capture en forma de grabación dicho vídeo contenido con el fin de comprobar que en el receptor se ha realizado correctamente la comunicación.

1.5. Organización de la memoria

Este documento se estructurará en siete capítulos, siendo el primero el actual, en el que se a cabo una introducción inicial, a modo de motivación, para comprender los motivos que han llevado a la elaboración de este proyecto, además de definirse la importancia y necesidad del tema que se aborda, y se dejarán explícitos los elementos de la investigación y diseño. Los capítulos sucesivos abarcan contenidos bastante diferenciados, los cuales son comentados a continuación.

- **Capítulo 2**: En el capítulo 2 se plasma el estado del arte de la percepción visual humana, describiendo el sistema visual humano desde el punto de vista biológico y físico, destacándose especialmente los aspectos referentes al enmascaramiento visual y al proceso de visión humana. Asimismo, se pretende realizar una breve comparación entre los sistemas de pantalla actuales y cámaras modernas con el ojo humano.
- **Capítulo 3**: El capítulo 3 recoge un estudio del estado del arte de los principios fundamentales de las pantallas y de las cámaras, explicando distintos tipos existentes y sus características.
- **Capítulo 4**: El capítulo 4 recoge un estudio del estado del arte de la esteganografía resaltando su aplicación en diversas ramas de la ciencia y la técnica. Se hace un análisis de los estegosistemas, caracterizándolos y clasificando los procesos y esquemas que intervienen, particularmente en el esteganografiado digital con imágenes.
- Capítulo 5: En el capítulo 4 se describirá el estado del arte de los S2C, aplicándole especial atención a las comunicaciones OCC que serán fundamentales en el diseño de nuestro sistema. Además, se describirán los principales objetivos del sistema a tratar junto con los principales problemas e inconvenientes del uso de esta tecnología, así como las soluciones planteadas.
- Capítulo 6: Este capítulo documentará el diseño de un método esteganográfico implementado en MATLAB, que pone en práctica la teoría matemática para el incrustado de la información. El algoritmo resultante se validará en los siguientes capítulos, donde se determinarán los logros alcanzados en cuanto a la perceptibilidad, señalando además el incremento de la robustez y seguridad del esteganograma obtenido en el proceso de codificación. Del mismo modo, en este capítulo se describirá el diseño del transmisor y del receptor de nuestro sistema, así como las decisiones tomadas en cada uno de ellos.
- Capítulo 7: A lo largo de este capítulo tendrá lugar una descripción del sistema implementado, utilizado para la realización de pruebas de funcionamiento y validación. Una vez hechas dichas pruebas, se expondrán los resultados obtenidos.

- **Capítulo 8**: Las conclusiones reflejarán los logros obtenidos con el diseño de S2C empleado, a partir de un análisis de los resultados y de los experimentos probados. Las recomendaciones se harán teniendo en cuenta las limitantes del método descrito, para su posterior perfeccionamiento por otros diseñadores e investigadores del tema; así como para el enriquecimiento de algunas de las temáticas abordadas.
- **Capítulo 9:** En este capítulo se expondrá de manera desglosada el presupuesto de este proyecto siguiendo las directrices técnicas del COITT (Colegio Oficial de Ingenieros Técnicos de Telecomunicación).

Además, se incluirán las referencias bibliográficas haciendo un listado, de toda la bibliografía consultada y siguiendo las normas establecidas. Por otro lado, la totalidad del código se encontrará en el apartado del anexo (digital).

2.1.Introducción

El sistema visual humano (SVH) se ha desarrollado y perfeccionado tras millones de años de adaptación al medio que lo rodea, permitiendo a los seres humanos interactuar de manera efectiva y precisa con su hábitat. El espectro visible, la visión central y periférica, las diferentes clases de células fotorreceptoras, y la visión binocular son algunas de estas adaptaciones. A lo largo de los años, los científicos han intentado definir modelos que expliquen lo mejor posible el funcionamiento del SVH. Un modelo muy conocido sobre la capacidad de la percepción de detalles en la visión es la *Función de Sensibilidad al Contraste* (del inglés *Contrast Sensitivity Function – CSF*), definida entre otros por Mannos y Sakrison [10], esta función intenta modelar la capacidad del sistema de visión de distinguir las diferentes frecuencias espaciales en función del contraste para un estímulo dado. Este modelo ha sido estudiado y. modificado a lo largo de los años, para contemplar, por ejemplo, variables como la velocidad y excentricidad.

El SVH tiene muchas limitaciones que pueden ser explotadas a la hora de representar diversas aplicaciones. El mundo real no es exactamente como lo percibimos, y aunque el proceso óptico y físico del sistema visual funciona prácticamente igual en todos los individuos, la percepción visual es mucho más compleja, puesto que consiste en la interpretación del estímulo recibido, y esta interpretación depende, en parte, de cada individuo.

Adquirimos conciencia del mundo que nos rodea a través de los sentidos. Los estímulos desencadenan sensaciones, pero la organización, interpretación y análisis de estas no depende exclusivamente de los sentidos, sino también del cerebro. A partir de los estímulos recogidos por los sentidos organizamos y recreamos la realidad y adquirimos conciencia de ella por medio de la percepción. El estímulo pertenece al mundo exterior y causa un efecto o sensación, mientras que la percepción es el proceso psicológico de la interpretación y depende, en gran parte, del «mundo interior» de cada individuo.

Si limitamos el estudio al campo visual, la percepción se define por el estímulo que produce la luz que, a su vez, nos crea una sensación que es analizada e interpretada en nuestro cerebro. Aunque el acto perceptivo tenga lugar de forma automática, es realmente complejo y tiene múltiples implicaciones. El mundo real no tiene que ser exactamente lo que percibimos por los ojos.

Finalmente, el proceso óptico y físico de la percepción visual, como se verá en los siguientes apartados, funciona mecánicamente de modo parecido en todos los humanos de vista sana. Las diferencias fisiológicas de los órganos visuales prácticamente no afectan al resultado de la percepción humana. Estas diferencias empiezan con la interpretación de la información que se recibe. Es decir, diferente cultura, educación, edad, memoria, inteligencia, e incluso estado emocional, pueden alterar el resultado final de interpretación. Se trata de una lectura de una interpretación inteligente de señales cuyo código no está en los ojos sino en el cerebro [11-15].

2.2. Espectro visible

La porción de rayos ultravioleta que llega hasta nosotros (de 280 a 400 nm) está en gran parte reducida por el ozono, pero viene en suficiente cantidad para causar reacciones biológicas familiares como un incremento en la producción de melanina y daños genéticos como cáncer de piel. Dado su potencial destructivo, es lógico en términos evolutivos que hayamos alejado nuestra sensibilidad de la luz ultravioleta.

El rango del espectro electromagnético visible para el humano está entre 400 y 750 nanómetros, pero no es casualidad que nuestro sistema visual pueda ver ese rango de frecuencias, sino otra adaptación. El medio en el que nos movemos (el aire) es prácticamente transparente a esta porción del espectro. La radiación visible es capaz de transmitir información desde los objetos distantes hasta nuestros ojos. La Figura 1 muestra el espectro visible al ojo humano, y su relación con el resto de las ondas electromagnéticas conocidas.



Figura 1. Espectro de ondas electromagnéticas.

En resumen, nuestras capacidades visuales son el resultado directo de una larga evolución en un entorno relativamente estable bajo la atmósfera y el Sol.

2.3.El sistema visual humano (SVH)

La información o estímulo viaja hasta nuestro sistema visual en forma de luz a través de un medio físico, comúnmente aire. La luz es transformada (enfocada) por las lentes de cada ojo y proyectada en la retina en la parte trasera de cada ojo. La retina consta de fotorreceptores, denominados conos y bastones, que convierten la radiación electromagnética (luz) en actividad neuronal (química y eléctrica). Los nervios ópticos conectados a los ojos transmiten esta información hasta la corteza visual del cerebro, en ambos hemisferios.

El campo de visión (FOV) está limitado obviamente por el movimiento que puedan realizar nuestros ojos (con la cabeza fija), por nuestra nariz y las cejas. Además, existe un gran solapamiento entre las imágenes que genera cada ojo, que el cerebro debe luego ordenar y juntar para generar una imagen coherente.

La estructura del ojo es manejada por 3 pares de músculos, y su equilibrio posicionamiento resultan en un constante e imperceptible movimiento, que hace que la información que recibe la retina cambie, aunque sea muy poco, cada 100 ms. Existen muchas imperfecciones en el sistema visual humano, como las aberraciones cromáticas, distorsiones y hasta fluidos en el ojo, que hacen que la luz no sea igualmente procesada, sin embargo el funcionamiento de la visión humana parece indicar que existen mecanismos temporales en el cerebro que permiten construir una imagen apropiada incluso ante la presencia de estos problemas.

La retina tiene aproximadamente 125 millones de fotorreceptores, que convierten la energía electromagnética en energía eléctrica y química. Los fotorreceptores se clasifican en conos y bastones, estos últimos son muy sensibles a la luz y a los cambios de luminosidad, pero se saturan fácilmente. Este tipo de células son más efectivas en el proceso de visión en condiciones de baja iluminación, llamada visión escotópica. Existen 120 millones de bastones, estos son más sensibles a longitudes de onda cercanas a los 500 nm y son hasta 10 veces más sensibles que los conos. Los conos, aproximadamente 6 millones, son mucho más efectivos en condiciones de alta iluminación (durante el día), en la visión fotópica, y son sensibles al color. Los conos se dividen en tres tipos, cada uno de ellos más sensible a diferentes longitudes de onda, desde las cortas (420 nm), las medias (530 nm) hasta las largas (560 nm) [16].

Los conos y bastones no están uniformemente distribuidos en la retina. Los conos están ubicados en una zona en el centro de la mácula lútea, llamada fóvea, de aproximadamente 1,5 mm cuadrados. Los fotorreceptores de la fóvea son los responsables de la agudeza visual. Cuando se enfoca un objeto, se centra la proyección de la luz que emite este justo en el centro de la fóvea, la

foveola, de aproximadamente 0,33 mm de diámetro. En la fóvea, existe una convergencia 1:1 con las células ganglionares (del nervio óptico), por ello es la zona donde se obtiene mayor resolución o nitidez visual. El ser humano enfoca cambiando la forma de la lente (la potencia focal en humanos es de 15 dioptrías).

- **Proceso temporal**. En cada fotorreceptor ocurre un proceso químico que dura algunos milisegundos. Todos los fotones que recibe cada fotorreceptor durante este proceso contribuyen al resultado, dando lugar a una respuesta promedio en el tiempo que dura el proceso. El efecto de ese promedio se denomina suavizado temporal o persistencia retiniana, y se puede comprobar fácilmente al observar luces que parpadean a gran velocidad. Cuando la luz parpadea lentamente, se percibe cada parpadeo individual correctamente, pero al aumentar la velocidad, y superar la tasa denominada CFF (*critical flicker fusion*) el *flash* se fusiona en una imagen continua. Esta CFF está alrededor de los 60 Hz. Bajo condiciones de iluminación bajas, el CFF disminuye levemente [17].
- Proceso espacial. Desde los más de 120 millones de fotorreceptores los estímulos deben transmitirse por aproximadamente un millón de células ganglionares. Estas últimas reciben los impulsos de múltiples fotorreceptores. La relación entre los fotorreceptores y estas células es muy diferente entre los conos y los bastones. Los conos, en la fóvea, tienen una relación de casi 1:1, requiriendo estímulos con suficiente energía para poder transmitir los impulsos, mientras que los bastones, en su mayor parte en la periferia de la fóvea, tienen una relación mucho mayor y muchos se conectan a la misma terminación de una célula ganglionar. La relación de muchos bastones a una sola célula ganglionar genera un efecto de amplificación del estímulo de la luz, por eso es que en la noche sólo los bastones participan en la visión, captando la poca luz disponible.

Los ojos son el instrumento central en el sistema de visión humana, el cual reacciona a la luz y así obtiene la percepción de las formas de los objetos, los colores, la profundidad y el movimiento, etc. La estructura de un ojo es similar a la de una cámara [18], en el sentido de que un ojo también tiene una lente (lente cristalina) y una placa sensorial (retina). Las unidades sensoriales (células receptoras) de la retina son mayores en el centro y menores en los bordes, creando así una visión central (también llamada fóvea) y periférica. La visión centralizada es responsable de la alta capacidad de agudeza visual, esencial para capturar los principales detalles visuales. El área periférica más grande está caracterizada por sensores de luz altamente sensibles (células de varilla) y por lo tanto es adecuada para detectar el movimiento, y se vuelve más efectiva en la oscuridad. A la vez que ofrece abundante información del mundo físico, nuestros ojos desnudos tienen límites físicos inherentes en la resolución, tanto espacial como temporal. En concreto,

nuestra percepción visual posee tres características ópticas, las cuales son respuesta centralizada, la fusión de parpadeo paso bajo y el efecto de matriz fantasma, que se relacionan con la resolución espacial, temporal y la sensibilidad al movimiento, respectivamente.

- **Respuesta centralizada**. Los ojos humanos no pueden discriminar los detalles demasiado finos, debido al límite físico de la mínima diferencia entre las células sensoriales adyacentes en la retina central. Típicamente, la excelente resolución espacial de los ojos humanos es de 0,07°, lo que corresponde de 1,2 mm a 1 metro de distancia de observación [19]. La resolución espacial también se ve afectada por el contraste espacial. El campo receptivo de los ojos humanos tiene diferentes respuestas en el centro y en los alrededores. Los campos receptivos pequeños (o grandes) son estimulados por altas (o bajas) frecuencias espaciales. Típicamente, nuestra percepción espacial sigue una característica de paso de banda (cercana a la de paso bajo): la mayor frecuencia es de unos 2-4 ciclos por ángulo visual (equivalente a un ciclo de contraste dentro de 5mm a 1m de distancia de observación). En resumen, los ojos humanos no pueden distinguir cosas demasiado pequeñas, y los detalles más finos se fusionan y aparecen como una especie de promedio. Un ejemplo es el control de la vista: cuando los detalles finos exceden los del ojo, uno sólo verá la forma aproximada.
- Fusión de parpadeo paso bajo. Las fluctuaciones temporales de la intensidad de la luz no son perceptibles para el ojo humano cuando esta supera una cierta frecuencia, denominada frecuencia de parpadeo crítica (CFF para abreviar en adelante) [20,21]. En cambio, el ojo humano sólo percibe la luminancia media. El comportamiento temporal del sistema de visión humano puede aproximarse como un filtro lineal de paso bajo a una frecuencia superior a la CFF. Un ejemplo común es la observación de la rueda de un coche girando. Cuando gira lo suficientemente rápido, dará como resultado la percepción semitransparente de esta. Cabe destacar que la CFF se ve afectada por muchos factores, incluyendo los contrastes de color, el movimiento, las formas de onda de luminancia, entre otros muchos. En escenarios típicos, se cree que la CFF de los ojos humanos es de unos 40-50Hz según los resultados de una amplia investigación sobre la visión [20, 22, 23]. Por ejemplo, los parpadeos que presenta un monitor de rayos catódicos (CRT para abreviar) de 60Hz no son perceptibles.
- Efecto de matriz fantasma. Describe otra característica espacial del ojo humano: la sensibilidad al movimiento. Si bien los objetos en movimiento rápido se acercan a la vista (ya sea por el movimiento del objeto o por el movimiento del ojo, como los ojos que giran), se puede notar el parpadeo incluso cuando la frecuencia de vis visualización es mucho más alta [24]. Por ejemplo, podemos observar el parpadeo de un LED a una frecuencia muy superior a la típica de un CFF si el LED se mueve en un entorno oscuro, pero sólo vemos el LED

constantemente encendido si está quieto. Cabe destacar que, a diferencia de la fusión del parpadeo, el origen del efecto de la matriz fantasma no ha podido ser determinado aún hoy en día. Estudios recientes descubren que a una menor amplitud de parpadeo se produce un mayor ciclo de trabajo y un mayor tamaño del haz lo hacen menos visible [25, 26].

2.4. Factores que determinan la sensibilidad al contraste

Se pueden encontrar varias definiciones del contraste en la literatura, aquí se mencionan las tres más empleadas:

• Weber. El contraste según Weber, directamente deducido de la ley general de Weber, intenta establecer una relación cuantitativa entre la magnitud de estímulo físico y cómo éste es percibido. La ley general de Weber establece que "el menor cambio discernible en la magnitud de un estímulo es proporcional a la magnitud del estímulo". En el caso del contraste, la función derivada de dicha ley es la siguiente:

$$C = \frac{I - I_b}{I_b} \tag{1}$$

donde $I e I_b$ representan la luminancia de las características y del fondo respectivamente. Este modelo es usado normalmente cuando se encuentran pequeñas características sobre un fondo de luminancia constante.

• Michelson. El contraste de Michelson [27] se define como:

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$
(2)

con L_{max} y L_{min} las luminosidades máximas y mínimas respectivamente. El contraste C está en el intervalo $\{0 \le C \le 1\}$. El contraste de Michelson se utiliza en patrones en los cuales las características oscuras y claras son equivalentes y ocupan proporciones similares en la imagen.

RMS. El contraste RMS no depende de la frecuencia espacial ni de la distribución de contraste de la imagen. Se define como el desvío estándar de las intensidades de los píxeles [28] y viene dado de la forma:

$$C = \sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ij} - \bar{t})^2}$$
(3)

siendo I_{ij} la intensidad de cada píxel de una imagen de *M* por *N*, y \overline{I} como el promedio de intensidades de la imagen.

Durante los años 60, algunos investigadores [29] desarrollaron algunas ideas sobre la habilidad del hombre de percibir los detalles, e intentaron encontrar los límites del sistema de visión humano. El resultado de estos estudios es un modelo muy conocido denominado *Contrast Sensitivity Function (CSF)* presentado por Mannos y Sakrison entre otros.

La función CSF indica los límites de percepción a diferentes frecuencias y contrastes de un estímulo visual. Si la frecuencia del estímulo visual es muy alta (detalles imperceptibles) entonces no se podrá reconocer más el patrón del estímulo, y la percepción será solamente de un color uniforme. Estos límites de resolución espacial están directamente relacionados con el número y el tamaño de nuestros fotorreceptores. La función CSF a la que llegaron Mannos y Sakrison es la siguiente:

$$A(\alpha) = 2,6(0,0192 + 0,144\alpha)e^{(-0,144\alpha)^{1,1}}$$
(4)

y en la Figura 2 se puede observar una representación de tal función. La curva delimita lo que las personas, con visión normal, pueden ver y lo que no pueden ver, usando la visión central.

Hasta hace poco se creía que la CSF de una persona no podía mejorar, y que a medida que el sujeto avanza en su edad su CSF indicaría una pérdida de agudeza visual. Estudios recientes de la Universidad de Rochester [28] mostraron que jugadores de videojuegos tienen mejor percepción de pequeños detalles (altas frecuencias) que personas que no juegan frecuentemente, demostrando que el SVH puede adaptarse y mejorar, en función de los estímulos cotidianos.



Figura 2. Función de sensibilidad al contraste.

La sensibilidad al contraste depende de muchos factores que tienen un origen tanto óptico como neural. Entre los factores ópticos más relevantes, debido a que determinan la función de transferencia de la modulación óptica, están el diámetro de la pupila [30, 31], la excentricidad [32, 33] y, naturalmente, el estado refractivo del sujeto [30, 34].

En definitiva, entre los factores neurológicos, que son los que dan más información sobre la naturaleza de los mecanismos fisiológicos que intervienen en la detección de los patrones acromáticos, consideraremos particularmente tres: la luminancia media, el tamaño del estímulo y la excentricidad. La luminancia media determina el estado de adaptación del sistema visual, el tamaño determina el número de ciclos que comprende el estímulo y la excentricidad determina las características de las matrices de sensores que actúan en cascada para realizar una determinada tarea visual.

2.5.Sensibilidad temporal

El SVH es menos capaz de distinguir detalles cuando el objeto que se proyecta directamente en la fóvea está en movimiento. El resultado de esto es la sensación de emborronado a medida que el objeto pasa frente a el punto de enfoque o cuando el sujeto mueve la cabeza para enfocar otro punto.

Murphy [35] propuso que la incapacidad del ojo de seguir objetos en movimiento de manera precisa es la causa del efecto de emborronado que se percibe, sin embargo, estudios más recientes

hechos por Tyler [36] y Nakayama [37] sugieren que los fotorreceptores limitan la sensibilidad temporal a los detalles, concluyendo que el proceso de detectar movimiento en sí mismo implica la integración de energía, en el tiempo durante el movimiento del objeto, y es justamente éste proceso de integración temporal y espacial el que puede destruir la información visual de determinadas características de alta frecuencia del objeto.

Daly [38] extendió el modelo de la CSF para agregar la variable velocidad. Esta versión se puede aplicar bien a imágenes estáticas y enfocadas con la visión central (fóvea), es decir, cual es la percepción del estímulo cuando éste está en movimiento y el sujeto está siguiéndolo con su vista.



Figura 3. (a) CSFs para velocidades de 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64 y 128 grados (derecha a izquierda). (b) La frecuencia espacial detectable más alta posible en función de la velocidad.

En la Figura 3 se pueden ver dos gráficas representadas por Reddy [39] que muestran como la CSF se desplaza en el eje de la frecuencia espacial al aumentar la velocidad del estímulo. La mayoría de los autores coinciden que al incluir en el modelo el movimiento del estímulo, el resultado es que la CSF sufre desplazamientos y escalados.

3.1. Introducción

Las cámaras y pantallas son omnipresentes en la vida de la mayoría de los seres humanos y su número continúa creciendo de manera exponencial con el paso del tiempo. Además de las cámaras de vigilancia instaladas en calles y negocios, la mayoría de las personas posee y lleva consigo distintos dispositivos que integran y combinan ambas tecnologías. Un ejemplo de estas son los ordenadores (tanto de escritorio o portátiles), teléfonos inteligentes, tabletas, monitores, sistemas de juego, televisores, etc. Con el rápido avance de la tecnología en los últimos años, las pantallas y cámaras han ido mejorando rápidamente. La mayoría de las pantallas LCD, especialmente las que tienen capacidad 3D, soportan una frecuencia de refresco de 120 Hz o superior. Por ejemplo, el monitor del *Macbook Air* de 13 pulgadas, que se utiliza en este trabajo, soporta 60 fotogramas por segundo (FPS). Otros monitores como el LG 47LH55 incluso soportan físicamente 240 FPS. Además, con la aparición de las pantallas basadas en diodos emisores de luz orgánicos (OLED), se soporta un tiempo de respuesta mucho más bajo (menos de 0,01 ms) en comparación con el LCD, lo que indica su gran capacidad para soportar una tasa de refresco bastante alta.

Las cámaras también están evolucionando a un ritmo considerable. Las cámaras de alta resolución para fines concretos pueden producir imágenes de un gigapíxel y alcanzar una velocidad de cuadro de captura de vídeo de más de 2.000 FPS. Incluso las cámaras de los teléfonos inteligentes convencionales pueden soportar imágenes de alta resolución a una velocidad de captura rápida, que supera lo que la retina está buscando. Por ejemplo, el Samsung Galaxy S5 admite una resolución de 16 megapíxeles y una tasa de captura de 120 FPS, mientras que el iPhone 6 admite una resolución de 8 megapíxeles y una tasa de captura de hasta 240 FPS. En el caso de este proyecto, se usará como dispositivo móvil el Huawei Mate 10, el cuál soporta una grabación de 30 FPS y 60 FPS en el modo dual.

3.2.Pantallas

Desde los receptores mecánicos hasta los televisores de pantalla plana ha habido un mundo de diferentes tecnologías que se han ido incorporando en el mundo de las pantallas. La historia de la televisión está plagada de circuitos analógicos que han ido convirtiendo en digitales. Un ejemplo de ello es el tubo de rayos catódicos que fue el primero que proporcionó el gran paso del desarrollo de la televisión y de las pantallas. Posteriormente aparecieron las pantallas planas de cristal

líquido o de plasma que han logrado sustituir dicha tecnología. Actualmente se puede encontrar una gran variedad de técnicas [40] que se presentan a continuación:

- Pantallas de cristal líquido (LCD). Se trata de un sistema eléctrico de presentación de datos formado por dos capas de cristales polarizados y entre ellas un material especial (cristal líquido) que tiene la capacidad de compartir propiedades líquidas y sólidas para orientar la luz a su paso y así formar la imagen. Las pantallas LCD se basan en cuatros principios básicos:
 - La luz se puede polarizar.
 - o Los cristales líquidos pueden transmitir y cambiar la luz polarizada.
 - La estructura de un cristal líquido puede ser cambiada por corriente eléctrica.
 - Hay sustancias transparentes que pueden conducir electricidad.

Por otro lado, para entender aún más esta tecnología hay que entender primero como funciona el componente principal de ella, el cristal líquido. Es una sustancia casi transparente que combina la propiedad de los líquidos y de los sólidos a su vez. Por un principio de física, un líquido no puede tener forma cristalina ya que entonces no adoptaría la forma del recipiente que lo contiene. En realidad, la forma cristalina adoptada debería llamarse semicristalina porque las moléculas ocupan un lugar fijo, pero con una amplia tolerancia sobre todo en el sentido de la orientación de su eje mayor.

Las moléculas del cristal líquido son de tipo gigante con forma de pequeñas varillas o cigarrillos, y normalmente (sin campo eléctrico aplicado) están alineadas todas paralelas unas con respecto a otras.

La luz que pasa a través del cristal líquido sigue la alineación de sus moléculas (sólidos), es decir, puede cambiar la polaridad de la luz. Al aplicarles electricidad (un campo eléctrico), cambian de alineación sus moléculas (líquidos) como bien se comentó con anterioridad. Realmente, cuando se aplica un campo eléctrico la sustancia se parece más a un cristal. En cambio, cuando no se aplica campo es una sustancia orgánica que tiene propiedades de un líquido (por ejemplo, la fluidez).



Figura 4. Estructura interna de una pantalla de cristal líquido.

La luz polarizada vibra en una sola dirección para cada momento, pero la dirección de vibración cambia con el tiempo. La dirección de vibración es única y constante con el tiempo. El filtro polarizador es como una rejilla que permite únicamente el paso de la luz que oscila en el mismo plano de la rejilla, la cual es la luz polarizada.

El procedimiento de esta técnica es el siguiente:

- 1. A medida que se enciende la pantalla, la luz de fondo del LED emite la luz blanca.
- 2. La luz entra en la placa de guía de luz (LGP), se refleja internamente y se distribuye uniformemente sobre la superficie superior del panel. La hoja difusora dispersa aún más la luz, por lo que no se observan puntos calientes fuera de la LGP.
- 3. El *Dual Brightness Enhancement Film* (DBEF) es una lámina que se encarga de reciclar la luz dispersada y la hoja de prisma garantiza que la luz esté enfocada y dirigida hacia el espectador.
- 4. El polarizador interior permite que la luz de la longitud de onda vertical pase mientras bloquea otras orientaciones.
- 5. Posteriormente, la luz polarizada verticalmente pasa a través de la capa de cristal líquido.

- 6. Luego, los cristales líquidos se manipulan aplicando el voltaje adecuado a través de la pantalla de cristal líquido (TFT) y el electrodo común. Los cristales líquidos pueden bloquear la luz blanca a un grado variable. El filtro en frente de cada subpíxel solo permite el paso de luz a través de un rango de longitudes de ondas apropiadas para su color.
- La luz pasa a través de los cristales líquidos y los filtros de color para producir los colores primarios.
- 8. La luz polarizada es filtrada por el polarizados superior. Solo se transmite la luz polarizada horizontalmente.



Figura 5. Formación de la imagen en una pantalla LCD.

Una pantalla LCD está formada por dos filtros polarizantes como se ha podido observar en la Figura 4. Estos filtros están formados con filas de cristales líquidos alineados perpendicularmente entre sí, de modo que al aplicar o dejar de aplicar una corriente eléctrica a los filtros, se consigue que la luz pase o no pase a través de ellos, de manera que el segundo filtro pueda bloquear o no el paso de luz que haya atravesado el primero. Dentro de los tipos de LCD podemos encontrar:

LCD de matriz pasiva o Dual Scan Twisted Nematic (DSTN). Por una malla de conductores a ambos lados del LCD pasa una corriente eléctrica que destuerce el cristal y deja pasar la luz. Nos referimos a matriz como la capa situada por debajo de los conductores, utilizada para activar los elementos de la pantalla. La matriz pasiva suele

estar construida por una base de tiras conductoras dispuestas de borde a borde de la pantalla, como se aprecia en la Figura 6.



Figura 6. Disposición de las tiras conductoras en una matriz pasiva.

Puesto que las tiras son relativamente largas, el tiempo empleado para activar cada elemento es mayor que en los modelos de matrices activas, ampliamente utilizados. Esto significa que se tarda más en refrescar la pantalla. En otras palabras, este tiempo incrementa con el aumento de la pantalla.

LCD de matriz activa o Thin Film Transistor (TFT). Cada uno de los píxeles tiene transistores y capacitores por cada punto que lo controlan, y llegan a generar 256 grados de brillo por píxel. Con ello se mejora el direccionamiento de cada píxel por separado. Este tipo de matriz se basa en una rejilla de transistores independientes que descansan en una capa situada por debajo de los elementos de la pantalla. El ángulo de visión es más ancho y la posición de los transistores obstruyen menos la iluminación que las tiras conductoras mencionadas con anterioridad.



Figura 7. Disposición de las tiras conductoras en una matriz activa.

Plasma. Esta tecnología en lugar de emitir electrones a través del tubo de rayos catódicos como su antecesor, dispone de una estructura diminuta de celdas colocadas entre dos láminas. En el interior de cada celda se encuentra una mezcla de gases (habitualmente Xenón, Neón y Argón), además de un pequeño condensador y un par de electrodos. Bajo esta disposición,

cuando se establece entre los electrodos una gran diferencia de potencial eléctrico se produce la ionización de los gases contenidos en la celda, generándose así un estado conocido como plasma (situación análoga a encender un tubo fluorescente). La mezcla de gases emite una luz ultravioleta, en lugar de luz visible, que excita la capa de fósforo de cada celda produciendo la iluminación de la pantalla. La estructura de una pantalla de plasma es la siguiente:



Figura 8. Diagrama de una pantalla de plasma.

Por otra parte, el funcionamiento de esta tecnología queda recogido en la Figura 11:



Figura 9. Funcionamiento de una pantalla de plasma.

• **Pantallas LED.** Las pantallas LED son una variante de las pantallas LCD (también llamadas TFT-LCD), que se diferencia de las LCD convencionales en que éstas utilizan retroalimentación mediante CCFL (*Cold Cathode Fluorescent Lamp*, lámpara fluorescente de cátodo frío) mientras que las pantallas LED emplean retroalimentación mediante diodos LED (*Ligth-Emitting Diode*, diodo emisor de luz).


Figura 10. Diagrama de bloques de una pantalla LED.

En esta tecnología tenemos una fuente de luz LED azul como principal origen de iluminación. Esta luz se filtra por la capa de fósforo que tiene el LED para conseguir una luz parcialmente blanca que funciona como retroiluminación en los paneles LCD.

Cada píxel individual se compone de tres subpíxeles RGB a modo de filtros que sólo dejan pasar el componente rojo, verde o azul de la luz blanca del sistema de retroiluminación. Para controlar dichas cantidades se usa un cristal LCD, con una transparencia variable para dejar pasar más o menos luz a través de los subpíxeles RGB. Si se deja pasar toda la luz por los tres subpíxeles RGB. Si se deja pasar toda la luz por los tres subpíxeles tendremos un punto blanco y si se bloquean el rojo y el verde tendremos un punto azul, y así sucesivamente.

Uno de los puntos débiles de esta forma de generar los colores en pantalla está precisamente en la luz blanca empleada para la retroiluminación, ya que se encuentra muy lejos de ser pura. Este suceso hace que los colores primarios tampoco lo sean, y por ende tampoco sea posible obtener colores precisos a partir de la combinación de diferentes proporciones RGB.

Dentro de las pantallas LED, existen distintos tipos [40] en función de la tecnología de retroiluminación LED empleada:

Edge LED: Los diodos LED van colocados alrededor de la pantalla, en los bordes del televisor, utilizándose un panel especial para distribuir la luz de forma uniforme por toda la superficie de la pantalla. Esta disposición lateral de los diodos hace que la pantalla sea muy delgada, pero no permite iluminar de forma precisa toda la pantalla, lo que afecta negativamente al contraste.

Full LED (también llamado Direct LED): Los diodos LED están distribuidos por todo el panel, no solo en los bordes de la pantalla, lo que permite un mejor contraste, brillo y calidad de imagen, aunque con un mayor consumo de energía y grosor que las pantallas explicadas con anterioridad.



Figura 11. Esquemas de Edge y Full LED.

OLED. Una pantalla OLED es un dispositivo monolítico, transistorizado que típicamente consiste en una serie de películas delgadas orgánicas intercaladas entre dos películas delgadas, los electrodos conductivos. Los materiales orgánicos y la estructura que los forman determinan los rasgos característicos del dispositivo: larga duración, colores nítidos y bajo consumo. Básicamente esta tecnología se basa en una matriz de diodos LED orgánicos. Estos diodos se caracterizan por su bajo consumo y su elevado rendimiento. Cada punto de la pantalla es un elemento orgánico. Con un pequeño impulso eléctrico generan la luz. La disposición es la de dos cristales y entre medias un ánodo, un cátodo y un líquido orgánico por el que pasan electrones dando dicho impulso.

La base de funcionamiento de la tecnología OLED queda ilustrado en la Figura 12:



Figura 12. Base de funcionamiento de la tecnología OLED.

MicroLED. El MicroLED es una tecnología que fusiona tanto lo mejor de LED como de OLED. Esta técnica consiste en LEDs autoemisores de luz a escala micrométrica (µm), es decir, son LED físicamente mucho más pequeños que los de un televisor OLED y que además producen luz por sí mismos. En esta tecnología no se incorporan componentes orgánicos, por lo que desaparece la degradación y las quemaduras de los paneles. Además, MicroLED ofrece la reducción del consumo energético y una calidad cromática muy alta, así como un nivel alto de brillo. Por contraparte, esta tecnología emplea un material inorgánico llamado nitruro de galio (GaN). Este cambio reduce la necesidad de una capa de polarización y encapsulación, haciendo que los paneles sean más delgados. Como resultado, los componentes de este sistema son pequeños, de ahí su nombre, ya que miden menos de 100 µm.

Esta tecnología está construida a partir de diodos emisores de luz. Esto significa que emplean la tecnología de autoemisión, por lo que cada subpíxel RGB produce su propia luz. Por lo tanto, este tipo de pantallas presentan relaciones de contraste muy altas y negros profundos.

• Quantum-dot (QLED). La tecnología QLED [41-43] nace de la evolución de las pantallas LCD LED a través de los *Quantum Dots* (QD, puntos cuánticos), para mejorar el rendimiento en áreas claves de la calidad de la imagen. Estos puntos, son partículas de

dimensiones muy pequeñas donde sus propiedades electrónicas son diferentes de otras mayores. Son tan pequeñas, de hecho, que a efectos prácticos se podrían considerar bidimensionales y poseen lo que los físicos denominan como espectro discreto de energía cuantizada. Dicho de una forma más sencilla, estas motas a escala nanoscópica tienen la particularidad de que emiten su propia luz de color cuando son iluminadas por una fuente externa como una lámpara LED. En función de los materiales y el tamaño, un QD brillará con uno u otro color. De esta manera, con el uso de esta tecnología el tamaño de la partícula dictaminará la longitud de onda de la luz que emite, lo que permite obtener hasta un billón de colores distintos para formar imágenes extremadamente nítidas.

Para entender que aportan los QD, es preciso recordar cómo funciona un panel LCD-LED tradicional que fue descrito en los puntos anteriores. Los QD entran en juego en la premisa de la pureza de la luz blanca. Esta tecnología soluciona el problema de la pureza del color manteniendo la retroiluminación LED, y los cristales LCD, pero discerniendo en lo demás.

En la retroiluminación por led azul ahora se elimina la capa de fósforo para conseguir una mayor intensidad y pureza, pero es ahora donde entra en juego otra capa que traduzca dicha intensidad y pureza lumínica en colores preciosos. La idea es que a cada subpíxel le llegue la luz del color que corresponda. Para conseguir los componentes R y G se usan puntos cuánticos con un tamaño de 6 y 3 nm. Posteriormente, al incidir la luz azul sobre estas partículas, brillarán con un color puro sin tonalidades parásitas. De este modo, una pantalla QLED está compuesta de tres colores de gran pureza: el azul de la retroiluminación, y el rojo y verde de los puntos cuánticos. Al final, lo que se consigue es un sistema capaz de emitir luz con longitudes de ondas perfectamente controladas para que se correspondan a colores puros. De esta forma, se combina lo mejor de la tecnología LCD-LED con las ventajas derivadas de contar con los puntos cuánticos.

Finalmente, a modo de resumen se muestra un esquema de las tecnologías comentadas a lo largo de este capítulo en la Figura 13 junto a una comparativa entre cada una de ellas, haciendo hincapié en sus fuerzas y debilidades.

En relación a las dos tecnologías de LCD se observa que la LCD convencional suele encontrarse con inconvenientes a la hora de representar los colores en los extremos de la pantalla, mientras que los TFT presentan colores y brillos mucho más uniformes, mejorando así el contraste. Por otra parte, en las pantallas LCD es común observar la estela de un objeto cuando se mueve rápidamente, mientras que en pantallas TFT, el tercer electrodo usado en esta tecnología permite un tiempo de respuesta más precioso cuando se tiene que representar movimientos veloces, es

decir, las imágenes poseen mayor nitidez. Del mismo modo en las pantallas LCD estándar el ángulo de visión varía entre 130°-140°, mientras que su análoga amplía dicho ángulo de visión hasta 160°.



Figura 13. Esquema de las principales pantallas en el mercado.

A continuación, se muestra una tabla donde se reflejan las ventajas e inconvenientes de las pantallas LCD:

Ventajas	Inconvenientes
El píxel es más pequeño que en los televisores	Las pantallas LCD pueden sufrir un tiempo de
de plasma. De este modo, se pueden encontrar	respuesta bajo (refresco), que tiene como
pantallas pequeñas con altas resoluciones.	consecuencia un efecto estela en aquellas
	imágenes con movimientos muy rápidos.
Las pantallas LCD no sufren parpadeos,	Los monitores LCD tienen un ángulo de
garantizando una visión óptima, la cual, no	visión estrecho, por efecto de la polarización
produce cansancio.	del cristal líquido.
Este tipo de pantallas ofrece una buena	El color negro siempre tiene una luminosidad
luminosidad, que asegura una mayor nitidez	residual de fondo.
de las imágenes.	
Tienen una buena eficiencia energética.	El contraste de los televisores LCD es inferior,
	tanto al obtenido en las pantallas de rayos
	catódicos como en las de plasma.

Tabla 1.	Ventajas e	inconvenientes	de las pant	tallas LCD.
----------	------------	----------------	-------------	-------------

En contraparte, podemos realizar una comparativa similar entre las pantallas LCD y las plasma. Como se comentó anteriormente, las pantallas LCD sólo podían alcanzar un ángulo de visión de 150°, mientras que las pantallas de plasma son capaces de obtener un ángulo de 170°. Por otra parte, aunque consigan unos contrastes más notables que las pantallas LCD, sufren el problema de quemado. Al permanecer mucho tiempo una imagen estática en la pantalla, esta quedará marcada incluso después de ser cambiada o al apagar el dispositivo. Su principal ventaja es la calidad de sus colores, siendo muy superior a la mayoría de las pantallas LCD. Del mismo modo que para las pantallas LED, se establecen una serie de ventajas y desventajas en el uso de dicha tecnología:

Ventajas	Inconvenientes		
	Los colores oscuros, especialmente los		
	matices negros, no están optimizados. Así		
	que, en una televisión plasma, cada elemento		
	que la constituye (píxel) posee dos estados:		
Muestran colores muy naturales, brillantes y	encendido y apagado. De este modo los tonos		
luminosos y una óptima escala de grises.	intermedios son generados a partir del tiempo		
	de encendido de cada uno de estos elementos.		
	Si el proceso no se realiza de forma idónea,		
	aparecerá un parpadeo en las áreas oscuras de		
	la pantalla.		
	El tamaño de cada píxel es mayor que en las		
Proporcionan un ángulo de visón máximo y	LCD, de esta manera, sólo se pueden		
no produce polarización.	encontrar televisores de plasma con		
	dimensiones superiores a 32 pulgadas.		
	Los paneles de plasma sufren el fenómeno		
	<i>"burn-in</i> ", un tipo de quemado de los píxeles.		
Ofrecen una luminosidad uniforme por toda la	Si un objeto permanece en la pantalla de		
pantalla.	forma estática mucho tiempo, se crea un tipo		
	de impresión fantasma que queda impresa		
	sobre el televisor.		

Tabla 2. Ventajas y desventajas de los televisores de plasma.

En resumen, a pesar de que esta tecnología ha mejorado bastante con el paso del tiempo, el impacto ambiental y su corta vida útil hacen que no sean un modelo muy eficiente de pantalla.

En el caso de la tecnología LED presenta grandes ventajas, ya que tiene muy poco coste energético y un tiempo de vida útil bastante largo en relación a sus antecesoras. Poseen una mejor definición y llegan a niveles de brillo inalcanzables para las tecnologías anteriores, además de mejorar el ángulo de visión. Su principal inconveniente es la temperatura, ya que esta tecnología es muy sensible a dichos cambios y haría que la intensidad de la luz de la pantalla pueda variar.

Finalmente, cuando se trata de niveles de negro, la tecnología OLED reina como el campeón indiscutible. Lógicamente, si OLED enciende o apaga los compuestos orgánicos en función de la necesidad, el nivel de negro será mejor dado que simplemente apagando dichos compuestos, se obtendrá un negro total, debido a la falta de luminosidad. Aún así, las pantallas QLED mejoran el rendimiento del nivel de negro de la pantalla LED, pero dependen de las luces de fondo que brillan detrás del panel LCD. Además, las pantallas QLED sufren un efecto llamado "purga de luz", donde la luz de fondo se filtra en lo que se supone que es una sección negra de la pantalla. Por otra parte, la tecnología MicroLED, al tener la misma filosofía de funcionamiento en cuanto a iluminación que la OLED, es decir, se tratan de autoiluminación controlada, puede obtener negros más puros que las pantallas QLED, simplemente apagando los elementos que generan luz.

En cuanto a brillo, la tecnología MicroLED puede tener mayores valores de luminosidad frente a la tecnología QLED y OLED. Específicamente, las pantallas QLED llegan a tener 2.000 nits (o lo que es lo mismo 2.000 candelas por metro cuadrado), y la MicroLED la supera notablemente llegando a valores de brillo de 100.000 nits.

Cuando se habla de espacio de color, la tecnología QLED es claramente la ganadora. Con la adición de la capa de puntos cuánticos se busca alcanzar una mayor gama de colores. Al mejorar la pureza de la luz de fondo, han permitido a los televisores QLED aumentar la precisión del color, el brillo y el volumen del color frente a la tecnología OLED y MicroLED.

El tiempo de respuesta se define como el tiempo que tarda cada diodo de la pantalla individualmente en cambiar de encendido a apagado. Con un tiempo de respuesta más rápido, se produce menos movimientos. OLED, con sus diodos más pequeños que funcionan como píxeles individuales supera a la tecnología QLED. En contraste, los diodos de QLED no solo son más lentos, sino que se sientan detrás del panel LCD e iluminan grupos de píxeles no individuales. Esto provoca un cambio global más lento entre los estados "encendido" y "apagado". En cuanto al tiempo de respuesta en la tecnología MicroLED supera incluso a la OLED.

Con las pantallas QLED y MicroLED, el mejor ángulo de visión está en el centro, ya que la calidad de la imagen disminuye tanto en color como en contraste cuanto más se mueve uno de

lado a lado. En contra, las pantallas OLED se pueden ver sin degradación de la luminancia en ángulos de visión drásticos: hasta 84 grados.

Por otra parte, en la actualidad, existen dos técnicas [44, 45] diferentes disponibles para interpretar el vídeo: barrido entrelazado y barrido progresivo. La elección de una de estas dos técnicas dependerá de la aplicación y el objetivo del sistema de vídeo y, en particular, de si será necesario captar objetos en movimientos y permitir la visualización al detalle de una imagen en movimiento. En primer lugar, comenzaremos hablando del barrido entrelazado. Las imágenes que se basan en este tipo de barrido utilizan técnicas desarrolladas para las pantallas de monitores de TV con tubo de rayos catódicos (ya es desuso), que constaban de 576 líneas visibles horizontalmente situadas a lo ancho de una pantalla de TV estándar. El entrelazado las divide en líneas pares e impares y, a continuación, las actualiza a 30 imágenes por segundo. El pequeño retraso entre las actualizaciones de una línea par e impar crea una distorsión. Esto ocurre porque sólo la mitad de las líneas sigue la imagen en movimiento, mientras que la otra mitad espera a ser actualizada.



Figura 14. Formación de una imagen mediante técnicas de barrido entrelazado.

Los efectos del entrelazado se pueden compensar ligeramente utilizando el desentrelazado. El desentrelazado es el proceso de convertir el vídeo entrelazado en una forma no entrelazada, eliminando parte de la distorsión del vídeo para lograr una mejor visualización. A este proceso se le conoce también como "duplicado de líneas". Esta característica elimina los problemas de distorsión de movimiento provocados por la señal de vídeo analógica de la cámara analógica. Este tipo de barrido ha sido de gran utilidad durante muchos años en el mundo de la cámara analógica, la televisión y el vídeo VHS, y aún lo sigue siendo para determinadas aplicaciones. Sin embargo, con el cambio de la pantalla de cristal líquido (LCD), los monitores que se basan en transistores de película delgada (TFT) y las cámaras digitales, se ha creado un método alternativo de aportar imagen a la pantalla, conocido como barrido progresivo. Este barrido, a diferencia del entrelazado, escanea la imagen entera línea a línea. En otras palabras, las imágenes captadas no se dividen en campos separados como ocurre en el barrido entrelazado, sino que cada fotograma está formado por una imagen única línea a línea. Por otro lado, las pantallas actuales sólo aceptan este tipo de barridos.



Figura 15. Funcionamiento del barrido progresivo y del entrelazado.

En una escena fija y con un monitor de calidad no notaremos diferencia entre una imagen entrelazada y una progresiva. Las diferencias aparecen cuando las imágenes muestran mucho movimiento o texturas móviles (por ejemplo, el humo, el agua, la afluencia de tráfico, ...). En estas situaciones la imagen entrelazada muestra una pérdida de definición, sobre todo en los bordes de cada objeto de la escena. La misma escena reproducida en progresivo se muestra tal y como ha sido grabada, sin pérdida de calidad alguna. De este modo se puede comprobar cómo van a llegar a nosotros dichas imágenes en función del formato. Estos formatos se denominan atendiendo a su resolución en vertical y añadiendo una "p" minúscula en el caso de ser imágenes progresivas o una "i" en el caso de ser entrelazadas (interlaced). Así tendremos resoluciones del tipo: 720i y 720p como resoluciones más bajas y 1080i o 1080p como resoluciones superiores.

En definitiva, el tipo de barrido entrelazado tiene como mayor ventaja que utiliza un menor ancho de banda porque se envía la mitad de la imagen en cada cuadro. Pese al reducido uso de ancho de banda, la velocidad de escritura es suficiente para que en la recepción se obtenga la representación de las imágenes sin que suceda parpadeo o se note distorsión en la imagen. A pesar de ello, los avances en codificación de vídeo han permitido tener calidades de imagen muy altas con anchos de banda muy reducidos. Un ejemplo de ello son los estándares de codificación de H.264, H265 y H.266 [46, 47]. El motivo por el que se utiliza el H.265 para los videos de ultra-alta definición es que permite reducir considerablemente la tasa de transmisión mientras que iguala en cuestiones de calidad los resultados de su predecesor H.264. Sin embargo, ambos algoritmos de compresión utilizan estrategias diferentes para comprimir el video, cuya principal diferencia radica en la manera en la que se realizan la compensación de movimiento entre imágenes consecutivas. Por otra parte, la nueva versión del códec H.266 [48] (también conocido como *Vesatile Video Coding* o VVC), ha llegado con la idea de aportar más facilidades a la hora de consumir vídeo en streaming. Las principales diferencias con sus antecesores son la reducción del tamaño de los

archivos a la mitad (comparado con en códec H.265), soporte para altas resoluciones (HD, 4K y 8K) y más versatilidad (resolución de forma adaptativa).

Por otro lado, el barrido progresivo es el que mayormente se utiliza en las pantallas de los televisores modernos, los monitores de ordenadores de escritorio o portátiles, las consolas de juego, los servicios de transmisión de vídeos por internet, etc. El motivo de su uso es que, al mostrar la imagen sin variaciones en todo momento, se obtiene un mayor detalle, así como un mejor contraste si se compara con el escaneo entrelazado. También se mejora la fluidez con la que se visualiza el contenido, se elimina el posible parpadeo y soporta métodos de compresión más eficientes gracias a los diferentes formatos en que se puede almacenar el contenido en alta definición.



Figura 16. Diferencia entre el uso de barrido progresivo y entrelazado.

En definitiva, en el caso de las imágenes entrelazadas, la velocidad de refresco permite engañar al ojo humano y hacerle creer que se están reproduciendo las imágenes completas, aunque sólo se reproduzcan la mitad de las líneas que la componen en cada fotograma. Por el contrario, en una reproducción progresiva si se proyectan simultáneamente todas las líneas de la escena y con los fotogramas dibujados al completo.

3.3.Cámaras

La función principal de la cámara es el de transformar la luz emitida por una escena en una tensión eléctrica, la señal de vídeo. Esto se logra descomponiendo dicha luz, una vez focalizada por el objetivo, en tres espectros correspondientes a las longitudes de onda de los rojos, verdes y azules, respectivamente, a través de un dispositivo óptico, el separador cromático. Estas tres luces se traducen a tensiones eléctricas, una para cada sensor de imagen, que, tras ser procesadas electrónicamente, son transmitidas hasta el receptor. En este, por medio de otro transductor, el convencional tubo de rayos catódicos o modernamente las pantallas de plasma o de cristal líquido, se reproducen tres luces: roja, verde y azul respectivamente, que han de causar una sensación visual lo más parecida posible a la de la visión de la escena directamente.

El sistema óptico [49] de la cámara lo constituye el objetivo, filtros neutros, los filtros de color y el separador cromático. Un objetivo con zoom no solo enfoca una escena formando su imagen en los sensores de la imagen, sino que también, sin perder el enfoque, es capaz de variar su ángulo de campo pasando de teleobjetivo a gran angular. Este control se hace de forma manual o automática, zoom in o zoom out. Existe otra posibilidad que es el de las tomas macros, que aportan al sistema un control total sobre la luz que entra en el objetivo por medio del diafragma o iris.



Figura 17. Diagrama de bloques del sistema óptico.

Existen dos tecnologías principales de captura de imagen, los *Charged Coupled Device* (CCD) y los sensores basados en la tecnología *Complementary Metal-Oxide Semiconductor* (CMOS). Tradicionalmente, los sensores CCD se han usado en aplicaciones donde se requieren los datos de imagen de alta calidad. Sin embargo, el desarrollo de la tecnología CMOS ha permitido que una gran cantidad de dispositivos de electrónica dispongan de una cámara de un tamaño reducido.



Figura 18. Diagrama de un sensor de imagen CCD.

Al contrario que en los sensores de tecnología CCD, las celdas de la matriz CMOS [50-52] son totalmente independiente de sus vecinas. La principal diferencia radica en que estos sensores realizan la digitalización píxel a píxel dentro del mismo sensor, por lo que la circuitería adicional al sensor es mucho más sencilla. Estos sensores sustituyen la película fotográfica por un circuito integrado con una matriz de elementos fotosensibles basados en dispositivos de estado sólido, que guardan electrónicamente la imagen.

Por tanto, los sensores fotográficos poseen un conjunto de receptores que, como su propio nombre indica, reciben los valores lumínicos, de forma que cada fotorreceptor recoge un nivel distinto de intensidad lumínica. La cuantía de este nivel depende de diferentes factores, como son las características de reflexión de luz de toma fotográfica, apertura del diafragma y estructura de los elementos intermedios u objetivos, con su número y calidad de lentes. Cuando se han recogido todos los datos, estos serán enviados en formato digital al procesador a través del circuito electrónico que recorre todo el chip. La conversión de formato analógico a digital se realiza siguiendo los siguientes pasos:

- Cada elemento de la imagen emite una intensidad lumínica analógica determinada, en función de las características de la imagen de estudio.
- II. Cada celda es sensible a la cantidad de fotones recibidos, registrando un valor específico por cada rango de intensidad luminosa recibida.

- III. El valor recibido por cada celda del sensor se envía al procesador, así como la información de identificación de la propia celda, es decir, la cuantificación de luminosidad exterior.
- IV. El procesador ordena todos los datos recibidos de cada una de las celdas estudiadas y configura la imagen digital.

La información de color la añade el procesador, y se obtiene a través de diversos procesos de interpolación.



Figura 19. Funcionamiento de filtro Bayer en procesado de imágenes.

Es importante comprender que los fotodiodos por si mismos son incapaces de interpretar colores en una imagen. El color como tal no existe físicamente, sino que es una percepción visual generada por el, a partir de las diferentes longitudes de onda de la luz. Lo anterior presupone un problema para la fotografía digital a color: al no existir el color, lo único que captan los fotodiodos son los niveles de luminosidad de la escena. Es decir, solo "ven" más o luz o menos luz, lo que en la práctica se traduce en una imagen en blanco y negro.

La solución a este problema es el mosaico de Bayer. La inmensa mayoría de las cámaras del mercado incorporan uno de estos filtros en el sensor de imagen. La función del filtro de Bayer es hacer que a cada fotodiodo que compone el sensor solo le llegue la longitud de onda correspondiente a uno de los colores primarios. Cada celda fotorreceptora del sensor fotográfico es recubierta de una malla de filtros con los tres colores básicos del espectro, es decir, rojo, azul y verde, poseyendo la parte de filtro correspondiente al color verde el doble de superficie que las dos anteriores. El filtro de Bayer está conformado por un 50% de filtros verdes, un 25% de rojos y un 25% de azules. Interpolando la información de cuatro fotodiodos vecinos (dos verdes, uno rojo y uno azul), según la cantidad de luz que se haya recibido de cada uno de estos colores, se obtiene un píxel de color "real".



Figura 20. Filtro de Bayer.

El fotorreceptor recibirá diferentes intensidades de color rojo, azul y verde en función de la onda luminosa que reciba, enviándose luego esta información, junto con la intensidad lumínica registrada y los datos de celda, al procesador de imagen, dando lugar a un píxel con una intensidad luminosa y con un color derivado determinado. La unión de cada uno de estos píxeles junto con su información de intensidad luminosa y de color conformará la imagen final o fotografía. Este proceso es observable en la imagen anterior adjunta (Figura 19). Una de las principales limitaciones de la matriz de Bayer es que requiere de la información de 4 fotodiodos para generar un píxel de color. Por ello, existen alternativas a esta metodología: sensor *Foveon* y *Multi-Shot*.

Desde el punto de vista de las comunicaciones ópticas [53], el receptor CMOS es un elemento clave en este ámbito, ya que es capaz de realizar la conversión óptica a eléctrica y de ofrecer multiplexación espacial inherente a un coste muy bajo. El receptor consiste en un fotodetector seguido de un preamplificador y, en ciertos casos concretos, uno o más post-amplificadores.

Por tanto, uno de los principales componentes del receptor óptico es el fotodetector, que puede estar compuesto por varios tipos de dispositivos como fotodiodos PIN, fototransistores y fotoconductores. El fotodiodo PIN, es quizás el más usual, ampliamente utilizado debido a su bajo coste y alta versatilidad. Se modela como una fuente de corriente en paralelo con un condensador, siendo corriente de salida del dispositivo proporcional a la potencia óptica entrante.

Cabe destacar que el rendimiento de un receptor óptico está determinado principalmente por la combinación preamplificador - fotodiodo.

Por otro lado, el amplificador de transimpedancia es el preamplificador más utilizado para los receptores ópticos de alta velocidad, ya que es capaz de combinar una ganancia de transimpedancia relativamente alta, necesaria para una alta sensibilidad, con una alta velocidad. Esta tecnología se basa en un amplificador de voltaje inverso con ganancia de lazo abierto. Como regla general, esto reduce tanto las impedancias de entrada como de salida del amplificador por la ganancia de bucle del amplificador. El amplificador de transimpedancia resultante es, por lo tanto, apto para aceptar una corriente a su entrada y proporcionar un voltaje en su salida.

Finalmente, en cuanto a los post-amplificadores, es importante destacar su gran relevancia. En la mayoría de las comunicaciones ópticas, se requiere un gran rango dinámico ya que la potencia óptica disponible depende fuertemente de la configuración de la fibra óptica, especialmente del número de conectores. Además, en las aplicaciones con multiplexación por división de tiempo, el presupuesto de potencia óptica puede variar de una franja de tiempo a otra. Para hacer frente a las diferentes amplitudes de señal, los post-amplificadores de los receptores ópticos son necesarios, estando basados en varias topologías diferentes según el caso concreto de estudio. Algunas incluyen el control automático de ganancia para ampliar el rango dinámico y el control automático de las fluctuaciones de los niveles de la señal de entrada, mientras que otros se basan en la limitación o recorte. Los post-amplificadores en los receptores presentados se basan en cadenas de inversores sesgadas. Éstas ofrecen la ventaja de un gran rango dinámico y una oscilación de salida completa de carril a carril que puede aplicarse directamente a una puerta digital.

Asimismo, cabe destacar que algunas de las características más relevantes de la tecnología CMOS digital estándar no modificada son el mínimo número de componentes externos necesario y las características del sesgo, integrado y automatizado. Además, todos los receptores tienen una oscilación de salida completa de carril a carril y son capaces de conducir un inversor mínimo.

Por otro lado, existen diferentes tipos de ruido en los circuitos integrados semiconductores [54], siendo los más relevantes:

- Ruido térmico: se debe a las variaciones producidas por la oscilación térmica de los portadores de electricidad en la conductividad de los materiales semiconductores o conductores. El valor cuadrático medio de este tipo de ruido se expresa en función de K, la constante de Boltzman, T, temperatura, B, el ancho de banda y R, la parte real del elemento

semiconductor considerado. Este tipo de ruido es también conocido como "ruido blanco", ya que posee un espectro plano.

- Ruido impulsivo (shot noise): se debe a la generación aleatoria de portadores que son capaces de atravesar una barrera de potencial, siendo un tipo de ruido típico de uniones semiconductoras. Al tratarse de procesos aleatorios, es capaz de provocar variaciones en el valor medio de la corriente que atraviesa la unión.
- Ruido 1/f o *flicker noise*: Actualmente, existe todavía cierta controversia sobre su origen exacto. Su existencia está estrechamente relacionada con la estructura de bandas de los semiconductores y los procesos de recombinación de portadores. Posee una dependencia inversamente proporcional a la frecuencia, haciéndolo por tanto especialmente relevante cuando esta posee valores muy bajos.

La conmutación de tensiones y corrientes es claramente la mayor fuente de ruido en los circuitos integrados actuales. La solución más simple para minimizar el ruido es reducir tanto los niveles como la frecuencia de las conmutaciones. Para ello, una de las soluciones más populares es la utilización de los circuitos lógicos CMOS, que, sin embargo, maximizan las excursiones (swings) de los niveles de corriente y tensión, y además concentran estas conmutaciones en intervalos muy pequeños de tiempo. Este tipo de ruido característicos de los circuitos CMOS puede clasificarse en función del origen de la perturbación, siendo el más relevante el ruido dV/dt, debido a conmutaciones en las tensiones de los nodos y cuyos fenómenos más relevantes son el crosstalk y el ruido dl/dt.

Los principales contribuyentes de ruido en los CMOS son el fotodiodo y el preamplificador, explicados anteriormente a lo largo de este documento. En un buen diseño, la contribución al ruido del preamplificador es mínima, ya que prácticamente no se añade ruido si no se utilizan componentes activos en el mismo. Este es el caso del preamplificador óptico más simple posible: una única resistencia RL que realiza tanto la conversión de una corriente a un voltaje como la preamplificación. El mayor inconveniente de esta tecnología es el limitado ancho de banda máximo alcanzable cuando el bajo ruido es importante. Por lo tanto, este receptor no es adecuado para aplicaciones de comunicación de alta velocidad. A continuación, se muestran algunas figuras [55] que establecen a modo de resumen las diferencias de los dos sistemas más relevantes en el campo de los sensores de imagen:



Figura 21. Estructura de los sensores CMOS y CCD.

Hay una diferencia fundamental entre la estructura de los sensores CMOS y CCD. Lo más importante es que el proceso de fabricación de CMOS es algo similar al proceso de fabricación de los chips de memoria. Esto significa que los sensores CMOS pueden ser producidos en muchas plantas de todo el mundo y de una manera muy económica en comparación con el proceso específico para la producción de los CCD. Con un rendimiento mejorado, los sensores CMOS están empezando a ser reconocidos por la industria y han conseguido imponerse ante los sensores CCD. En el caso de sus aplicaciones, el sensor CMOS está disponible para diversas aplicaciones, mientras que los sensores CCD siguen siendo el estándar común para cámaras broadcast, aunque se espera un aumento del uso de la tecnología CMOS en este tipo de cámaras.

Temas	CMOS	TI CCD
La tecnología más reciente	er	No
	51	No hay novedades
Sonsor All-digital	e:	No
Sensor All-digital	31	Tecnología analógica y antigua
	Sí	No
Altamente confiable	Baia velocidad de reloi, baio consumo, baia	Alta velocidad de reloj: alto consumo
	temperatura (altamente confiable)	Alta temperatura (menos confiable)
	Sí	No
Bajo costo de mantenimiento	Alta estabilidad de píxeles (mínimos defectos de píxeles y bajos costos de mantenimiento)	Altos defectos de píxeles (altos costos de mantenimiento)
Valor comercial alto	Sí	No
	Valor comercial alto y baja depreciación	Valor comercial bajo > 2011 y alta depreciación

Figura 22. Consideraciones de los sensores CMOS Y CCD.

Aplicaciones	CMOS	TI CCD
Teléfonos móviles/Cámaras Fotográficas	Todas estas cámaras se basan en diseños de chip único completamente integrado, y ahora se ofrecen con 4-8 megapíxeles por lo menos: • Diseño de chip único • Bajo consumo de energía • Control de sensibilidad electrónica flexible • Amplio rango dinámico	No hay productos
DSLR/Cinematografía Digital	Todas las cámaras DSLR y cámaras de 35 mm de alta gama en el mercado actual utilizan un diseño de sensor único y un separador de colores integrado	No hay productos
Broadcast	Xensium de 2/3 pulgadas con diseño de beam splitter óptico de Grass Valley Cámaras selectas de broadcast de Ikegami La próxima generación de cámaras de broadcast de 2/3 pulgadas	La mayoría de las cámaras de broadcast de gama alta de 2/3 pulgadas de fabricantes japoneses. No hay nuevos desarrollos previstos

Figura 23. Aplicaciones de los sensores CMOS y CCD.

El *global shutter* de los CCD expone cada *frame* durante 1/50 segundos y muestra algunos efectos borrosos cuando existen objetos en movimiento en la escena. La reducción de los tiempos de exposición para cada *frame* se utiliza (por ejemplo, para deportes) para obtener imágenes más nítidas, sin embargo, esto crea algunos efectos de obturación.



Figura 24. Luminosidad en los sensores CMOS y CCD.

La diferencia en el diseño de sensores muestra una notable diferencia en el rendimiento bajo condiciones de luz extremas. El CCD, debido a su diseño con columnas de transporte en la parte baja de la imagen, muestra los efectos de desbordamiento con luces altas, que son visibles como rayas verticales blancas o incluso de color, en la parte superior e inferior del toque de la luz. Un nivel de dichas características (tipo Smear) es de -135 dB. Por otro lado, los sensores CMOS, debido a su estructura, no muestran ningún efecto de Smear por luces de alta intensidad, ni rayas.



Figura 25. Intensidad media (Strobe flash) para sensores CMOS y CCD.

El *Rolling shutter* del sensor CMOS y el *global shutter* del CCD muestran un comportamiento diferente a los destellos de luz cortos. En un CCD, un *Strobe flash* muestra un efecto en el frame durante el cual ocurrió, y puede ser visto como un *frame* con un brillo aumentado. Por otro lado, en un sensor CMOS, el *Strobe flash* muestra un efecto en el frame en el que ocurrió, y en un frame siguiente. Con CMOS, el primer *frame* tiene el brillo aumentado desde el punto en que el flash se produjo, hasta la parte inferior de la imagen. En el segundo *frame*, existe un aumento de brillo desde la parte superior de la imagen hasta el punto en que el flash se produjo en el frame previo. Esto puede verse fácilmente en el modo de imagen fija cuando uno se mueve de una imagen a otra. Sin embargo, en una secuencia de vídeo funcionando a 50 o 60 *frames* por segundo, este efecto es casi invisible, percibiéndose como un aumento de brillo en un *frame* completo (similar al CCD).

Hay un efecto más problemático con el sensor CMOS si se utilizan tiempos de exposición cortos para la lectura de imágenes. En este caso, el brillo de sólo una parte de la imagen se incrementará, y esto puede ser más perturbador que un cuadro completo de brillo diferente. Dado que la reducción del tiempo de exposición se utiliza para el control de la sensibilidad en las cámaras de teléfonos móviles y algunas otras aplicaciones de cámaras de gama baja, este efecto puede ser visto regularmente en esas aplicaciones. Sin embargo, la reducción del tiempo de exposición se utiliza raramente en aplicaciones con cámaras de sistema. Por lo tanto, el comportamiento de sensores CMOS con tiempos de exposición cortos no es relevante.

Formatos HD	CMOS	TI CCD
	Diseñado con 1920x1080 píxeles activos en sensor de formato progresivo	Diseñado con 1920x1080 píxeles activos en sensor de formato progresivo
1080i50/60	El sensor puede cambiar a lectura entrelazada	El sensor puede cambiar a lectura entrelazada
720p50/60	Escalado para horizontal y vertical	Escalado para horizontal y vertical
1080p24/25/30/50/60	Disponible para productos de cinemato- grafía digital	Lectura directa de píxeles
Alta velocidad > 50/60 fps	Frame rates extremamente altas con veloci- dades de 10X a más posibles	Velocidad máxima de 3X disponible

Figura 26. Cambios de formatos para sensores CMOS Y CCD.

Los sensores de CCD y CMOS son básicamente, sensores de formato progresivo con el estándar de 1920x1080 píxeles, con un total de 2,2 millones de píxeles efectivos. Ambos tipos de cámaras pueden cambiar entre lectura progresiva y entrelazada. Por lo tanto, no se necesita conversión externa para obtener un formato entrelazado. Otros formatos HD, como 720p, se pueden derivar a través de la conversión de formato externa. Sólo con sensores CMOS es posible la opción de cambio de formato nativo con un modo de lectura de "región de interés". Los framerates más bajas y más altas que el estándar de TV son posibles tanto con sensores CCD, como CMOS. La gran ventaja de CMOS es la baja velocidad del reloj interno para broadcast normal, lo cual lo hace ideal para diseños de cámara ultra lenta.



Figura 27. Defectos de los píxeles en sensores CMOS y CCD.

La baja velocidad del reloj interno de CMOS mantiene el sensor a una temperatura relativamente baja, y como resultado se mostrarán menos defectos de píxeles visibles con el tiempo. Por otro lado, los CCD son bien conocidos por defectos de pérdida de píxeles, que podrían aumentar los costos de mantenimiento con el tiempo. A pesar de que los sensores CCD han sido el estándar por excelencia para la captación de señales, han llegado al final de su ciclo de vida sin nuevos avances tecnológicos previstos. Las características de rendimiento de los CMOS en comparación con los CCD son excelentes, ofreciendo muchos más beneficios para el usuario. Pero el beneficio real es la imagen. Los sensores CMOS producen una calidad de imagen igual o mejor que los CCD de hoy.

4.1.Introducción

Hasta el momento se ha hablado de la esteganografía como el conjunto de métodos para ocultar información. En realidad, esta definición no es del todo correcta, ya que a este conjunto de métodos se le conoce precisamente como métodos de ocultación de la información (del inglés, *information hiding*), siendo la esteganografía un subconjunto de los mismos. La ocultación de la información es la ciencia que engloba cualquier método que permita esconder cualquier tipo de información, sea cual sea su naturaleza, sus medios, o sus fines. En definitiva, la esteganografía es la rama de la criptología que trata sobre la forma de escribir información secreta de tal modo que sólo quien la envía y quien la recibe sepan de su existencia.

En el marco de las telecomunicaciones, el enmascaramiento de la información adquiere gran utilidad e importancia para el anonimato y la seguridad de transmisiones secretas. Gracias a la digitalización, los avances tecnológicos del software, el hardware computacional y la expansión de las redes de telecomunicaciones, se genera y transmite a diario un gran torrente de información, el cual hace que sea posible el ocultamiento de metadatos, así como el transporte desapercibido de información confidencial. La esteganografía se desarrolla con diferentes técnicas y variantes, desde tiempos remotos, con el objetivo de proveer privacidad en las comunicaciones a distancia; así pues, un sistema esteganográfico será aquel que enmascara el contenido en un medio para evitar que el mensaje sea percibido, de modo tal que ninguna parte fuera del emisor y el receptor conozcan la existencia de este.

Actualmente existe un creciente número de aplicaciones que utilizan esta tecnología, no sólo para llevar a cabo comunicaciones privadas a través de Internet, sino también en la protección de la propiedad intelectual (copyright), en votaciones digitales, en transacciones electrónicas y para el espionaje en entornos militares o económico-industriales. Diferentes esquemas esteganográficos compiten entre sí para lograr mayor robustez al ruido, alta capacidad de almacenamiento de información, seguridad ante la detección no autorizada e imperceptibilidad de la imagen esteganografiada.

El principal objetivo de los sistemas esteganográficos es garantizar la total imperceptibilidad del mensaje oculto; con este propósito se han desarrollado diversas técnicas con alto grado de complejidad y éxito, pero ninguna es totalmente invulnerable, debido a la continua evolución de las técnicas de esteganoanálisis, orientadas a la adquisición de la información.

Debido a la importancia que ha adquirido esta ciencia en los últimos tiempos, este trabajo surge con motivo de documentar los algoritmos y procesos que llevan a cabo la esteganografía en imágenes y realizar una descripción del diseño e implementación de un método para el cifrado de la información de manera imperceptible, segura y robusta; para ello se analizan y comparan las principales técnicas aplicadas en los sistemas esteganográficos, comenzando por las más simples, desarrolladas en el dominio espacial, y finalizando con el estudio de los complejos algoritmos que requieren específicas transformaciones de la señal. De esta forma, se da margen a la aparición de nuevos y más avanzados métodos que en las situaciones requeridas y gracias a la existente tecnología logran su cometido.

En todo sistema esteganográfico se pueden identificar los siguientes elementos:

- *La carga útil, mensaje o marca*: son los datos o mensajes que se desean transportar de forma oculta. Es el fichero u objeto que se pretende embeber. Este fichero puede ser texto, imagen, audio o vídeo, bien sea cifrado o sin cifrar.
- La portadora o fichero que cubre a la carga útil (estegomedio): es el fichero que oculta la carga útil. Es el objeto visible en el que el mensaje embebido se oculta. Análogamente, puede ser de texto, imagen, audio o vídeo.
- *El paquete u objeto esteganográfico, señal esteganografiada o fichero con la marca digital (esteganograma)*: es el resultado de incrustar la carga útil en la portadora. Es la portadora con la carga útil, mensaje o marca de agua oculta embebida.
- *Clave secreta (llave) o par de claves pública-privada esteganográficas*: es la clave o claves esteganográficas utilizadas para controlar el proceso de ocultar la información y extraerla intacta en el receptor. Aportan seguridad al sistema.

4.2. Esteganografía en la era digital

El proceso de ocultamiento de la información en un sistema esteganográfico comienza por identificar el medio de recubrimiento, el estegomedio, que no es más que le medio en donde el mensaje será ocultado. Los diferentes tipos de estegomedios son [56]: archivos de imagen, sonido, texto o vídeo, archivos ejecutables, páginas web, campos no usados de paquetes de redes (TCP/IP), espacios no utilizados de discos y particiones escondidas.

Cualquier tipo de fichero es apropiado para ocultar datos en su interior, pero los más utilizados han sido las imágenes y ficheros de sonido. Las imágenes digitales son adecuadas para esconder datos por muchas razones; están siempre presentes en Internet, ilustrando un sitio Web, o de adjunto en un correo electrónico; lo que no llama la atención. Aún con las técnicas modernas de compresión, las imágenes son todavía relativamente grandes y pueden variarse de forma imperceptible. Existen varios métodos desarrollados para la esteganografía con imagen, más que para cualquier otro estegomedio [57].

Cada tipo de formato de fichero gráfico se puede utilizar para desarrollar métodos esteganográficos concretos [58]; pero todos ellos tienen características comunes que permiten su uso activo en la esteganografía:

- Contienen una cabecera que identifica el tipo de fichero del que se trata, que posee información para interpretarlo como puede ser el tamaño de la imagen o el número de colores.
- Una vez descomprimidos los datos de un fichero mediante el algoritmo concreto de cada formato, los datos del fichero indican el color específico de cada píxel de la imagen.

En función de los colores de la imagen se utilizarán más o menos bits que indicarán las cantidades de rojo, verde y azul que se usará para representar cada color en cada píxel. Debido a que la paleta de colores puede ser modificada según la imagen, es necesario almacenarla en un fichero. De los 24 bits de cada píxel, se utilizan 8 bits por color; en el caso de imágenes a escala de grises, cada píxel utiliza sólo 8 bits para establecer un valor entre 0 y 255 que representa el nivel o luminancia de este.

4.3. Estegosistema: descripción y características

Como se mencionó con anterioridad, la esteganografía consiste en ocultar un mensaje secreto en un medio de cobertura para que sólo el destinatario previsto sea consciente de la existencia del mensaje oculto. Para mayor seguridad, el mensaje suele ser codificado antes de ser incrustado en el medio de cobertura. El resultado de este procedimiento de codificación es una secuencia pseudo-aleatoria de 0 y 1, que luego se incrusta en la cubierta. Al incrustar los bits del mensaje en una imagen, normalmente se introducen algunos artefactos visuales y se cambian las estadísticas de la imagen. Las estadísticas de la imagen son analizadas por métodos de estegoanálisis, mientras que los artefactos visuales pueden ser detectados por el ojo humano. La Figura 28 representa los pasos de inserción y extracción en un sistema esteganográfico general, en el cual el transmisor (también llamado codificador) aplica primeramente alguna transformación a la imagen portadora, en la que el mensaje **m** será enmascarado, obteniendo el vector **X** que representa a los valores ya transformados de la imagen. Esta transformación (DFT - Transformada discreta de Fourier, DCT- Transformada discreta del coseno, DWT-Transformada discreta de Wavelet) es opcional en dependencia del método que se aplique. Seguidamente, se lleva a cabo el algoritmo esteganográfico del esquema en cuestión que puede ser tres tipos esencialmente: los que modifican el bit menos significativo, los de espectro extendido y los que aplican modulación del índice de cuantificación; al cual se le introduce la llave de codificación **k**, útil también en el proceso de descifrado, y es obtenida como salida la señal esteganografiada **S**, que con una transformación inversa resulta ser la imagen con el mensaje o marca incluida.

De esta manera, el esteganograma viaja a través de un determinado canal, el cual introduce ruido o es objeto de ataques intencionados (para obtener la información escondida o simplemente destruirla) o ataques no intencionados (manipulaciones al fichero como es la compresión o filtrado), hacia el receptor. Una vez realizada o no, la misma transformación que en el codificador, es obtenido el esteganograma con ruido **S'**, que junto a la original llave **k** forma las entradas del esquema de encubrimiento inverso para de esta forma extraer el mensaje. El estegosistema anterior puede contener técnicas criptográficas para mayor seguridad, o códigos de corrección de errores agregando redundancia al mensaje para incrementar la robustez de este.



Figura 28. Pasos de inserción y extracción del mensaje en un sistema esteganográfico.

La esteganografía adaptativa utiliza las características de la imagen y la información del contenido para mejorar el proceso de incrustación. Por ejemplo, se pueden introducir artefactos menos detectables en la imagen de la portada por el método de incrustación. En algunos casos, estos artefactos se detectan con dificultad por los métodos de esteganalización debido a los cambios triviales introducidos en las estadísticas de las imágenes. Sin embargo, los artefactos pueden ser perceptibles visualmente para el ojo humano. Por lo tanto, la indetectabilidad perceptiva es muy importante en la esteganografía. Esto significa que la aplicación de un método de esteganografía a una imagen no debe dejar ningún fenómeno visualmente perceptible. Para apoyar esta idea y mejorar la robustez de los métodos de marca de agua, se propusieron modelos perceptivos para describir el sistema visual humano.

Las características de un estegosistema, representante de un esquema determinado por un algoritmo específico, dan una medida de la eficiencia de este, por lo que se usan para hacer comparaciones entre los métodos existentes y establecer pautas de superioridad o deficiencias. Estas son:

 Imperceptibilidad: Mide la diferencia que existe entre la imagen original (el estegomedio) y la imagen esteganografiada con el mensaje oculto (esteganograma), la cual no debe ser distinguible por el sistema visual humano (SVH). Se mide en términos de la relación señal a ruido pico, PSNR. Un PSNR mayor que 30 dB, implica que la diferencia entre ambas no es distinguible por el SVH.

$$PSNR = 10 \, log_{10}(\frac{256^2}{MSE}) \tag{5}$$

$$MSE = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} (Io - Iw)^2}{N_1 N_2}$$
(6)

Las siglas MSE representan el error de distorsión cuadrático medio existente debido a la inserción y los vectores **Io** y **Iw** son la imagen original y marcada, respectivamente, y los valores **N1**, **N2** representan el tamaño de las imágenes.

• *Robustez*: Número de modificaciones que el medio puede soportar antes de que se pierda la información oculta (antes de que el ruido del canal o un adversario destruya o modifique la información escondida).

Se mide en BER (razón de error de bit). Un número sustancial de bits erróneos es dado típicamente por un BER de 0.15.

$$BER = \frac{n \text{úmero de bits erróneos}}{n \text{úmero de bits transmitidos}} \tag{7}$$

Muchas técnicas intensifican la robustez de sus mensajes implementando un codificador de control de error cíclico, para la recuperación de estos, y para la detección de errores en la transmisión.

- *Capacidad de inserción*: Cantidad de información que puede ser ocultada en determinado estegomedio. La capacidad se puede medir en bytes de información, respecto al estegomedio, determinándose, de esta forma, el número de bits insertados en un byte o píxel (para el caso de imágenes a color) de la cubierta.
- Seguridad: Mide la dificultad de detección de la información escondida. La seguridad va a aumentar con el nivel de dificultad del proceso de cifrado esteganográfico, generalmente, y se puede incrementar aún más con procesos criptográficos incluidos en cada uno de los algoritmos realizados.

4.4.Introducción a los métodos y técnicas esteganográficas empleadas

Los métodos esteganográficos se clasifican según la forma de inserción en [56]:

- *Métodos de adición*: Se oculta el mensaje secreto en las secciones del medio portador que pueden ser ignoradas por la aplicación que lo procesa.
- *Métodos de sustitución*: Se modifican ciertos datos del medio portador por los datos del mensaje secreto.
- *Métodos de generación*: Se crea el esteganograma a partir de la información secreta, sin contar con un medio portador previamente.

Los esquemas de sustitución son los más aplicados para la esteganografía con imágenes, a causa de la gran variedad de imágenes existentes con diferentes matices y tonalidades pictóricas; y gracias a que contienen una mayor información que el propio mensaje a empotrar, el resultado de los algoritmos aplicados es muy similar a la imagen original de la cubierta utilizada en el proceso de codificación. Debido a la gran diversidad de formas de llevar a cabo este "arte", los métodos se pueden dividir en dos grandes grupos según el dominio de la imagen utilizado [58]:

- *Basados en el domino espacial de la imagen*: abarcan técnicas que aplican la modificación de los bits menos significativos de cada píxel, o modificándolos mediante expresiones aritméticas, con puertas lógicas o mediante la manipulación del ruido.
- Basados en el dominio de la transformación: están relacionados con algoritmos de modificación y transformación de la imagen. Ocultan los mensajes en las áreas más significativas de la cubierta y pueden manipular las propiedades de la cubierta como es la luminosidad. Utilizan técnicas como son la transformada discreta del coseno (DCT), la transformada discreta de Fourier (DFT), o de Wavelet (DWT), entre otras.

Estas transformadas se pueden desarrollar por bloques, o sobre la imagen entera. De la primera forma se divide la imagen en bloques más pequeños (los dos tamaños más populares de estos bloques son 8x8 y 16x16), y la transformada se realiza individualmente en cada bloque. La ventaja de usar el dominio transformado es que es generalmente más fácil balancear la distorsión introducida por el ocultamiento y da mayor robustez contra el ruido o el ataque, que en el dominio espacial.

La esteganografía es un arte complejo y con muchos matices. Sin llegar a la combinación de esteganografía y criptografía, es posible el uso de determinadas técnicas avanzadas que permiten aumentar la eficacia de una información oculta [59], como son:

• Uso de múltiples claves: técnica heredada directamente de la criptografía, pero con distinta forma de aplicación. Consiste en usar distintas codificaciones para cada porción arbitraria del mensaje a ocultar. Así en el dominio espacial, por ejemplo, una frase de cinco palabras puede tener una clave de codificación para cada una de las palabras: en la primera se resta una unidad en los ceros y se suma una unidad en los unos, en la segunda se realiza lo mismo, pero invirtiendo el orden de los bits, en la tercera se realiza una XOR de los bits.

Las claves, también reconocidas como llaves esteganográficas, son muy utilizadas por los esquemas para incrementar seguridad. Si el receptor no tiene exactamente la misma llave usada en el algoritmo de empotrado de la información, entonces no podrá obtener los datos ocultos, aunque conozca el procedimiento inverso correcto. Estas pudieran ser también datos sobre la imagen original, sobre la naturaleza del mensaje oculto (si es una imagen: sus dimensiones, si es un fichero: el tipo o cabecera del mismo), o simplemente algún dato adicional del algoritmo codificador.

- *Esteganografía en capas*: establece una relación lineal entre los elementos ocultos. Por ejemplo: la codificación de la segunda palabra o letra de un mensaje depende de la primera palabra o letra (puede depender del último valor de la cifra, del último valor modificado, de la posición). De esta forma el esquema, que aplica la esteganografía en capas, establece un orden estricto de decodificación que impide obtener completamente el mensaje sin la primera parte, con lo cual únicamente se debe comunicar la pauta a seguir para encadenar los fragmentos en el decodificador.
- Adición de ruido: en un esteganograma todo el fichero es considerado ruido, pero se puede añadir ruido en el proceso de esteganografiado. Además de modificar los bits necesarios para inyectar los datos, se modifican unos cuantos bits aleatorios del mensaje de forma que, aún teniendo el fichero original, un posible atacante deba conocer el sistema de codificación usado.
- *Uso de distintas magnitudes*: aunque lo habitual es variar en 1 bit de determinado byte de la imagen original, nada impide variarlo en más bits. Así, se establecen algoritmos más complejos que incrementan la seguridad del mecanismo de empotrado, como, por ejemplo: en el procesamiento de una frase de cinco palabras, para ocultar la primera de las palabras se suma 1 bit en la codificación de la primera letra, 2 bits en la codificación de la segunda, 3 bits en la tercera, hasta que vuelva a aparecer una modificación de 1 bit, que significará el inicio de otra palabra.

En este caso de trabajo, para la esteganografía en imágenes existen diferentes enfoques, como la esteganografía de conservación de estadísticas, la esteganografía basada en modelos y la incrustación de máscaras a modo de métodos de procesamiento natural. El principal logro en la esteganografía en imágenes es que el mensaje secreto no sea reconocido por terceros que no tengan derechos de acceso. El receptor debe extraer la información oculta mediante el algoritmo que utiliza el emisor.

En la estenografía basada en modelos, se asume un modelo para la imagen de la portada [60]. El procedimiento de incrustación cambia los elementos de la imagen (valores de los píxeles o coeficientes de transformación) para que no violen el modelo de cubierta asumido. Por ejemplo, Sallee [61] propuso una distribución Cauchy generalizada (GCD) para modelar la distribución de cada coeficiente de la transformada del coseno discreto (DCT) y asumió que todas las imágenes obedecen a este modelo. La idea de utilizar un modelo para la estadística de imágenes en su enfoque era una idea novedosa. Pero, desafortunadamente, los métodos de estegoanálisis detectan fácilmente las imágenes estego-generadas por este método [62, 63]. Otra de las técnicas más

empleadas en la esteganografía de imágenes es la LSB (del inglés *Least Significant Bit*), en la que el mensaje secreto es cambiado a binario para cubrir el interior de la imagen.

4.5. Técnicas de espectro ensanchado para la esteganografía con imágenes (SSIS)

SSIS es un esquema ciego donde la imagen original no se necesita para recuperar la información oculta [64]. El destinatario sólo necesita una llave para revelar el mensaje. Este sistema proporciona la habilidad de esconder una cantidad significante de bits de información dentro de las imágenes, evitando la detección por un observador. Este objetivo defiende la maximización de la capacidad y la minimización de la perceptibilidad.

Técnicas de comunicación de espectro ensanchado, codificación para el control de errores y procesamiento de imágenes se combinan para lograr el esquema SSIS. El proceso del sistema de transmisión se encuentra proyectado en la Figura 29. Dentro del sistema, el mensaje, después de la encriptación opcional, es codificado con una baja razón de código de corrección para no incrementar el mensaje debido a redundancias, produciendo el mensaje codificado **m**. El remitente introduce la llave 2 en un generador pseudo-aleatorio de ruido de banda ancha, generando una secuencia extendida **n**. Luego, llegamos al esquema de modulación que se usa para extender el espectro de banda estrecha de **m** en la secuencia **n**, componiendo la señal **s**, la cual es introducida en un entrelazador espacial. Los entrelazadores también pueden usar una llave para dictar el algoritmo de entrelazado. La señal resultante se combina con la imagen de la cubierta para producir la imagen esteganografiada, la cual es apropiadamente cuantificada para conservar el rango dinámico inicial de la imagen, obteniendo así la parte transmisora del estegosistema.



Figura 29. Proceso de codificación SSIS.

En el receptor, el sistema decodificador, mostrado en la Figura 30, usa técnicas de restauración de imagen y desentrelazado para construir una estimación de la señal con la información s^{\wedge} , de la imagen recibida. El destinatario mantiene la misma llave 2 que usó el remitente, regenerando con ella la secuencia extendida **n**. El mensaje codificado se decodifica y se construye una estimación **m**^{\wedge} de este mensaje codificado. La estimación del mensaje se descifra con el decodificador de errores, y entonces desencripta si es necesario, para que los datos ocultos sean revelados.



Figura 30. Proceso de decodificación SSIS.

5.1.Introducción

En los últimos años hemos sido testigos del rápido crecimiento de las pantallas electrónicas desplegadas en el mundo cibernético. En efecto, estamos rodeados de tales dispositivos en varios factores de forma, que van desde la pantalla de un teléfono o de una tableta, el monitor de un ordenador, un televisor, un tablero electrónico de publicidad y pantallas aún más grandes. Estas pantallas se han convertido en una fuente primaria de información desde el punto de vista del usuario. Por ejemplo, la reproducción de vídeo ha contribuido al 79% del tráfico de Internet [65]. En estos escenarios de uso común, la función principal de una pantalla es transmitir en formación información a los ojos humanos. Así, la pantalla establece la comunicación entre ella misma y los ojos. Se establece así la comunicación de pantalla a ojo (*Screen-to-eye*). A menudo es muy deseable transmitir aún más cierta información a un usuario mientras mira la pantalla. Un ejemplo de la vida real es remitir al espectador a una página web de películas para obtener información adicional. Otro caso similar es el de los vendedores que presentan información añadida (por ejemplo, especificaciones de productos y ventas) en los anuncios de televisión. Con la creciente prevalencia de las cámaras, la comunicación visible de pantalla a cámara ha surgido rápidamente como un canal de comunicación conveniente e improvisado

Durante este periodo se ha producido la adopción universal de los enlaces visibles entre pantalla y cámara como un valioso canal lateral para la comunicación entre dispositivos [5, 7, 29, 66]. Así pues, se introducen códigos de barras dinámicos, por ejemplo, códigos de respuesta rápida (QR), para aumentar la capacidad de transporte de datos por el canal visible entre la pantalla y la cámara de vídeo. Sin embargo, debido al espacio limitado de la pantalla, el contenido visual favorable para el ser humano (normalmente en forma de imágenes) y el contenido visual favorable para los dispositivos (por ejemplo, los códigos QR) se ha disputado la asignación del espacio del monitor y han dado lugar a desafíos estéticos. En consecuencia, hemos visto que un código QR únicamente puede tomar una pequeña área de todo el espacio. Esto no sólo limita su capacidad de transporte de información, sino que también induce a esfuerzos adicionales para poder capturar el código (por ejemplo, estar suficientemente cerca del código QR). La experiencia del usuario para la representación de la pantalla también se considera una distracción. Por lo tanto, surge una pregunta clave: ¿Podemos eliminar esta diferencia restaurando la visualización del frame completo para los usuarios, estableciendo simultáneamente la comunicación de datos entre el dispositivo de pantalla y el frame completo? Fundamentalmente, exige un nuevo paradigma de

comunicación en modo dual, que permita la entrega simultánea de contenido de vídeo primario y otra información sin perjudicar la experiencia de visualización del usuario.

En este proyecto, se presenta un sistema que permite la comunicación visible en modo dual, de cuadro completo, mediante la multiplexación del canal visible para transportar datos sobre el contenido de vídeo normal. La Figura 31 ilustra el concepto de multiplexación de cuadro completo, donde la pantalla muestra el contenido de video multiplexado del video original y los cuadros de datos (por ejemplo, códigos QR). El usuario aún puede ver el video como de costumbre, sin notar los marcos de datos incrustados. La cámara captura los datos transportados por los cuadros multiplexados y decodifica la información relevante. La información se codifica en patrones visuales y se muestra en la pantalla. Los dispositivos equipados con cámaras capturan posteriormente la imagen mostrada por el monitor y recuperan la información de los datos.



Figura 31. Concepto de multiplexación de fotograma completo: Screen-to-eye para videos y Screen-tocamera para la comunicación de datos a través del mismo canal visible simultáneamente.

5.2. Antecedentes de las OCC

La creciente necesidad de comunicaciones móviles de datos está viéndose aún más incrementada debido a la omnipresente conectividad del IoT, tal y como se menciona en capítulos anteriores, y el crecimiento de los servicios digitales como los medios sociales y los contenidos de vídeo. Según el último informe de Ericsson, se espera que el tráfico móvil total aumente con una tasa de crecimiento anual del 42% [67]. En 2023, el tráfico de datos móviles globales mensuales superará los 100 exabytes (EB). Para satisfacer esta demanda que aumenta exponencialmente, se debe utilizar ya sea el aumento del ancho de banda o la mejora de la eficiencia espectral. Sin embargo, el aumento de la eficiencia espectral es lento y no puede satisfacer esta demanda insaciable. La explotación del nuevo espectro se convierte en una única solución, ya que se demanda una banda de espectro mucho más amplia de terahercios (THz). En esas circunstancias, las comunicaciones ópticas inalámbricas han suscitado un gran interés de investigación en los últimos años debido a sus muchas propiedades deseables, entre ellas una gran cantidad de espectro disponible (de 350

nm a 1550 nm), una gran eficiencia energética, una normativa independiente y una seguridad de las comunicaciones debidamente controlada [68-70].

De ahí a que las comunicaciones inalámbricas ópticas (del inglés Optical Wireless Communications - OWC) aparecieran como una alternativa y una opción complementaria a las comunicaciones de radiofrecuencia (RF) existentes. Las OWC cubren las bandas de ultravioleta (UV), infrarroja (IR) y visible. Basándose en la utilización del espectro, los sistemas OWC pueden dividirse en cuatro categorías: comunicaciones ópticas en el espacio libre (FSO), comunicaciones de luz visible (VLC), fidelidad a la luz (Li-Fi), y comunicaciones de cámaras ópticas (OCC) [71]. Los sistemas de comunicación FSO consisten en un transmisor de diodo láser (LD) y un fotodiodo (PD). Típicamente se basa en bandas UV o visibles, ofrece una alta tasa de transmisión a larga distancia, y puede ser usado para la reconfiguración de redes de comunicación. Sin embargo, las comunicaciones de los FSO necesitan una alineación estricta, y, en consecuencia, su costo en los equipos es alto. El VLC basado en bandas visibles es un transmisor de LD o diodo emisor de luz (LED) y un receptor de PD basado en tecnología de comunicación de medio alcance. VLC es capaz de ofrecer una alta tasa de datos dentro de un rango de decenas de metros, pero no considera el acceso de múltiples usuarios. Sin embargo, todos los receptores de las tecnologías FSO, VLC y Li-Fi, consisten en PD, que raramente existen en los dispositivos receptores actuales y el costo de comercialización para cambiar dicha estructura de los receptores es alto. Por lo tanto, el desarrollo de un sistema práctico de OWC para aprovechar los beneficios inmediatos utilizando dispositivos comerciales ya existentes sigue siendo una cuestión pendiente [72]. Esto motiva el concepto de OCC. Las diferentes características y limitaciones de OCC y VLC se presentan en la Tabla 3.

OCC es una tecnología de comunicación que utiliza sensores ópticos de imagen como receptores basados en bandas IR o visibles, y, por lo tanto, también se denomina como comunicaciones de sensores de imagen [73, 74]. Comparada con otros tipos de tecnologías OWC, la OCC tiene muchas ventajas. En primer lugar, no necesita ninguna modificación en el receptor, en el que los actuales teléfonos inteligentes, cámaras digitales, cámaras de vehículos traseros y cámaras de vigilancia son todos posibles candidatos a receptores, por lo que OCC puede facilitar un gran número de aplicaciones.

71

Características	OCC	VLC
Rango de transmisión	Cientos de metros (con telescopio)	Pocos metros
Longitud de onda	UV, IR y luz visible	Luz visible
SNR	Alto	Bajo
Receptor	Cámara	Fotodetector
Velocidad de datos	Menor que VLC (en kbps)	11.67 kbps – 96 Mbps
Protocolo	IEEE.802.15.7r1	IEEE.802.15.7
Implementación MIMO	Fácil	Difícil

Tabla 3. Comparación de OCC y VLC.

En segundo lugar, los receptores del OCC se basan en millones de píxeles, y estos proporcionan un número alto de grados de libertad (DoF) para transmitir datos y manejar el acceso de un número masivo de usuarios. Asimismo, en tercer lugar, los sensores de imágenes de las cámaras actuales pueden normalmente tratar con tres colores que permiten la transmisión sobre el dominio de color. Debido a estas ventajas en cuanto a costo, popularidad y capacidad de transporte de información, OCC ha atraído mucho interés en áreas como IoT, la localización en interiores, la captura de movimiento y los sistemas de transporte inteligente (STI). Sin embargo, OCC tiene sus limitaciones, que incluyen, entre otras, una baja tasa de datos, la cual conlleva a su vez una alta tasa de muestreo en el receptor, el *jitter* de la tasa de captura y el bloqueo aleatorio (igual que en VLC). La comunidad de investigadores de OCC ha propuesto diferentes soluciones para superar las limitaciones ya mencionadas. Por lo tanto, el tema central de este apartado es presentar las diversas limitaciones de OCC y sus soluciones.



Figura 32. Categorías de OCC.

OCC puede dividirse en tres categorías, como se ilustra en la Figura 32. La primera categoría es la de un solo LED como transmisor y la cámara como receptor (LED2C), la segunda categoría consiste en las matrices del LED como transmisor y la cámara como receptor (LEDA2C), que también se denomina MIMO visual [75]. La tercera consiste en una pantalla como transmisor y una cámara como receptor (S2C) que es la que utilizaremos en este trabajo [76].



Figura 33. Estructura del receptor de la cámara óptica, procedimiento de comunicación y aplicaciones.

Basándonos en la clasificación anterior, los posibles candidatos para un transmisor en los sistemas OCC son los LEDS, las pantallas de televisión, las pantallas de monitor, semáforos y las luces de los coches (Figura 33). La luz de todas estas diferentes fuentes pasa por el canal inalámbrico y llega a la lente del receptor. Después, la luz pasa a través del obturador que controla la exposición del sensor de la cámara. Existen dos tipos diferentes de obturador: el obturador global y el obturador rodante. El obturador global controla los sensores de la imagen simultáneamente, mientras que el obturador rodante es responsable de los sensores de la imagen para mostrarla fila por fila [77]. Se genera una imagen de la fuente capturada en el plano imagen de la cámara, que puede procesarse adicionalmente utilizando técnicas de procesamiento para extraer los datos requeridos para diversas aplicaciones. En los sensores de imagen, la luz se transforma en una señal electrónica y se cuantifica para obtener una imagen y se comprime a un formato de calidad específico. Mediante el procesamiento de imágenes y los métodos de extracción de datos, la información transportada en el dominio color/espacio/temporal de la luz, puede decodificarse para su localización y navegación. La variación del canal de luz puede utilizarse para el reconocimiento del movimiento. Aunque la velocidad de los datos es muy limitada debido a la reducida velocidad de fotogramas, pueden transmitir alguna información sobre localización a baja velocidad sobre el reconocimiento de la ubicación, y también puede realizar una comunicación segura cerca del campo.

Por otro lado, los enlaces OCC necesitarían de varias etapas de procesamiento de imágenes para asegurar una recepción correcta, como la detección y el seguimiento de la fuente de luz (la región de interés, ROI), la ecualización de la potencia óptica a lo largo de la ROI y el cálculo del umbral para la demodulación.
Un canal basado en cámara es análogo a un canal RF MIMO, donde cada píxel de la cámara actúa como una antena receptora y los elementos emisores de luz (pantallas) como las antenas de transmisión. En este tipo de canales la calidad de la señal en cada elemento de la antena receptora es una función de la pérdida de trayecto en el canal, la interferencia multitrayecto y la interferencia de otras antenas de transmisión, también llamada interferencia cocanal [78]. Un canal de estas características tiene un desvanecimiento insignificante, pero experimenta una pérdida de trayectoria en la energía de luz e interferencia de otros elementos emisores de luz, que se manifiestan como distorsiones visuales en la salida de la cámara, es decir, en la imagen. Estas distorsiones son principalmente un derivado del proceso de formación de imágenes de la cámara y pueden modelarse (de forma determinista) utilizando la teoría clásica de formación de imágenes en cámaras.

La calidad de la señal en el receptor también se ve influenciada por el ruido del canal. El ruido en los sistemas basados en cámaras se manifiesta como una señal eléctrica ficticia en forma de corriente en cada píxel de la cámara. La corriente de ruido se genera debido a los fotones de la iluminación ambiental y de los circuitos de imagen del transmisor y receptor [79]. La corriente de ruido en un píxel suele considerarse independiente de la señal cuando la iluminación ambiental es suficientemente alta en comparación con la señal de transmisión; por ejemplo, en salas de oficinas o al aire libre [80]. A la salida de la cámara, la corriente de ruido en cada píxel es una cantidad cuantificada y se manifiesta como fluctuaciones en la intensidad (valor digital de la salida del sensor) de ese píxel. Por consiguiente, la energía de ruido acumulada en cada píxel puede ser cuantificada usando el valor medio de la varianza en la intensidad del píxel.

5.3. Trabajos previos en sistemas Screen-to-Camera

En un mundo de los dispositivos inteligentes son cada vez más numerosos, equipados con pantallas y cámaras. El hecho de permitir que las pantallas y las cámaras sean capaz de comunicarse ha atraído un interés cada vez mayor. La idea es simple: la información se codifica en un marco visual que se muestra en una pantalla, y cualquier dispositivo equipado con una cámara puede apuntar a la pantalla y obtener inmediatamente la información. Al funcionar en la banda del espectro de la luz visible, la comunicación entre la pantalla y la cámara está libre de interferencias electromagnéticas, lo que ofrece una prometedora alternativa de comunicación fuera de la banda para la adquisición de información de corto alcance. Asimismo, las comunicación entre pantalla y cámara también aprovecha el canal de luz visible. Naturalmente, las técnicas de VLC pueden ser utilizadas también en este campo en desarrollo. La mayoría de

los trabajos de comunicación Screen-to-camera basados en VLC hacen hincapié en la capacidad e invisibilidad. VRCodes [81] y HiLight [82] son los primeros trabajos de este campo, que utilizan el cambio cromático y de luminancia entre frames, respectivamente, para modular los datos incrustados. Por otro lado, TextureCode [83] optimiza la selección del área de incrustación para reducir eficazmente la obstrucción. Uber-in-Ligth [84] aprovecha el parpadeo (*flicker*) cromático para modular los datos y propone un método mejorado para la demodulación basado en MUSIC para mejorar la precisión de la transmisión. Las comunicaciones Screen-to-camera basadas en tecnología VLC requieren un esquema de extracción de datos diferente (extracción basada en vídeo) a los códigos de barras (extracción basada en imágenes). Del mismo modo, recientes investigaciones han dado lugar a innovadores diseños de códigos de barras que aumentan la velocidad de los datos [85,86] o mejoran la fiabilidad de la transmisión [87-90]. Estos esfuerzos son muy interesantes; sin embargo, comúnmente requieren la visualización de imágenes codificadas visibles, que interfieren con el contenido de la pantalla y crean una experiencia de visualización desagradable.

Los sistemas mencionados han tenido como objetivo el de conseguir una comunicación discreta entre la pantalla y la cámara, que permita a la pantalla cumplir simultáneamente un doble papel: el de mostrar el contenido y la comunicación. En última instancia, se imagina un sistema de comunicación entre la pantalla y la cámara que transmita y reciba datos en tiempo real, al tiempo que garantiza que la comunicación se produzca de forma discreta, independientemente del contenido que se muestre en la pantalla, ya sea una imagen, una película, un videoclip, una página web, una interfaz de un juego o cualquier otra aplicación. A medida que el usuario interactúa con la pantalla y cambia el contenido, la comunicación se mantiene. Por lo tanto, la comunicación se realiza totalmente como una funcionalidad adicional para la pantalla, sin poner ninguna limitación a la funcionalidad original de esta (mostrar el contenido). Como se mencionó anteriormente, recientes investigaciones han realizados valiosos progresos en el diseño de comunicaciones discretas entre pantalla y cámara [91-94]. Sin embargo, sigue existiendo una brecha fundamental en la capacidad de realización, principalmente porque todos los diseños existentes requieren modificaciones directas de los valores de color de los píxeles (RGB). Esta metodología no puede permitir una comunicación discreta en tiempo real sobre un contenido arbitrario en la pantalla que pueda generarse sobre la marcha con las propias interacciones del usuario. La razón es doble. En primer lugar, la modificación de los valores RGB de los píxeles en el tiempo tiene que apoyarse en la Unidad de Procesamiento Gráfico (GPU) aprovechando directamente las bibliotecas de esta. Sin embargo, el sistema operativo no suele permitir que una aplicación de terceros prevea o modifique el contenido de la pantalla de otras aplicaciones o interfaces del sistema (por ejemplo, la pantalla de inicio de un teléfono inteligente o una tablet). Por lo tanto, para lograr la comunicación en tiempo real, los diseños existentes se limitan al contenido de la pantalla dentro de una sola aplicación o archivos independientes (por ejemplo, imagen o vídeo). En segundo lugar, si bien el procesador principal (CPU) puede preconfigurar y modificar los valores RGB de los píxeles, la modificación de estos valores en la CPU conlleva un retraso considerable de cientos de milisegundos, lo que hace que estos diseños sean incapaces de soportar la comunicación en tiempo real sobre contenidos dinámicos como el vídeo. De esta premisa surgen los sistemas mencionados con anterioridad.

5.4. Idea básica del modelo de estudio

Para superar la barrera mencionada en el apartado anterior, se propone un nuevo paradigma de diseño para la comunicación discreta entre pantalla y cámara. El sistema que se pretende tratar se basa en la brecha de capacidad entre el ojo humano y los dispositivos electrónicos. Se sabe que el sistema de visión humana tiene un límite físico, mientras que las pantallas y cámaras modernas han excedido nuestros ojos en términos de resolución temporal, es decir, velocidad. Específicamente, el sistema de visión humana tiene una resolución temporal de hasta 40-50Hz más allá de la cual no podemos capturar objetos que se mueven más rápido. En contraste, las pantallas modernas, especialmente aquellas con capacidad 3D, admiten 120FPS o una frecuencia de actualización más alta. Las cámaras de los teléfonos inteligentes (smartphone) pueden tomar imágenes de alta resolución a altas velocidades de cuadro. Por ejemplo, el *Samsung Galaxy* S5 ofrece una resolución de 16 megapíxeles y una tasa de captura de 240FPS. Además, la capacidad del dispositivo avanza a un ritmo mucho más rápido que el sistema de visión humana.

Existen dos desafíos cuando se diseña la comunicación visible en modo dual de fotograma completo usando este sistema. Primero, el requisito de no afectar el canal primario de *Screen-to-eye* que implica una restricción rígida para la comunicación en modo dual. Además, el sistema de visión humana es sensible a los parpadeos. Por lo tanto, necesitamos encontrar un esquema de multiplexación de cuadros apropiado para combinar un flujo de datos con un vídeo arbitrario. El objetivo es permitir que el video presentado se perciba sin distorsión de color, artefactos y parpadeos cuando se muestra en una pantalla de alta frecuencia.

En segundo lugar, el canal secundario *Screen-to-camera* podría experimentar una posible interferencia del contenido de video original en el canal primario. Se necesita diseñar un esquema de codificación efectivo de trama de datos. Debe ser resistente tanto a los cambios en el contenido de video como a las limitaciones conocidas de comunicación entre la pantalla y la cámara (por ejemplo, falta de coincidencia de velocidad de fotogramas, efecto de obturador, mala calidad de captura). Este sistema aprovecha la discrepancia de capacidad y las características distintivas del

sistema de visión humana y los dispositivos (pantalla y cámara). Por ejemplo, las pantallas pueden mostrar el contenido más rápido de lo que perciben los ojos humanos; las cámaras tienen obturador, pero los ojos humanos no, etc. El vídeo y los datos transportados por los cuadros compuestos operan en diferentes escalas de tiempo. El contenido del vídeo se percibe a un ritmo lento debido a los límites físicos del ojo humano, mientras que los cuadros de datos se muestran a una velocidad rápida, que sólo puede ser capturada y decodificado por la cámara.

En el diseño de nuestro proyecto, superamos el primer desafío aprovechando la propiedad de fusión de parpadeo del sistema de visión humana, es decir, la exhibición de un efecto de filtro paso bajo al ver escenas que cambian más rápido que la frecuencia de parpadeo crítica, y la superior resolución temporal de las pantallas. De este modo, ideamos un concepto de *frame* complementario, incrustando un frame de datos en un par de frames de vídeo multiplexados. Un fotograma multiplexado tendrá artefactos obvios, pero cuando se reproduce lo suficientemente rápido, los artefactos de los frames complementarios se anulan cada uno. Se propone además una técnica de suavizado de frames para incrustar gradualmente los frames de datos en el contenido de vídeo. Para combatir el segundo desafío, diseñamos una estructura jerárquica del marco de datos y también esquemas especiales de codificación. En particular, el esquema de codificación utiliza los estados de encendido y apagado de un patrón artificial de tablero de ajedrez para señalar un -1/1 bit, donde el valor -1 actúa como nuestro 0 lógico.

En la comunicación entre pantalla y cámara, la información se modulará en función de la intensidad de la luz de los píxeles del transmisor (pantalla) que se reciben, y se decodificarán a partir de la intensidad de los píxeles de la imagen capturada por la cámara en el receptor. La intensidad del píxel capturado por la cámara es una cantidad digital Q1 que es proporcional a la cantidad de corriente de fotones generada en el píxel a partir de la energía lumínica acumulada en su área (cuanto más pequeña sea el área, menor será la intensidad de luz que acumule). Cuando el píxel de la pantalla que emite la luz está en el foco del objetivo de la cámara, todos los rayos de luz del píxel de la pantalla se enfocan en un píxel de la cámara y, por lo tanto, no se produce ninguna pérdida de energía en el píxel. Por otro lado, si el píxel de la pantalla es perturbado (en posición y/u orientación) por el enfoque de la cámara o incurre en una pérdida de energía debido al tamaño finito de la apertura del objetivo de la cámara, no todos los rayos de luz convergerán en el píxel de la cámara, lo que dará lugar a una reducción de la energía acumulada y, por tanto, a un menor valor de intensidad del píxel. La pérdida de la intensidad de la luz recibida en un píxel de la cámara da lugar a la deformación visual del tamaño o la forma del píxel de la pantalla; un efecto que se denomina distorsión de la perspectiva. Para controlar la transparencia de los píxeles se utilizó el canal alfa, ya que regula el cambio de transparencia de los píxeles para que sólo sea perceptible para las cámaras, pero no para los ojos humanos. Además, al controlar la translucidez de cada píxel independientemente, se permite que haya múltiples elementos transmisores simultáneos en la pantalla. Esto crea un canal de comunicación MIMO entre las pantallas y las cámaras, que puede aumentar la velocidad de los datos y mejorar la fiabilidad de la transmisión.

En resumen, nuestro sistema tiene dos canales de comunicación visual concurrentes, el canal primario *Screen-to-eye* para los humanos y el canal secundario *Screen-to-Camera* para los dispositivos. Ambos canales se originan en la misma pantalla, coexistiendo y contrarrestándose el uno al otro. Se ha implementado este sistema y realizado un estudio preliminar. Los experimentos realizados confirman que dicho sistema permite una comunicación visible de modo dual, tanto a los humanos como a los dispositivos, simultáneamente.

6.1. Justificación y objetivos del modelo

Hoy en día, las cámaras actuales se utilizan con frecuencia para leer códigos QR, que pueden considerarse a su vez como una forma de comunicación visual en la que la cámara actúa como un receptor. El uso generalizado de estos códigos motiva la construcción de nuevas aplicaciones de comunicación con cámaras, donde las pantallas podrían ser moduladas para enviar dichos códigos variables en el tiempo para ser decodificados por la cámara posteriormente.

Los patrones visuales espaciales (por ejemplo, los códigos de respuesta rápida QR anteriormente mencionados) no son del todo consumibles por los seres humanos, que se sienten más cómodos con los contenidos de fácil acceso, como textos, imágenes y vídeos. Cuando una pantalla se convierte en la misma fuente, tanto para el contenido primario favorable para las personas como para la información lateral óptima para las cámaras, ambos compiten por la pantalla. Reconociendo esa controversia y el hecho de que la entrega de contenido consumible por los seres humanos es más bien un propósito primario, en las prácticas actuales se ha recurrido a algún tipo de compromiso, ya sea espacial o temporal.

En el caso del código QR, el código tomaría una pequeña área y residiría en una esquina de toda la pantalla (Figura 34(a)), u ocuparía toda la pantalla a su vez después de los contenidos para humanos (Figura 34(b)). Sin embargo, la anterior comprobación espacial no sólo limita su capacidad de transporte de información, sino que también incurre en un trabajo adicional para capturar adecuadamente los códigos (por ejemplo, estar lo suficientemente cerca de los códigos QR).



(a) Spatial compromise



Figura 34. Las prácticas actuales de comprometer las comunicaciones de Screen-to-eye y de Screen-to-Camera. a) Un pequeño código QR en la esquina (en el caso de un partido de hockey sobre hielo); b) Dos frames consecutivos, uno para cada canal.

Sin embargo, el anterior compromiso espacial no sólo limita la capacidad de transporte de información, sino que también implica un trabajo adicional para capturar adecuadamente los códigos, como bien se dijo en el apartado anterior. Aunque pequeño, el aspecto del código QR se sigue considerando una distracción, lo que perjudica las experiencias de los usuarios. Sin embargo, esa práctica sólo es viable en los escenarios y en los dispositivos interactivos. En lo que cierne a los objetivos, tenemos dos concretos para nuestro caso de estudio. En primer lugar, el flujo de datos con información a través del canal Screen-to-camera no debería afectar a la percepción del usuario del contenido entregado a través del canal primario. En segundo lugar, tratamos de lograr una alta tasa de datos en el canal secundario de Screen-to-camera, lo que es vital para permitir nuevos escenarios de aplicación para la comunicación visual de datos en el futuro. Para ello, tenemos que gestionar eficazmente las interferencias del canal primario de Screen-to-eye. Hay que tener en cuenta que el contenido visual del canal primario inevitablemente interferirá con el canal secundario. Los elementos de la pantalla y la cámara poseen una gran cantidad de píxeles que pueden aprovecharse para enviar un elevado volumen de datos a través de códigos en 2D que varían con el tiempo (secuencia de imágenes). Por ejemplo, un usuario podría apuntar con una cámara a una pantalla de un ordenador de escritorio o a la pantalla de un smartphone que muestra el código variable en el tiempo para descargar un archivo. Investigaciones recientes han explorado más a fondo esta dirección diseñando prototipos en los que los códigos 2D variables en el tiempo pueden ser transmitidos desde los monitores [95,96] y pantallas de teléfonos inteligentes [97] a una cámara como receptor. Aunque estos trabajos han diseñado y medido la capacidad de información para una solución puntual específica, todavía no está claro cuánto se puede mejorar en estas situaciones o si hay algún límite en el rendimiento de dichos sistemas.

Por último, el sistema deberá:

- No ser detectable por los usuarios. La comunicación de datos entre la cámara y la pantalla deberá estar completamente oculta para los usuarios y no requerirá mostrar ninguna imagen codificada visible en la pantalla. Por lo tanto, la comunicación de datos no interferirá con ningún contenido que el usuario esté viendo en la pantalla, y se podrá incorporar fácilmente a cualquier pantalla existente (por ejemplo, la de los teléfonos inteligentes, ordenadores portátiles, ...) sin sacrificar su funcionalidad original, es decir, la visualización del contenido.
- Soportar cualquier escena. La comunicación de datos puede producirse sin importar el contenido de la pantalla, ya sea una imagen, videoclip, una película, una escena de juego, una ventana de aplicación de varias capas, o la pantalla de inicio de un teléfono inteligente o una tableta. Además, la comunicación continuará a medida que el contenido de la pantalla cambie sobre la marcha (por ejemplo, el usuario navega por una página web, juega y ve una película). Para que sea compatible con todos los tipos de contenidos, la comunicación debe ser independiente del contenido de la pantalla, que puede generarse sobre la marcha y variar con el tiempo a medida que el usuario interactúa con la pantalla.

• Operar con dispositivos comerciales. El sistema puede transmitir y recibir datos utilizando los dispositivos inteligentes existentes (por ejemplo, teléfonos inteligentes o tabletas). Estos dispositivos tienen una potencia de cálculo limitada, y sus cámaras no son tan sofisticadas como las cámaras réflex de una sola lente (SLR) de alta gama. Por lo tanto, los diseños de codificación y decodificación tienen que ser ligeros y robustos, de modo que los dispositivos inteligentes como transmisores y receptores puedan procesar los datos a la vez que soportan el contenido dinámico como el vídeo.

6.2. Limitaciones y soluciones del modelo

A la hora de diseñar un sistema S2C se deberá de tener en cuenta los siguientes objetivos para realizar correctamente un sistema de comunicación S2C.

- Procesamiento de datos dinámicos en tiempo real. El sistema puede transmitir y recibir datos sobre la marcha, en lugar de preprocesar una imagen o un archivo de vídeo para incrustar los datos. Esto es necesario cuando los datos llegan sobre la marcha o cuando el contenido de la pantalla no se conoce de antemano (por ejemplo, una escena de videojuegos. El reto es que cuando la pantalla muestra vídeo (por ejemplo, películas, videoclips), la frecuencia de fotogramas es de al menos 24 fotogramas por segundo (FPS). Por lo tanto, el sistema tiene que terminar de codificar los datos en un cuadro en un plazo de 42 ms antes de que la pantalla muestre el siguiente cuadro. Cuando el vídeo se reproduce a 60 FPS o 120 FPS, la codificación tiene que terminar dentro de 16 ms o 8 ms, por ejemplo.
- Operar con dispositivos inteligentes. El sistema S2C puede transmitir y recibir datos en tiempo real utilizando los dispositivos inteligentes existentes (por ejemplo, teléfonos inteligentes, tabletas). Estos dispositivos tienen una potencia de cálculo limitada, y sus cámaras no son tan sofisticadas como las cámaras réflex de una sola lente (SLR) de alta gama. Por ello, los diseños de codificación y descodificación tienen que ser ligeros y robustos, de modo que los dispositivos inteligentes como transmisores y receptores puedan procesar los datos a tiempo y al mismo tiempo soportar contenido dinámico como un video.



Figura 35. Ilustración de la distorsión de la perspectiva en el canal *screen-camera*. Los píxeles de la pantalla visualizados son borrosos y se reduce el tamaño en la vista frontal completa y también en forma en la vista angular.

- Distorsiones. Las distorsiones que dependen de la perspectiva de la cámara se producen debido a la naturaleza del mecanismo de obtención de las imágenes en la cámara y se manifiesta como una deformación del tamaño y la forma del objeto capturado (píxel de la pantalla que emite luz) en la imagen, lo que da lugar a una compresión o ampliación visual de la proyección del objeto en la imagen. Cuando la pantalla está a una distancia de desenfoque del objetivo de la cámara (o en un ángulo oblicuo), estas distorsiones se hacen evidentes provocando una interferencia entre los píxeles adyacentes de la imagen de la cámara, lo que se denomina como interferencia entre píxeles o IPI. El efecto combinado del ruido de fondo y el IPI degradará la calidad de la señal recibida, y, por lo tanto, reducirá la capacidad de información en los canales basados en OCC. Por ejemplo, consideremos que los bloques de píxeles de una pantalla son iluminados por un patrón de tablero de ajedrez y capturados por una cámara como se muestra en la Figura 35. Podemos observar que las distorsiones de la perspectiva hacen que los píxeles de la pantalla se deformen en tamaño cuando la pantalla no está en el foco de la cámara, y en forma cuando no está alineada frontalmente.
 - Escalado de la perspectiva. Si el píxel de la pantalla estuviera en el foco, y suponiendo que la pantalla y la cámara tienen la misma resolución, su imagen en la cámara debería ocupar la misma área que un píxel. Pero en realidad, los rayos de luz del píxel pueden no terminar exactamente en los límites del píxel de la cámara y existirán algunas zonas que acumulen interferencias. Estas áreas de desalineación y la geometría del píxel de la pantalla dependerán de la perspectiva y se debe a la distorsión debida a la escala de la perspectiva del área del píxel.
 - Lens-blur. Podemos observar en la Figura 35 que los píxeles de la pantalla se visualizan borrosos, especialmente en las regiones de transición entre los bloques blancos y negros. Este efecto de desenfoque se atribuye a la lente de la cámara y se denomina como

desenfoque de la lente (*lens-blur*). Este efecto se modela típicamente utilizando la función de dispersión puntal (PSF) [98], que representa la respuesta de un sistema de imágenes a una fuente puntual. En el canal *screen-camera* esto se traduce como la distorsión de los píxeles en las regiones de transición entre los píxeles más brillantes (alta intensidad) y los más oscuros (baja intensidad), y conduce a la interferencia (IPI) entre los píxeles vecinos, tal y como se puede observar en la figura anterior. Dado que el área y la máxima energía que puede ser muestreada en cada píxel de la cámara es finita, la IPI conduce a una reducción efectiva de la energía de la señal por píxel.

Para resumir, los diseños de comunicación S2C discretos existentes requieren modificaciones directas de los valores de color de los píxeles de contenido (RGB) y no transmiten datos dinámicos sobre un contenido dinámico arbitrario, en el que los datos o el contenido de la pantalla llegan sobre la marcha. Además, al crear una capa opaca con valores RGB modificados, estos diseños pueden anular la optimización de la representación existente. Para mitigar el problema, los diseños actuales tendrán que renunciar a la comunicación en pantalla completa y sacrificar la velocidad de los datos o la fiabilidad de la transmisión. Además, la integración de la comunicación con el contenido de la pantalla limita en gran medida sus escenarios aplicables.

Los píxeles de la pantalla y del sensor de imagen de la cámara pueden no tener el mismo tamaño exacto, e incluso si es así, pueden no ser exactamente rectángulos similares y pueden persistir algunas micro-desviaciones que pueden causar desalineaciones en los píxeles de la imagen. Esto puede causar que un píxel de la pantalla no se alinee con un píxel de la cámara, incluso si el píxel de la pantalla estuviera en el foco de la cámara. Tales desalineaciones causarán una desviación en el factor de distorsión para cada píxel a medida que la perspectiva cambia. Sin embargo, se puede suponer que esas desviaciones son insignificantes si se comparan con las distorsiones debidas a la escala de la perspectiva. Al considerar un factor de distorsión promedio sobre la imagen de la cámara, tales micro-desviaciones serán casi insignificantes.

Como solución se propone una comunicación desvinculada usando el canal alfa. Al operar en este canal el sistema ya no necesita modificar directamente los valores RGB de los píxeles, logrado el mismo efecto. Dado que los valores alfa son mezclados por la GPU, la codificación es casi instantánea, lo que es crítico para soportar la comunicación en tiempo real sobre cualquier contenido dinámico. En segundo lugar, al realizar la comunicación en una capa de imagen separada, el sistema hace que la comunicación discreta esté disponible universalmente y sea verdaderamente paralela al contenido que se reproduce en la pantalla, sin importar si se trata de un contenido estático, dinámico o de varias capas, e independientemente de la velocidad de fotogramas a la que se reproduce y de la resolución. Los usuarios podrán usar la pantalla tal como

está, mientras que la comunicación entre la pantalla y la cámara se produce detrás de la escena en tiempo real, de forma discreta.

Finalmente, para superar estas limitaciones, se propuso un nuevo paradigma de diseño para una comunicación S2C discreta. Motivados por el hecho de que la estrecha integración de la comunicación y el contenido de la pantalla es problemática, planteamos desacoplar la comunicación S2C de las capas de las imágenes del contenido de la pantalla. Además, diseñamos una codificación de datos ligera que aprovecha la potencia de la GPU y no requiera modificaciones directas de los valores RGB del contenido. Y lo que es más importante, dado que la mezcla de alfa es una característica que ofrece la GPU, los valores alfa pueden modificarse llamando a una función del sistema sin necesidad de modificaciones a nivel del sistema operativo.

6.3. Descripción del algoritmo propuesto

El objetivo de este proyecto, tal y como se ha mencionado en apartados anteriores, es el de construir una comunicación simultánea entre los usuarios y los dispositivos de pantalla. Se establece un canal de datos secundario encima del canal de vídeo primario, sin que ello produzca una interferencia en la experiencia de visualización del usuario, y sin dejar de poder transmitir la información. En nuestro escenario de uso, una cámara captura los datos de la pantalla, mientras el usuario está viendo un programa de televisión sin una degradación notable de la calidad.



Figura 36. Diseño de nuestro sistema. Usamos una pantalla de 60 Hz y vídeos de 50 FPS. Vi y Di representan el i-ésimo frame del vídeo original y de datos.

La idea clave de la propuesta es explotar la brecha de percepción entre la visión humana y los modernos sistemas de cámaras. La Figura 36 ilustra el diseño general que asegura dos objetivos de diseño en las comunicaciones *screen-camera* y *screen-eye*. Primero, nuestro sistema protege la experiencia de visualización de vídeo de la pantalla para los usuarios, mientras que incrusta los

bits de datos. Segundo, permite la comunicación de datos bajo las restricciones impuestas por el canal de vídeo primario.

Por consiguiente, el algoritmo propuesto requiere de dos partes esenciales: el codificador esteganográfico, que se encargará de la introducción del mensaje dentro de la imagen utilizada como cubierta, que realiza extensos análisis y comparaciones para la toma de decisiones de sus procesos y, por lo tanto, necesita un mayor tiempo de ejecución; y el decodificador esteganográfico, mecanismo inverso, pero mucho más sencillo y rápido, pues solamente requiere de una o más llaves, enviadas a través de otras vías, para el descifrado de la información útil.

En la descripción siguiente se utiliza el término de píxel para señalar un valor entre 0 y 255, como si se tratase de una imagen a escala de grises, para una mejor compresión del algoritmo, pero hay que tener en cuenta que el programa trabaja con imágenes a color RGB, que son matrices de 3 dimensiones, por lo que realmente el píxel está formado por tres valores entre 0 y 255 (24 bits) representantes del nivel de luminancia de cada color. Seguidamente, se ilustra una serie de conceptos que serán de vital importancia a la hora de entender el funcionamiento y la arquitectura del proyecto:

- Concepto de *frame* complementario. Siguiendo la propiedad de filtro paso bajo del sistema visual humano, si colocamos dos frames de vídeo seguidos, uno detrás del otro, siendo uno de ellos el inverso del otro, percibiremos un cuadro de color promedio uniforme. Esta observación origina los frames complementarios, concepto central de nuestro sistema. Primero, definimos los píxeles complementarios. Si dos píxeles p y p* se complementan entre sí con respecto al nivel de luminancia v, si sus valores de píxeles suman 2v, es decir, vp + vp* = 2v. En consecuencia, dos frames P y P* serán cuadros complementarios con respecto al nivel de luminancia v, si todos sus píxeles son complementarios respecto a este. Es fácil ver que, después de un filtrado paso bajo, dos frames complementarios permiten así agrupar adecuadamente los frames de vídeo y los frames de datos (por ejemplo, V ± D) sin producir distorsiones de color y cambios notables en el nivel de luminancia del vídeo.
- Fusión de parpadeo. Se presenta una nueva forma de incrustar datos en los frames de vídeo. En el capítulo 2 se expuso que el SVH es propenso a detectar parpadeos y a la interferencia visible. Aplicando el concepto explicado con anterioridad, generamos dos frames de datos más vídeo (Vi+Di, Vi+k+Di*) con respecto al frame de vídeo original Vi. Así mismo, obtenemos cada frame de datos Di, calculamos su frame complementario Di*, y luego integramos ambos en el frame del vídeo original. Nuestro estudio muestra que este esquema

es efectivo para lograr un efecto insignificante en las experiencias de visualización del usuario. La decodificación podrá llevarse a cabo siempre y cuando Vi y Vi+k sean suficientemente parecidos. El algoritmo que se propone tiene esto en cuenta desde un punto de vista de comunicaciones.

Suavizado de bloques. Una secuencia frames de datos variables {Di} podría imponer interferencias en la visión cuando cambia bruscamente de uno a otro (por ejemplo, de V1+D1* a V1+D2). Esto se debe a que la transición abrupta (efectos de matriz fantasma) puede causar parpadeos. En estos experimentos, se ha utilizado la forma de onda mostrada en la Figura 37 para mitigar dicho efecto, la cual se basa en una forma de onda triangular bipolar suavizada mediante un factor γ menor a la unidad. Por otra parte, la forma de onda dependerá del número de frames, *L*, que dure nuestro símbolo y vendrá determinada por la fórmula:

$$h[n] = R^{\gamma}[n] \tag{9}$$

$$R[n] = \begin{cases} \frac{1}{\frac{L}{4}-1} n & \sin \in \left[0, \frac{L}{4}-1\right] \\ 1 - \frac{1}{\frac{L}{4}-1} \left(n - \frac{L}{4}+1\right) & \sin n \in \left[\frac{L}{4}, \frac{3L}{4}-1\right] \\ -1 + \frac{1}{\frac{L}{4}-1} \left(n - \frac{3L}{4}+1\right) & \sin n \in \left[\frac{3L}{4}, L-1\right] \end{cases}$$
(10)



Figura 37. Modelamiento de la forma de onda del suavizado temporal.

Finalmente, se generan múltiples frames de datos intermedios, que se multiplexan con los frames de vídeo para asegurar una transición imperceptible entre dos frames de datos diferentes. Observe que, el anterior suavizado también facilitaba la mitigación del desajuste entre la tasa de refresco de la pantalla y la tasa de captura de la cámara. Si los datos cambian demasiado rápido, la cámara no puede capturar todos los datos, debido al efecto de "*rolling shutter*". De esta manera, con la transición suave se consigue reducir las interferencias entre frames de datos adyacentes.

• Canal Alpha. El canal alpha es una técnica estandarizada ampliamente utilizada en los gráficos por ordenador. Diversos autores han utilizado el canal alpha para formar una imagen compuesta con transparencia parcial o total [99-101]. Mientras que el elemento de la imagen almacena los valores de intensidad de color en los diodos rojos, verdes y azules (RGB) para cada píxel de la imagen, el canal alfa almacena un valor (α) entre 0 y 1 para indicar la translucidez del píxel. Un valor de α de 0 significa que el píxel de la imagen es totalmente transparente, y 1 (valor por defecto) significa que el píxel es totalmente opaco. Como ejemplo, la Figura 38 muestra la pantalla de un Samsung Note 3, donde se establece uniformemente los valores de α de la parte derecha de la pantalla en 0,5, lo que resulta en una apariencia más oscura.



460 440 420 400 400 0.4 0.5 0.6 0.7 0.8 0.9 1 α

(a) Changing α values of the screen from 1, 0.99, to 0.5

(b) Color intensity perceived by the receiver (camera)

Figura 38. El impacto de cambiar los valores de los píxeles de α . (a) Muestra una imagen con diferentes valores de α : el valor predeterminado 1 (la izquierda), 0,99 (el centro) y 0,5 (la derecha), utilizando un Samsung Note 3; (b) Muestra la intensidad de color del cuadro capturado por el Samsung S5 al variar los valores de α en la pantalla del transmisor.

Como se muestra en la figura anterior, el aumento del valor de α lleva a un aumento lineal de la intensidad del color percibida por la cámara del receptor. Esto se debe a que el valor de α se utiliza para multiplicar los valores de color RGB. Sin embargo, al usar el canal alpha completo hace que aumente la cantidad de energía que puede percibir un observador. Motivado por las observaciones anteriores, el sistema propuesto únicamente usa el canal alpha en el canal azul, reduciendo considerablemente la perceptibilidad a costa de precisar un cierto de contenido de azul en las imágenes para poder codificar.

6.4. Arquitectura del transmisor

El sistema propuesto está implementado en MATLAB. y consta de un emisor y un receptor. En este caso se presta especial atención al emisor. Éste toma como entrada un flujo de vídeo y un flujo de datos, genera el flujo multiplexado y luego reproduce el flujo de vídeo. El sistema de codificación tiene como cometido tomar el flujo de datos de entrada, distribuirlos en bloques y combinar dichos bloques con los diferentes frames de una señal de vídeo, para de esta manera, integrar la transmisión de la información en la emisión de imágenes emitidas por una pantalla. El objetivo es que la señal de datos quede embebida en las imágenes emitidas de una forma inapreciable por el espectador, como se ha comentado en los apartados anteriores de este trabajo. El conjunto de bloques en los que se ha desarrollado el codificador se detalla a continuación en la Figura 39.



Figura 39. Diagrama de bloques del sistema transmisor.

Decisor. En las comunicaciones inalámbricas, la relación señal/interferencia (SIR) es un parámetro predominante que caracteriza el rendimiento del sistema. Especialmente para las cuestiones de asignación de recursos radioeléctricos, como el traspaso, la asignación dinámica de canales, el control de potencia, etc. El sistema debe tener en cuenta las mediciones de la SIR. La información precisa de la calidad de los enlaces radioeléctricos es la clave para prestar servicios de alta velocidad de datos. Sin embargo, en la práctica, la estimación de la SIR requiere de muchos cálculos. El mayor desafío para su estimación es encontrar una forma eficiente de separar la señal deseada de la interferencia.

Este apartado se encarga de cuantificar la calidad de la señal en el transmisor, usando un promedio de la SIR para determinar si el símbolo puede ser codificado o no a partir de un umbral dado. La SIR, como se mencionó anteriormente, es una pieza importante en el mundo de la ingeniería de las comunicaciones, ya que indica la calidad de un enlace entre un transmisor y un receptor. La mayoría de los métodos actuales resolvieron el problema de la estimación del SIR separando la señal portadora de información y la interferencia más el ruido de la señal total recibida. Una vez que se conoce la parte de la señal o la parte de la interferencia más el ruido, podemos obtener fácilmente el SIR del canal.

En lo referente al proyecto, la interferencia vendrá dada por la diferencia entre las imágenes iniciales y finales de cada símbolo. La potencia dependerá de los píxeles que superen el umbral de sensibilidad del canal azul (los valores mínimos y máximos de azul para asegurar la detección del código es de 50 y 250), de la conformación de pulsos y del valor de alpha utilizado. De esta forma obtenemos la SIR:

$$\Delta \mathbf{I} = I_i - I_f \tag{11}$$

$$I = E[\Delta I^2] \tag{12}$$

$$\beta = \frac{\sum \delta_i}{N_x N_y} \tag{13}$$

$$\delta_i = \begin{cases} 1 & si \ l_i^b \in \mathfrak{S}, l_f^b \in \mathfrak{S} \\ 0 \end{cases}$$
(14)

$$\mathfrak{S} = (s_{min}, s_{max}) \tag{15}$$

$$S = \beta \cdot \alpha^2 \tag{16}$$

$$SIR = 10 \log\left(\frac{S}{I}\right)$$
 (17)

Donde, ΔI es la imagen diferencia de la imagen inicial (I_i) y final (I_f) de un símbolo cualesquiera, *I* es la interferencia, N_x y N_y son el número de píxeles X e Y de la imagen respectivamente, \mathfrak{S} contiene los rangos donde debe de estar acotada la imagen en relación a los máximos y mínimos niveles de azul, *S* es la señal de potencia, β es la cantidad de píxeles que superan el umbral de detección del canal azul y α es el valor de intensidad por el que va multiplicado nuestro código en el canal azul, referente a la invisibilidad de este. Finalmente, al codificar únicamente aquellos fragmentos de vídeo que presenten una SIR estimada superior a un umbral, se mejorará a priori la calidad de la recepción, reduciendo así la BER (*Bit Error Rate*) en el receptor.

• Encoder. Se encarga de codificar los datos entrantes en códigos, que denominaremos códigos expandidos). Para ello se divide el flujo binario en cadenas de N bits (en nuestro caso 4 bits, aunque se harán pruebas con otros valores) y las expande, siguiendo estrategias de Espectro Ensanchado por Secuencia Directa (DSSS de sus siglas en inglés, *Direct Sequence Spread Spectrum*). El concepto de expansión hace referencia a que para cada bit de la cadena se generan M bits. Cada código, por tanto, representa un bit, y se sitúa en una celda de la imagen. De ahí que el código expandido tenga una longitud M de 16 bits. Por tanto, si se codifican 4 bits, se tienen 4 celdas independientes en la imagen, y si cada celda cuenta con 16 ranuras, la longitud del código final será de $L = N \cdot M$. Estos códigos de 4 bits (que corresponden a una celda) siguen dos configuraciones: el código seleccionado para el bit 1 será el que corresponde espacialmente a la configuración del tablero de ajedrez.

Del mismo modo, el encoder utiliza un esquema diferencial, usando matrices de Hadamard de rango 2^{n+1} y la distribuimos en toda la imagen porque tienen energía nula y poseen la característica de ortogonalidad. En principio codificamos N bits en toda la imagen, pero esta se puede subdividir. A continuación, se emplean las técnicas de codificación empleadas:

 Código Hadamard. Estos códigos [102] son equivalentes a las secuencias de Walsh o a los códigos Reed-Muller de primer orden. Se han utilizado ampliamente en los sistemas de comunicación para la sincronización y la difusión del ancho de banda. Sin embargo, la ganancia de este tipo de codificación es relativamente baja.

Una de Hadamard de orden n es una matriz H de tamaño $n \ge n$ con entradas +1 y -1, tal que:

$$H * H^t = n \cdot I \tag{18}$$

Es decir, una matriz es de Hadamard si el producto interno de dos distintas filas es 0 y el producto interno de una fila consigo misma es n.

$$H = \begin{pmatrix} + & + \\ + & - \end{pmatrix}_{2x2}, \quad H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}_{2x2}, \quad n = 2, \quad H^{t} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$
$$H * H^{t} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 2 I = nI$$
(19)

90

De la definición se sabe que cualesquiera de las dos columnas o filas son ortogonales. Esta propiedad no cambia si permitamos filas o columnas o si multiplicamos alguna fila o columna por -1 (estas matrices se llaman equivalentes). Esto significa que el producto escalar de dos filas o dos columnas distintas entre sí será cero. En definitiva, estos códigos tienen parámetros óptimos para transmitir y modular información a través de un canal ruidoso.

Codificación Diferencial. La codificación Manchester [103] es un método de codificación de datos digitales de una serie de bits arbitrarios, sin que se presenten largas cadenas de ceros y unos, evitando así ambigüedades en la recepción. Además, este tipo de codificación ayuda a eliminar el efecto de parpadeo de la luz.

En este tipo de codificación, cada bit de información es transmitido en un tiempo fijo de (periodo T), el periodo de bit se divide en dos intervalos iguales, donde: un uno binario es representado por un pulso donde la primera mitad del periodo tiene un estado a nivel alto y la segunda mitad tiene un estado bajo. Por otro lado, un cero binario se representará por un pulso donde la primera mitad del periodo tiene un estado bajo y la segunda mitad es un estado alto.

Este método asegura que cada periodo de bit codificado tenga una transición a la mitad. La transición contiene la información, así: para un uno binario se realiza una transición alto/bajo y para un cero se realiza una transición bajo/alto. No siempre ocurre una transición al inicio o final del periodo del bit codificado. Cuando ocurre solo sirve para ubicar la señal en el estado correcto para la transición de la mitad del periodo. Estas transiciones tienen como objetivo realizar dos funciones: sincronizar la información y transmitir los datos.

Finalmente, la codificación Manchester es una codificación sin memoria, pues no requiere información del bit anterior codificado. También es conocida como una codificación bifase, donde un 1 binario se representa por una onda cuadrada con una fase 0° y un 0 binario es representado por una onda cuadrada con una fase de 180°.

Técnicas de espectro ensanchado. CDMA [104] proviene de utilizar técnicas de espectro ensanchado, en particular secuencia directa. La idea general del proceso es muy simple: dada una señal digital que se desea transmitir, se aplica una técnica de secuencia directa para obtener una señal digital con un ancho de banda mayor. Dicha señal es transmitida y

el receptor recupera la señal original mediante el uso de un correlador y la misma clave que se utilizó para la transmisión.

En CDMA, el transmisor asigna una secuencia de código única que se usa para la codificación de la señal, que contiene la información. En la parte del receptor, conocida ya la secuencia, se decodifica la señal obtenida y se recuperan los datos originales. En este tipo de técnica, las señales comparten tiempo y frecuencia, y se diferencian entre sí mediante la asignación de códigos binarios ortogonales.

El ancho de banda de transmisión es mucho mayor que el ancho de banda de la señal que contiene la información, por lo que el proceso de codificación realiza un ensanchamiento del espectro de la señal original. En general, cualquier técnica de espectro ensanchado debe verificar dos condiciones: el ancho de banda de transmisión debe de ser mucho mayor que el ancho de banda de información y que el ancho de banda de transmisión es estadísticamente independiente de la señal que contiene la información. Del mismo modo, este tipo de técnicas presentan una serie de características que las hacen únicas. En primer lugar, partiendo de las propiedades de la correlación cruzada entre señales, confiere a las interferencias naturaleza determinista, no aleatoria. En segundo lugar, este tipo de técnicas dotan a los sistemas de protección contra interferencias.

- Masker. Este apartado se encarga de adaptar los códigos expandidos a la imagen en sí. Dicho con otras palabras, este bloque se encarga de generar una máscara 2D que tiene las mismas dimensiones que la imagen. Para la generación de estas máscaras, el bloque grid representará un grid de colocación de las celdas en la imagen (cada celda será la representación 2D de un bit del código correspondiente). Esta máscara deberá pasar por una serie de rutinas de filtrado que permitan reducir la variabilidad espacial y temporal del código, esto es, que permitan suavizar los bordes y las transiciones de las celdas, de tal forma que se minimice la capacidad de detección de los códigos por el ojo humano.
- Blender. Se encarga de mezclar la imagen con la máscara generada por el Masker. En este caso, se utiliza una mezcla aditiva. Este tipo de mezcla se explica como la combinación de determinadas cantidades de luz (rojo, verde y azul), con objeto de crear nuevas composiciones de colores. Si se mezclan las tres fuentes de luz en su máxima intensidad, el ojo humano percibirá el color blanco como resultado. La mezcla de los mismos colores con menor intensidad se percibirá como un gris neutro y si se apagan las tres fuentes se logra el negro. Por otra parte, si sólo una de las tres fuentes de luz está apagada y las otras dos emiten con su intensidad máxima, se pueden obtener mezclas como el amarillo (rojo y verde), el cian (azul

y verde) o el magenta (rojo y azul). Las distintas combinaciones de dos o tres colores primarios de fuentes luminosas en sus diferentes intensidades permiten reproducir en cualquier pantalla la mayoría de los colores. Por tanto, la mezcla aditiva se utiliza en los monitores de ordenadores, televisores, tabletas, etc. La pantalla de un monitor está compuesta por un cierto número de píxeles, y cada píxel contiene tres pequeñas fuentes luminosas RGB. La mezcla de los colores de estas tres fuentes luminosas le da al píxel su color específico. Por ende, la imagen generada será la imagen original más la máscara superpuesta, la cual está multiplicada por un cierto factor, el factor alpha, α . Este factor, tal y como se detalló anteriormente, es un parámetro de crucial importancia, puesto que juegan un papel determinante tanto para que los códigos puedan ser detectados por la cámara, como para que éstos no sean apreciables por el ojo humano. La discusión del valor de este factor se detallará en los siguientes apartados.

- Localizadores. Los localizadores o códigos ancla, son códigos específicos que se añaden a las máscaras de los códigos a transmitir y que tienen dos propósitos:
 - Por un lado, permiten ubicar el contenido a decodificar en la imagen del vídeo.
 - Por otro lado, indican al receptor si hay un código incrustado en la imagen, de tal forma que este pueda proceder a la decodificación de los datos de una forma más sencilla.

Durante la fase de pruebas se determinó la necesidad de indicar al receptor la presencia de información incrustada. Para ello se incluyeron 4 anclas visibles en las esquinas de la imagen (Figura 40). Estas anclas no estarían visibles en el sistema final, ya que sólo serían de utilidad en el apartado de pruebas. Algunos autores usan anchor bits para facilitar el proceso de detección y seguimiento de la ROI [105], pero en este trabajo no se ha llevado a cabo este procedimiento, ya que no era el objetivo del mismo. En el primer acercamiento a estos sistemas de transmisión se consideró necesario el uso de alguna estrategia para indicar al receptor de que las imágenes que estaba detectando contenían algún código útil. Como solución se proponen estos localizadores, los cuales siguen el mismo patrón que el código utilizado para la transmisión del símbolo. Sin embargo, para indicar su color al verde. Concretamente, en la primera imagen con código se utilizará el localizador rojo, y en el final del símbolo, el mismo localizador, pero de color verde. Cuando no existan códigos útiles no habrá ningún localizador en la pantalla. Estos códigos únicamente han sido añadidos de cara a la fase pruebas con el fin agilizarla, realizando estas pruebas de

una forma más correcta y sencilla. En lo que respecta al sistema final, estos códigos anclas no están implementados.



Figura 40. Tipos de localizadores usados en las pruebas.

6.5. Arquitectura del receptor

A diferencia del transmisor, el receptor toma como entrada los frames capturados, detecta la existencia de patrones de tablero de ajedrez y decodifica los frames de datos recuperados. El proceso de recepción se puede dividir en dos etapas, descubrimiento y adquisición. El receptor utilizará la longitud del código del transmisor para diferenciar las diferentes tramas de frames de información. En la etapa de descubrimiento, el receptor correlaciona la información que tiene del transmisor como una plantilla y si excede de un valor umbral de detección impuesto, se considera que la fuente de información no habrá sido detectada con éxito. Por otro lado, en la etapa de adquisición, el receptor correlaciona la plantilla con la ROI. Cuando se encuentra una plantilla, se sincroniza. Esto reduce la carga de cálculo necesaria para el preprocesamiento de toda la imagen. Finalmente, se realiza el proceso de binarización y el ensamblaje de datos.

A continuación, se exponen algunas consideraciones que deben de tenerse en cuenta a la hora de implementar la recepción de la señal. El conjunto de bloques en los que se ha desarrollado el receptor se detalla a continuación:



Figura 41. Diagrama de bloques del sistema receptor.

- Adquisición de la imagen. La imagen es capturada. Esta consiste en una matriz de 1280x720x3 que contiene la información de los canales RGB. Durante la transmisión, la adquisición de la imagen se realiza con el mínimo tiempo de exposición disponible de la cámara. Durante el estado de detección de la ROI, el tiempo de exposición se incrementa para facilitar el algoritmo de detección de la ROI.
- Detección de la ROI. El uso de la ROI para detectar objetos ha sido el foco de investigaciones durante muchos años. Casi todos los algoritmos de ROI utilizan dos enfoques: basado en objetos y basado en características [106]. El enfoque basado en características busca píxeles que tienen características ópticas significativas y luego usa el análisis y un objeto para hacer la ROI. Es capaz de capturar casi todos los píxeles del objetivo basándose en la similitud de las características ópticas. Dado que no todos los píxeles tienen una característica marcada que se muestre en los píxeles ópticos, la detección mediante métodos basados en características no suele ser suficiente para identificar todo el objetivo. Por el contrario, el método basado en objetos puede detectar la ROI mejor que el método anterior, ya que utiliza la información sobre la estructura y las formas para la detección. Los métodos basados en ROI se desplegaron procesando sólo un subconjunto de píxeles para hallar LEDs en las imágenes. A partir de ahí, se detecta la posición de las señales y decodifica en el lado del receptor. En este trabajo, la ROI ha sido definida de manera supervisada, indicando los cuatros puntos que delimitan el área de la imagen y calculando apropiadamente las zonas de muestreo de cada uno de los chips que conforman la transmisión.
- Discusión sobre el efecto del movimiento del receptor. Las aplicaciones de comunicación *Screen-to-camera* normalmente implican un cierto grado de movimiento, por ejemplo, cuando la cámara se sostiene en la mano, o cuando la pantalla o la cámara están situadas sobre un objeto en movimiento. El movimiento debido a las sacudidas de la mano o a los movimientos laterales puede causar un cambio drástico de enfoque entre la pantalla y la cámara. En tal caso, se pueden asumir algunas oscilaciones en los píxeles, especialmente cuando la cámara no es estable, donde los píxeles parecen interferir entre sí, causando eventualmente un efecto visual borroso en la imagen, lo que se conoce como *motion-blur*. Este efecto es más notable en cámaras global shutter que en cámaras Rolling shutter, por lo que en general puede despreciarse en este trabajo. En definitiva, este fenómeno afecta de tal forma que los símbolos codificados no estarán perfectamente alineados y se generará una interferencia.

6.6.Implementación

La implementación de los algoritmos utilizados queda reflejados en el ANEXO digital. En este anexo se podrán encontrar las funciones principales del transmisor y del receptor usados en el sistema. Además, se podrán encontrar las funciones complementarias que hacen que dicho sistema consiga una codificación visualmente imperceptible para los usuarios. A modo de resumen se presentan las funciones que han sido utilizadas para la creación de este sistema:



Figura 42. Diagrama del flujo de trabajo de la algoritmia utilizada.

7.1. Objetivos de la validación experimental

El objetivo de estos experimentos es el de la evaluación de la calidad de servicio (QoS) de las comunicaciones esteganográficas, al mismo tiempo que se cumplen los criterios de visibilidad impuestos, con el fin de no afectar a la calidad de la experiencia de usuario (QoE).

7.2. Configuración experimental

En esta sección se explica el entorno de pruebas usado para la demostración de la funcionalidad del algoritmo propuesto. Dentro de los trabajos llevados a cabo, se incluye un escenario de experimentación del funcionamiento del sistema, que se realizó según la configuración que se muestra en la Figura 43. En esta figura se observa que la pantalla está a una distancia suficiente (70 centímetros) para que el sistema de captura sea capaz de englobar toda la escena. El móvil se encuentra en una posición central enfrentado al monitor.



Figura 43. Ejemplo de la configuración experimental.

En este experimento se utilizaron 4 vídeos codificados y comprimidos a 50 FPS de una duración de 15 segundos, y se grabaron con una cámara convencional del móvil Huawei Mate 10. Dichos vídeos (Figura 44) Pueden clasificarse atendiendo a la cantidad de movimiento aparente: bajo, medio y alto. Por otro lado, análogamente al movimiento, podemos dividir los vídeos en función de sus niveles de intensidad o colorimetría, en los que encontramos: vídeos con colores fríos, colores neutros, colores cálidos y vídeos con alto nivel de contraste. Los primeros, son todos aquellos en los que esté presente el azul, verde, turquesa, cían o violeta, es decir, los relacionados

con una temperatura baja. Por otro lado, los colores cálidos son la otra mitad del espectro, el color rojo, naranja y amarillo. Del mismo modo, los colores neutros son todos aquellos colores que no compiten en un espacio y que combinan con cualquier color adicional. Funcionan como base para agregar color y texturas. Entre ellos se encuentran el blanco, gris, negro, marrón y beige. Por último, los vídeos con alto nivel de contraste equivalen a una gran diferencia relativa en la intensidad entre dos puntos de la imagen.

En este proyecto se usó como pantalla transmisora de información la del *MacBook Air* de 13 pulgadas del 2015, la cual es una pantalla retroiluminada por LED y compatible con millones de colores. La resolución de pantalla utilizada durante los experimentos fue de 1440 x 900 pixels. Por otro lado, se empleó la cámara que se encuentra integrada en el dispositivo móvil *Huawei Mate 10* como dispositivo de captura. Los vídeos fueron capturados con una resolución de 1280 x 720 pixels, a 30 fps y la configuración de la exposición y el balance de blancos de la cámara se configuran de forma automática (configuración predeterminada en los dispositivos *Android*).



a) sea.mp4



b) walk.mp4



c) flower.mp4



d) birds.mp4

Figura 44. Vídeos utilizados para codificar.

7.3. Descripción de los experimentos

El motivo principal de los experimentos fue determinar la capacidad y la calidad de un canal de estas características. Se realizó una evaluación en dos aspectos. Por un lado, se llevó a cabo un estudio de usuario para validar si nuestro diseño asegura una visualización normal de vídeo sin

degradación de la calidad o interferencias, y, por otro lado, evaluamos el rendimiento de la comunicación de datos.

Todos los experimentos se llevaron a cabo bajo las mismas condiciones de iluminación ambiental con las mediciones tomadas en el interior de un laboratorio equipado con iluminación fluorescente en el techo. Fijamos la cámara en un trípode para asegurar la menor cantidad de error en la medición de la distancia y el ángulo entre la pantalla y la cámara. Hemos utilizado un generador de datos pseudo-aleatorios con una semilla preestablecida para generar los cuadros de datos originales.

Nuestro sistema tiene varios parámetros sintonizables: el factor alpha, la SIR umbral y el tamaño del símbolo (FP). De esta forma, se realizarán todas las combinaciones posibles con los valores de alpha (3, 5 y 10), de SIR umbral (10%, 50% y 90%) y de longitud del símbolo (7, 14, 27). Se han elegido estos valores en función del contraste de Michelson y del teorema del muestreo de Nyquist. De esta manera, por cada vídeo para codificar, se crearán 27 vídeos en respuesta de estas combinaciones. Una vez creados los 27 vídeos, se unieron todos en un único archivo de vídeo con una separación entre vídeo y vídeo de 2 segundos de negro, ya que esto facilita la tarea de recepción. Posteriormente, una vez renderizado el vídeo final con todos los vídeos codificados, fue grabado por nuestro sistema de captación (móvil) y decodificado por el sistema receptor (decoder), obteniendo así las métricas que determinarán la calidad de la codificación. Cuando se realiza la codificación se guardan las variables de los bits enviados, y con estas variables se comparan con las métricas obtenidas en el receptor.

7.3.1. Sostenibilidad de la QoE

La evaluación de la calidad subjetiva se realizó a través de un sencillo estudio de percepción visual sobre un conjunto de 90 usuarios, de edades comprendidas entre 18 y 60 años. El experimento consistió en presentar a los usuarios un conjunto de vídeos, entre los que se encontraban el original y un subconjunto de aquellos con información embebida (considerando únicamente la longitud de los símbolos y la intensidad). Los vídeos se presentaron de manera desordenada, y los sujetos tuvieron que identificar aquellos en los que percibían algún tipo de distorsión (parpadeo). El diseño experimental es ciego, ya que los usuarios no conocían en qué posición se encontraba en vídeo original, pero sí los diseñadores del experimento.

7.3.2. Mejora de la QoS

La calidad de servicio (QoS) en este trabajo se analizó a través de una métrica de estimación de la probabilidad de la presencia de un símbolo en el buffer de imágenes presentado en el capítulo de Diseño del Sistema. Esta métrica se puede observar en la siguiente ecuación.

$$M(x) = N \frac{p(C_{max}|x)}{\sum p(C_j|x)}$$
(17)

Donde N es el número de símbolos de la codificación de Hadamard, C_i es el símbolo con mayor probabilidad y $p(C_j | x)$ es la probabilidad de detectar el símbolo j-ésimo dada la información almacenada en el buffer x. Esta estimación de probabilidad se ha obtenido empleando una etapa *softmax* a la salida del correlador de *Hadamard*. Se observó que esta métrica era capaz de discernir apropiadamente tanto la presencia de información en el vídeo (sincronización), como el símbolo concreto. En esencia, M(x) deberá ser elevada cuando dentro del buffer exista un símbolo, y deberá ser cercana a la unidad cuando no lo haya (todos los símbolos son equiprobables).

7.4. Análisis de datos

7.4.1. Test de hipótesis sobre la perceptibilidad de los códigos

Una vez recopilados los datos del experimento de QoE, se procedió a realizar una comparación uno a uno entre los resultados de proporción de percepción de parpadeo entre cada uno de los vídeos codificados y el vídeo original. De esta manera el problema se reduce a hacer un test de hipótesis paramétrico de igualdad de proporciones. Matemáticamente, se quiere empleará la siguiente hipótesis nula.

$$H0:\pi_{orig} = \pi_{code} \tag{18}$$

Donde π_{orig} y π_{code} son las proporciones anteriormente mencionadas para los vídeos original y codificado respectivamente. Este test se puede llevar a cabo mediante una *T-Student*, donde el estadístico de contraste se corresponde con la siguiente ecuación.

$$Z = \frac{p_{orig} - p_{code}}{\sqrt{p \left(1 - p\right) \left(\frac{1}{n_{orig}} + \frac{1}{n_{code}}\right)}}$$
(19)

Donde n_{orig} y n_{code} son los tamaños muestrales de los grupos original y codificado, y p es la proporción muestral expresada en la siguiente ecuación.

$$p = \frac{p_{orig}n_{orig} + p_{code}n_{code}}{n_{orig} + n_{code}}$$
(20)

Para cada valor de Z obtenido se calculará el p-value asociado, o la probabilidad de equivocarse al rechazar la hipótesis nula. El p-value es la probabilidad de equivocarse al rechazar la hipótesis nula. En nuestro caso, interesa que el p-value sea alto para así no poder asegurar que existe un efecto que modifique la QoE.

7.4.2. Análisis de detectabilidad

Para demostrar que el procedimiento de inserción de información basado en la predicción de la SIR en el receptor funciona, la distancia entre las funciones de distribución correspondientes a la presencia de símbolo, $F(M \mid symbol)$ y la ausencia de los mismos $F(M \mid no_symbol)$, debe crecer respecto al nivel umbral de SIR.

Esto implica que la probabilidad de tener un símbolo que se detecte con facilidad debe aumentar con la selectividad del algoritmo de decisión de inserción de información desarrollado en este trabajo.

7.5.Resultados

7.5.1. Resultados de perceptibilidad

Una vez realizado el estudio se obtuvieron los siguientes datos, en los que "SI" significa que el espectador es consciente de alguna distorsión en la imagen y "NO" es que no ha notado ningún elemento que moleste su visualización.

VÍDEO 1 - SEA		VÍI	DEO 2 – W	ALK	VÍDI	EO 3 - FLOWER VÍDEO 4 -			DEO 4 - B	IRDS	
	SI	NO		SI	NO		SI	NO		SI	NO
Vídeo			Vídeo			Vídeo			Vídeo		
1	42,2 %	57,8 %	1	48,9 %	51,1 %	1	16,7 %	83,3 %	1	12,2 %	87,8 %
Vídeo			Vídeo			Vídeo			Vídeo		
2	11,1%	88,9 %	2	16,7 %	83,3 %	2	20 %	80 %	2	47,8 %	52,2 %
Vídeo			Vídeo			Vídeo			Vídeo		
3	21,1 %	78,9 %	3	32,2 %	67,8 %	3	14,4 %	85,6 %	3	6,7 %	93,3 %
Vídeo			Vídeo			Vídeo			Vídeo		
4	64,4 %	35,6 %	4	23,3 %	76,7 %	4	20 %	80 %	4	17,8 %	82,2 %
Vídeo			Vídeo			Vídeo			Vídeo		
5	14,4 %	85,6 %	5	33,3 %	66,7 %	5	15,6 %	84,4 %	5	46,7 %	53,3 %
Vídeo			Vídeo			Vídeo			Vídeo		
6	13,3 %	86,7 %	6	57,8 %	42,2 %	6	13,3 %	86,7 %	6	13,3 %	86,7 %
Vídeo			Vídeo			Vídeo			Vídeo		
7	27,8 %	72,2 %	7	16,7 %	83,3 %	7	21,1 %	78,9 %	7	8,9 %	91,1 %
Vídeo			Vídeo			Vídeo			Vídeo		
8	73,3	26,7 %	8	20 %	80 %	8	18,9 %	81,1 %	8	20 %	80 %
	%										
Vídeo			Vídeo			Vídeo			Vídeo		
9	14,4	85,6 %	9	18,9 %	81,1 %	9	15,6 %	84,4 %	9	52,2	47,8 %
	%									%	
Vídeo			Vídeo			Vídeo			Vídeo		
10	37,8	62,2 %	10	62,2 %	37,8 %	10	15,6 %	84,4 %	10	13,3	86,7 %
	%									%	

Una vez obtenido dichos datos, se calculó para cada uno de ellos la probabilidad de equivocarse o rechazar la hipótesis nula, obteniendo así los siguientes resultados en función de cada vídeo:

Vídeos codificados	Orden de visualización	p-value
sea_FPS27_A10	Vídeo 1	0
sea_FPS27_A3	Vídeo 2	0.6532
sea_FPS14_A5	Vídeo 3	0.1691
sea_FPS7_A10	Vídeo 4	0
sea_FPS7_A3	Vídeo 5	0.8313
sea_original	Vídeo 6	1
sea_FPS27_A5	Vídeo 7	0.0182
sea_FPS14_A10	Vídeo 8	0
sea_FPS14_A3	Vídeo 9	0.8313
sea_FPS7_A5	Vídeo 10	0.0003

Tabla 5. Resultados obtenidos para los vídeos de "sea".

Cuanto más cercano esté a 1 el porcentaje de error significará que más cercano está al vídeo original, ninguna percepción de distorsión en la imagen. Por el contrario, cuando más cercano esté a 0, significará que el código incrustado es perceptible para los usuarios. Los valores marcados en verde serán aquellos que determinen la calidad de la QoE (valores mayores a 0.25).

En el caso de "sea" se puede observar que todos los vídeos codificados con un valor de alpha 10 o 5 son perceptibles por los usuarios, esto se debe a que el vídeo tiene muchas componentes de azul y conlleva a una mayor energía. Por otro lado, los datos codificados con un valor de alpha 3 funcionan muy bien independientemente de la longitud del símbolo, ya que no todos los usuarios son capaces de verlo. A medida que aumenta el alpha empeora la percepción visual del usuario para el vídeo de "sea".

Tabla 6. Resultados obtenidos para los vídeos de "walk".
--

Vídeos codificados	Orden de visualización	p-value
walk_FPS27_A10	1	0.0001
walk_FPS7_A5	2	0.5688
walk_FPS27_A5	3	0.0657
walk_FPS27_A3	4	0.5923
walk_FPS14_A5	5	0.0466
walk_FPS7_A10	6	0
walk_FPS7_A3	7	0.5688
walk_original	8	1

walk_FPS14_A3	9	0.8525
walk_FPS14_A10	10	0

En el caso del vídeo de "walk", se puede observar que se obtienen mejores valores que en el caso de "sea". Los vídeos codificados con valores de alpha 10 y 5 siguen siendo visibles por los usuarios, pese el movimiento que existe en este vídeo. Además, se sigue observando que, para valores de longitud de símbolo elevadas (sea el caso de 27 y 14) se siguen obteniendo peores resultados, que pueden ser debidos al alto dinamismo del vídeo en cuestión. En definitiva, los valores de alpha 3 siguen liderando los resultados, ya que son casi imperceptibles por los usuarios.

Vídeos codificados	Orden de visualización	Porcentaje de error
birds_FPS7_A5	1	0.4731
birds_FPS27_A10	2	0
birds_FPS27_A3	3	0.5835
birds_FPS14_A5	4	0.0827
birds_FPS7_A10	5	0
birds_FPS7_A3	6	0.3500
birds_original	7	1
birds_FPS27_A5	8	0.0370
birds_FPS14_A10	9	0.3500
birds_FPS14_A3	10	0

Tabla 7. Resultados obtenidos para los vídeos de "birds".

En este caso, se observa lo mismo que en los casos anteriores. Los vídeos de alpha 10 siguen siendo percibidos por los usuarios, mientras que algunos de los vídeos codificados para alpha 5 pasan desapercibido por algunos usuarios. Del mismo modo, los vídeos codificados para alpha 3 siguen siendo los que menos ven los usuarios. En el caso de la longitud de símbolo, sigue la misma metodología que en los vídeos anteriores, en la que para valores elevado se consigue una mayor perceptibilidad. Los datos tan pobres conseguidos a la hora de esconder el código en el vídeo pueden deberse a la poca presencia de azul en este.

Vídeos codificados	Orden de visualización	Porcentaje de error
flower_FPS14_A10	1	0.8415
flower_FPS27_A10	2	0.4424
flower_FPS7_A5	3	0.8222
flower_FPS27_A5	4	0.4424
flower_FPS27_A3	5	1
flower_FPS14_A5	6	0.6619
flower_FPS7_A10	7	0.3431
flower_FPS7_A3	8	0.5594
flower_original	9	1
flower_FPS14_A3	10	1

Tabla 8. Resultados obtenidos para los vídeos de "flower".

Finalmente, observando la tabla anterior, podemos observar que en los vídeos codificados de "flower" se obtienen los mejores resultados, llegando a engañar completamente a los espectadores. En este caso, se obtiene los mejores valores para los símbolos más largos. Esto se debe a que el vídeo codificado no tiene mucho movimiento, pero si un gran contraste, lo que ayuda al sistema a esconder de mejor forma los datos. De esta forma, los mejores valores para codificar este vídeo son valores de intensidad bajos y longitudes de código elevadas.

En definitiva, el estudio realizado ha cumplido con los objetivos y las expectativas puestas en él, ya que se esperaba obtener este tipo de resultados, ya que para valores de alpha elevados la mayoría de las veces será visible por el usuario, mientras que la elección del tamaño del código dependerá del movimiento que haya presente en el vídeo en cuestión.

7.5.2. Resultados de QoS

A la hora de determinar un valor de SIR umbral dependerá del tipo de vídeo usado para codificar, del valor de alpha y de la cantidad de frames por símbolos. De esta forma, para cada tipo vídeo y para cada combinación de los valores anteriores se deberá de realizar un estudio por cada vídeo estableciendo inicialmente una SIR para la cual todos los bits de código sean transmitidos en su totalidad (por ejemplo, -50 dB). Posteriormente, se realizaría una función de distribución acumulativa, la cual mostraría los valores de SIR en su totalidad. En el caso de uso de este proyecto, se han usado los valores de 10%, 50% y 90% para la realización de las pruebas. Esto significa que sólo se codifican el 10%, 50% y 90% de la totalidad de los bits de datos.

Las siguientes dos figuras muestran la función de distribución acumulativa de los diferentes valores de SIR con sus distintas combinaciones pertenecientes a un vídeo llamado 'sea' de duración 15 segundos a 50 FPS.



Figura 45. SIR obtenida para una longitud de símbolo 7 y valor de alpha 3.



Figura 46. SIR obtenida para una longitud de símbolo 27 y valor de alpha 10.

Se puede observar en dichas figuras que, a medida que aumenta el valor de alpha aumenta el valor de la SIR. A modo de ejemplo se puede observar la siguiente tabla, en la que quedan reflejados los valores de SIR para todas las combinaciones que se han hecho en este trabajo sobre el vídeo "sea.mp4":

		Valor de intensidad				
SIR (sea)	(alpha)					
		Baj	Baja (3)		Alta (10)	
		10%	-9.7958 dB	-5.3455 dB	0.66178 dB	
	Baja (7)	50%	-8.3432 dB	-3.8872 dB	2.1144 dB	
Longitud del código (framesPerSymbol)		90%	-6.9365 dB	-2.4777 dB	3.5211 dB	
	Media	10%	-12.5411 dB	-8.125 dB	-2.1044 dB	
	(14)	50%	-11.4917 dB	-7.0645 dB	-1.0439 dB	
		90%	-10.0723 dB	-5.6092 dB	0.481 dB	
	Alto	10%	-13.6265 dB	-9.1639 dB	-3.1689 dB	
	(27)	50%	-12.9613 dB	-8.5366 dB	-2.5037 dB	
		90%	-11.6055 dB	-7.0007 dB	-0.99915 dB	

Tabla 9. Valores de SIR obtenidos para el vídeo "sea".

Observando la tabla anterior, se puede afirmar que a mayor valor de *alpha*, se obtienen mejores valores de SIR para la codificación, y que a mayores valores de *framesPerSymbol* se obtienen peores valores de SIR, ya que los frames cambian más y esto crea una mayor interferencia.

Los valores de la tabla anterior se han obtenido gracias a las gráficas procedentes de la función de distribución acumulativa mostradas previamente. Bajo las restricciones de: 3 valores de alpha, 3 de tipos de longitudes de código y 3 porcentajes de codificación; se han creado 9 vídeos y gracias a la función de distribución acumulativa se han obtenido 27 valores de SIR. Por cada vídeo original se han codificado 27 vídeos, en función de las restricciones para las posteriores pruebas. Finalmente, este proceso se llevó a cabo con todos los 4 vídeos explicados con

anterioridad, ya que cada uno deberá de tener un valor de SIR determinado por las condiciones de alpha y de la longitud del código usadas.

Del mismo modo que en el caso del vídeo "sea", se muestran el resto de tablas obtenidas para los otros 3 vídeos restantes:

		Valor de intensidad							
SIR (flower	SIR (flower)			(alpha)					
		Baja (3)	Media (5)	Alta (10)					
		10%	-16.6176 dB	-12.0208 dB	-6.16 dB				
Longitud del código	Baja (7)	50%	-9.7641 dB	-5.2073 dB	0.69351 dB				
(framesPerSymbol)		90%	-6.3999 dB	-1.8536 dB	4.0577 dB				
		10%	-18.9811 dB	-14.2625 dB	-8.5235 dB				
	Media (14)	50%	-12.1986 dB	-7.6028 dB	-1.741 dB				
		90%	-9.2525 dB	-4.7035 dB	1.2051 dB				
		10%	-20.8635 dB	-16.4265 dB	-10.5872 dB				
	Alta (27)	50%	-14.0811 dB	-9.6442 dB	-3.7979 dB				
		90%	-11.3288 dB	-6.8918 dB	-0.98357 dB				

Tabla 10. Valores de SIR obtenidos para el vídeo "flower".

	Valor de intensidad								
SIR (birds	SIR (birds)			(alpha)					
	Baja (3)		Media (5)	Alta (10)					
		10%	-7.5725 dB	-3.1355 dB	2.8851 dB				
	Baja (7)	50%	-3.3654 dB	1.0716 dB	7.0922 dB				
		90%	90% 0.3668 dB	4.8038 dB	10.8244 dB				
	Media (14)	10%	-9.0803 dB	-4.6278 dB	1.3772 dB				
Longitud del código (framesPerSymbol)		50%	-5.4986 dB	-1.0332 dB	4.959 dB				
		90%	-2.9464 dB	1.5191 dB	4.8038 dB 10.8244 dB 4.8038 dB 10.8244 dB -4.6278 dB 1.3772 dB -1.0332 dB 4.959 dB -1.5191 dB 7.5111 dB -5.297 dB 0.70652 dB -1.8283 dB 4.1654 dB				
		10%	-9.7511 dB	-5.297 dB	0.70652 dB				
	Alta (27)	50%	-6.2921 dB	-1.8283 dB	4.1654 dB				
		90%	-3.7508 dB	0.7362 dB	6.7068 dB				

Tabla 11. Valores de SIR obtenidos para el vídeo "birds".

Tabla 12. Valores de SIR obtenidos para el vídeo "walk".

SIR (walk)	Valor de intensidad (alpha)				
		Baja (3)	Media (5)	Alta (10)	
		10%	-23.2521 dB	-18.8151 dB	-12.7945 dB
	Baja (7)	50%	-17.1782dB	-12.7412 dB	-6.7206 dB
		90%	8.6148 dB	13.0517 dB	19.0723 dB
		10%	-24.917 dB	-20.4918 dB	-14.4595 dB
	Media	50%	-19.0328 dB	-14.6191 dB	-9.1301 dB
---------------------	------------	-----	-------------	-------------	------------
Longitud del código	(14)				
(framesPerSymbol)		90%	4.438 dB	8.8705 dB	14.8956 dB
	A 140 (27)	10%	-9.7511 dB	-20.8067 dB	0.70652 dB
	Alta (27)	50%	-6.2921 dB	-17.2074 dB	4.1654 dB
		90%	-3.7508 dB	5.3066 dB	6.7068 dB

Respecto a los valores obtenidos, se puede observar que todos ellos siguen la misma metodología, ya que medida que al aumentar el valor de la longitud del código y de alpha, se obtienen peores datos, consiguiendo así una codificación más restrictiva. Del mismo modo, se puede observar que no todos los vídeos obtienen los mismos resultados de SIR para la misma codificación empleada. Por ejemplo, si observamos los SIR umbrales de los vídeos "birds" y "walk" tienen una diferencia en algunos casos de hasta 15 dB. Estas diferencias se deben a las propiedades del vídeo (movimiento y colorimetría). El vídeo de "birds" tiene muy poco movimiento si lo comparamos con "walk". Además, este último tiene colores frios, que vienen muy bien a la hora de la codificación, mientras que "birds" es un vídeo con colores neutros y cálidos. En definitiva, las diferencias producidas en laso 4 vídeos se deberá a la cantidad de azul que haya y al tipo de transiciones que se produzca (movimiento lento, medio y dinámico). Por otro lado, se puede observar que en cada vídeo existen las mismas diferencias de SIR umbrales entre distintos tipos de codificación. Por ejemplo, el en vídeo "sea" se puede observar como las diferencias de SIR umbrales están acotadas en torno a 2-3 dB de diferencia, mientras que en el caso de "flower" van de 9-10 dB. Finalmente, gracias a dichos valores de SIR se pudo codificar todas las combinaciones de los 4 vídeos usados. Una vez realizada la codificación, se realizó una decodificación para comprobar que el sistema funcionaba correctamente, cumpliendo los criterios de detección y perceptibilidad del código.

Por otro lado, los vídeos codificados presentan el siguiente aspecto (en función del valor de alpha utilizado):



a) Vídeo codificado con alpha 3

b) Vídeo codificado con alpha 10

Figura 47. Vídeo "flower" codificado con valores de alpha a) 3 y b) 10.



a) Vídeo codificado con alpha 3

b) Vídeo codificado con alpha 10

Figura 49. Vídeo "sea" codificado con valores de alpha a) 3 y b) 10.



a) Vídeo codificado con alpha 3

b) Vídeo codificado con alpha 10

Figura 48. Vídeo "flower" codificado con valores de alpha a) 3 y b) 10.



a) Vídeo codificado con alpha 3

b) Vídeo codificado con alpha 10

Figura 50. Vídeo "birds" codificado con valores de alpha a) 3 y b) 10.

Si tomamos los vídeos *frame* a *frame* no se notan notables diferencias, hay que fijarse muy bien para poder notar el código incrustado. Si se observan las Figura 50 b) y Figura 48 b) se puede visualizar en el fondo de la escena cambios de iluminación que no existen en la codificación a), ya que es casi imperceptible para el ojo humano. Aunque *frame* a *frame* no se noten diferencias entre las imágenes mostradas, si se visualiza el vídeo en conjunto si pueden existir parpadeos, especialmente en las dos últimas figuras mencionadas.

• Análisis de los datos obtenidos.

En la parte del decodificador se obtuvieron una serie de métricas las cuales, después de un postprocesado y comparación con los bits codificados enviados por el transmisor se obtuvieron una serie de gráficas. Para evaluar dichas gráficas se han utilizado 5 puntos de vista: el valor de la SIR, el valor de la intensidad alpha, la longitud del código, la cantidad de azul en el vídeo y el lo que ocurre cuando se utilizan vídeos con altos niveles de contraste. En primer lugar, evaluaremos como actúa la SIR en las siguientes gráficas:



Figura 51. Métricas obtenidas en función de la SIR del vídeo "birds" con alpha 3 y longitud de código 7.



Figura 52. Métricas obtenidas en función de la SIR del vídeo "sea" con alpha 3 y longitud de código 7.

La Figura 51 pertenece a un vídeo con colores neutros mientras que la Figura 52 pertenece a un vídeo con una gran presencia de azul. Ambos vídeos tienen un movimiento muy escaso, por lo que es normal que sus gráficas tengan tantas similitudes. Además, ambas gráficas han sido sacadas bajo las mismas condiciones (alpha 3 y longitud de código 7). Se puede observar a partir de ambas figuras que se realiza una correcta codificación, ya que se está produciendo un desplazamiento a la derecha. Este desplazamiento implica que el decodificador está detectando símbolos, ya que la curva de color rojo es mucho menor que la azul en todos los casos (símbolo versus no símbolo). Como se ha comentado a lo largo de este trabajo, la SIR es uno de los parámetros que más influye a la hora de detectar un código, ya que en función de su grado (alto, medio o bajo) se introducirán más o menos datos en el vídeo. Del mismo modo, se demuestra que, a cambio de una menor tasa de datos (SIR del 90%), aumenta las probabilidades de detección del código. Es decir, a mayores valores de porcentaje de codificación el sistema se vuelve más restrictivo, pero se consiguen métricas mejores.

Por otra parte, otro de los factores que determina la calidad de la comunicación implementada es el factor alpha. Este factor se evaluará en las siguientes gráficas:



Figura 53. Métricas obtenidas en función de alpha del vídeo "birds" con SIR al 10% y longitud de código 7.



Figura 54. Métricas obtenidas en función de alpha del vídeo "sea" con SIR al 10% y longitud de código 7.

Estás métricas han sido tomadas de los mismos vídeos que en la evaluación anterior, para un valor de SIR de 10% en el que se transmiten el 90% de los datos y una longitud de símbolo muy pequeña. Se puede observar que el valor de alpha no es un parámetro relevante en la calidad de codificación, ya que en ambos vídeos existe un desplazamiento de las curvas a la derecha y la curva azul se encuentra adelantada en todo momento. El valor de alpha cobra vital importancia en el apartado perceptual del código, ya que hasta para un valor de alpha 3, aunque el espectador no sea capaz de distinguirlo, la cámara si lo es. Por otro lado, el valor de alpha está directamente relacionado con la SIR, como se puede haber visto en las tablas anteriores. A mayor valor de alpha se obtendrá un valor de SIR con más calidad, ya que interviene en la decodificación haciéndola más sencilla. De cara a las métricas obtenidas no se nota especial diferencia entre ellas, con lo que este factor sólo es perceptible visualmente. La principal y única diferencia se puede observar en amplitud y esto puede deberse al modo en que usamos el canal alpha, metiendo sólo datos en la componente azul. Es por ello, que siempre deberán tener algún matiz de azules los vídeos a codificar. Al ser "sea" un vídeo con sólo azules, tendrá mucha más energía que "birds", ya que es un vídeo con una mezcla entre colores neutros y cálidos.





Figura 56. Métricas obtenidas en función de la longitud del símbolo del vídeo "birds" con SIR al 10% y alpha 3.



Figura 55. Métricas obtenidas en función de la longitud del símbolo del vídeo "birds" con SIR al 10% y alpha 3.

De las gráficas anteriores se puede observar que a medida que aumenta el valor de la longitud del código se obtienen peores valores en la codificación. Esta diferencia se debe a que no es lo mismo recuperar el dato durante 7 frames que 27. Durante este periodo de tiempo, aunque corto, podrían ocurrir diversos eventos como: una mala sincronización entre símbolos, desenfoque de la cámara, cambios de luminosidad en la escena, etc. Es por ello por lo que se recomienda utilizar símbolos cortos para realizar la codificación, incluso por debajo del mínimo utilizado en este trabajo. En definitiva, se puede observar que los valores obtenidos no son tan buenos como las métricas anteriores, pero aún así se siguen obtienendo los datos decodificados correctamente. Nuevamente, las diferencias entre las métricas dependen del tipo de vídeo a codificar.

Una vez comentados los parámetros de los que depende la codificación, pasamos a comentar cómo afecta el tipo de vídeo empleado en función de su movimiento. En el caso de la evaluación anterior, se utilizó para la comparación dos vídeos: el de "sea" y el de "birds". Hasta este punto no se había hecho hincapié en el tipo de movimiento, ya que ambos vídeos son imágenes estáticas con muy poco movimiento en ellos. Para esta nueva evaluación se utilizó los otros dos vídeos restantes: "walk" y "flower". Lo distintivo de ambos vídeos es que "walk" es una secuencia con mucho movimiento de personas caminando y "flower" es un vídeo con un altro contraste, ya que está grabado a contraluz, y de bajo movimiento. De este modo, se realiazaron las mismas evaluaciones que en el caso anterior.



Figura 57. Métricas obtenidas en función de la SIR del vídeo "walk" con alpha 3 y longitud de código 7.



Figura 58. Métricas obtenidas en función de la SIR del vídeo "flower" con alpha 3 y longitud de código 7.

Las gráficas anteriores fueron obtenidas a un valor de alpha 3 y una longitud de símbolo 7. De ellas se puede obtener que, en el caso del vídeo "flower", al existir tan poco movimiento y una gran cantidad de luz, ayuda a la hora de realizar a decodificación, adquiriendo los datos en su totalidad. Se sigue observando que, a cambio de una menor tasa de datos, aumenta las probabilidades de detección del código considerablemente. Por otra parte, en el caso del vídeo "walk" se puede observar que se obtienen los peores valores comparado con el resto de las gráficas, ya que, como se comentó anteriormente, es un vídeo con muchas transiciones en la que es difícil incrustar el código para relaizar una correcta decodificación. Así, se seguirán obteniendo valores, pero no con la misma calidad que en los anteriores casos (aparición de falsos positivos).



Figura 59. Métricas obtenidas en función de alpha del vídeo "walk" con SIR al 10% y longitud de código 7.



Figura 60. Métricas obtenidas en función de alpha del vídeo "flower" con SIR al 10% y longitud de código 7.

De la evaluación del factor alpha se sacan las mismas conclusiones que en los vídeos anteriores. Este parámetro mejora la SIR y hace que la captura de símbolos sea mucho más exacta con valores de alpha elevados. Sin embargo, un valor de alpha elevado puede ser perceptible en función del vídeo empleado en la codificación.



Figura 61. Métricas obtenidas en función de la longitud del símbolo del vídeo "walk" con SIR al 10% y alpha 3.

Se puede observar que las métricas obtenidas cumplen lo comentado anteriormente. Para símbolos cortos se realiza una buena captación de los datos, mientras que si vamos aumentando el tamaño sólo llegamos a captar un símbolo para una longitud de 27. Al tener este vídeo tanto movimiento es normal que con símbolos tan altos obtengamos resultados mediocres. Una recomendación ante vídeos de este tipo es acortar la duración del símbolo. Si los vídeos tienen movimiento elevado se deberá establecer una restricción, mientras que, para los vídeos con movimientos lentos o imágenes estáticas, se podrá ser más permisivo con esta premisa. En el caso del vídeo "flower" se puede observar que con altos valores de contraste en la imagen se pueden obtener los datos para longitudes de símbolos elevadas. Además, como en la imagen existe muy poco movimiento, se pueden decodificar más de un dato. Finalmente, en ambas gráficas las curvas están desplazadas a la derecha (en mayor o menor medida), con lo que se concluye que se ha realizado una buena métrica para la decodificación de los símbolos.



Figura 62. Métricas obtenidas en función de la longitud del símbolo del vídeo "walk" con SIR al 10% y alpha 3.

En definitiva, se ha comprobado que, en la mayoría de los casos, las mejores métricas se obtendrán para valores elevados de SIR (90 %), ya que en estos casos se limita la transmisión de datos por velar por la integridad del símbolo. En relación a la longitud del símbolo siempre será mejor utilizar símbolos cortos porque no podremos controlar el movimiento o las transiciones de los vídeos. Con longitudes cortas evitamos estos problemas. Por otro lado, el valor de intensidad dependerá de la forma en que la cámara capte la información, en este caso de estudio se ha comprobado que para un valor de alpha 3 ha sido suficiente en la mayoría de las situaciones. En definitiva, no existe una manera exacta a la hora de codificar los vídeos, dependerá en la mayoría de los casos de las características de estos. Se pueden seguir una serie de recomendaciones, pero siempre puede variar ya que no podemos controlar todas las características del vídeo en todo momento.

Los objetivos de este Trabajo Fin de Grado fueron: el estudio de los sistemas S2C, la evaluación de sistemas de codificación que permitieran utilizar una pantalla como fuente de datos sin afectar a la percepción visual y el diseño y evaluación de un sistema de test. Todos estos objetivos fueron llevados a cabo con la implementación de un algoritmo en MATLAB. El sistema propuesto permite experiencias de visualización de vídeos para los usuarios mediante la multiplexación de fotogramas de vídeos y de datos a la vez que se presentan por pantalla. El diseño implementado fue mucho más ambicioso que los objetivos que se habían presentado para este trabajo, ya que en un principio sólo se tenía pensado realizar la codificación en el lado de transmisor, sin tener en cuenta la forma y el modo que llegarían los datos al receptor. De este modo, se creó un sistema receptor en el que se crearon una serie de métricas, que ayudaron a determinar la calidad del enlace entre transmisor y receptor, es decir, que los datos llegaran de forma correcta.

El diseño propuesto aprovecha la propiedad de fusión de parpadeo del sistema de visión humana y la capacidad de las pantallas modernas para mostrar información. El vídeo y los datos transportados por los *frames* operan en diferentes escalas de tiempo. El contenido del vídeo se percibe a un ritmo lento debido a los límites físicos del ojo humano, mientras que los cuadros de datos se muestran a una velocidad rápida, que sólo puede ser capturada y decodificado por la cámara. En definitiva, se optó por utilizar técnicas de espectro ensanchado para reducir las interferencias producidas por el canal y un esquema diferencial usando matrices de Hadamard, ya que tienen energía nula y poseen la propiedad de ortogonalidad. En definitiva, se ha desarrollado un sistema de codificación diferencial S2C que emplea códigos espaciales ortogonales basados en secuencias de Hadamard y un algoritmo de decisión de inserción de información basado en la predicción de la SIR en el receptor.

Se ha realizado un estudio a 90 participantes sobre la perceptibilidad del código en vídeos codificados, donde se han cumplido con los objetivos propuestos de invisibilidad del código, ya que la mayoría de los participantes no eran capaces de distinguir a ciencia cierta el código en la mayoría de los casos. En relación a los factores que determinan el sistema propuesto podemos encontrar 3: la SIR, el factor alpha y la longitud del símbolo. Se ha comprobado que, en la mayoría de los casos, las mejores métricas se obtienen para valores elevados de SIR, ya que en estos casos se limita la transmisión de datos por velar por la integridad del símbolo. En relación a la longitud del símbolo siempre será mejor utilizar símbolos cortos porque no podremos controlar el movimiento o las transiciones de los vídeos. Por otro lado, el valor de intensidad dependerá de la forma en que el sistema de captura obtenga la información. En la mayoría de los casos de prueba

un valor de alpha 3 ha sido suficiente. En definitiva, no existe una manera exacta de codificar los vídeos, ya que dependerá en la mayoría de los casos de las características de estos. Finalmente, se ha demostrado que nuestro sistema mantiene la calidad de visión normal para los usuarios, mientras que permite una comunicación visual de alta velocidad para los dispositivos.

La mayoría de las limitaciones de este proyecto se han encontrado en el lado de la recepción, ya que este proceso depende de diversos factores como es la posición de la cámara, la compensación del movimiento, la distancia entre transmisor y el receptor, etc. Se ha usado un algoritmo sencillo de estimación global de SIR a nivel de *frame*, ya que la SIR a nivel local puede tener grandes variaciones. Futuras versiones del sistema de decisión de inserción de información deberán tener en cuenta esta característica para mejorar la QoS. Además, en este trabajo se ha probado únicamente un tamaño de código, aunque el código fuente del software desarrollado permite utilizar cualquier profundidad de código. Por otro lado, se ha utilizado toda la pantalla para transmitir un símbolo, lo cual no explota en su totalidad la capacidad del canal de comunicaciones. Se barajó otro esquema de codificación, pero se desestimó por cuestiones de tiempo, ya que consistía en la multiplexación espacial de varios símbolos como los desarrollados en este trabajo en toda la pantalla, empleando un esquema de subgrids.

En este capítulo se estima el presupuesto del proyecto en base a las recomendaciones y pautas establecidas por el Colegio Oficial de Ingenieros de Telecomunicación (COITT) establecidas en el 2008. A su vez, dicho presupuesto se desglosa en los siguientes apartados:

9.1. Recursos Materiales

Entre los recursos materiales utilizados para ejecución y desarrollo de este trabajo final de grado se incluyeron tanto los recursos hardware como los recursos software, que a su vez pueden tener costes asociados al requerimiento de licencias para su uso. Para estimar el coste de amortización, se estipula un periodo de 4 años, presuponiendo un sistema de amortización lineal, en el que se desprecia el material inmovilizado de forma constante durante el periodo de evaluación. Para realizar dicho cálculo, se hace uso de la siguiente ecuación:

$$Coste \ de \ amortización = \frac{Valor \ de \ adquisición - Valor \ residual}{Años \ de \ vida \ útil}$$
(21)

Como el presente proyecto tiene una duración de 300 horas distribuidas de manera aproximada en 4 meses, y este periodo es inferior a los 4 años del coste estipulado de amortización, dicho coste será el derivado de los 4 meses en los que se desarrolla el proyecto.

Entre los recursos hardware empleados para el desarrollo de este proyecto se encuentran:

- o Ordenador portátil Macbook Air 13" 2015.
- o Trípode.
- o Móvil Huawei Mate 10.

Recurso	Valor de adquisición (€)	Valor residual (€)	Coste de amortización (€)
Ordenador portátil	990	390	50
Móvil	595	189	33,83
Trípode	4,95	2,32	0,22
	Total:		84,05

Tabla 13. Recursos hardware.

El coste de amortización asociado a los recursos hardware asciende a ochenta y cuatro euros y cinco céntimos (84,05 €).

En lo que respecta a los recursos software, entre las herramientas utilizadas se encuentras: MATLAB R2019b y Microsoft Office 2019. En este caso, las herramientas utilizadas disponen de licencias para su uso, pero al pertenecer a la comunidad universitaria de la Universidad de las Palmas de Gran Canaria no conllevan costes de amortización asociados, ya que se han otorgado de manera gratuita.

9.2. Trabajo tarifado por tiempo empleado

El proyecto se ha desarrollado en un tiempo de 4 meses, en los que se incluyen tareas de formación, especificación, desarrollo y documentación necesarias para la realización de este. Para hacer una estimación de los honorarios que debe recibir el ingeniero por el trabajo realizado se sigue las recomendaciones del COITT, empleando la siguiente expresión:

$$H = C_t \cdot 74,88 \cdot H_n + C_t \cdot 96,72 \cdot H_e \tag{22}$$

Donde: H son los honorarios totales por el tiempo dedicado, Ct es un factor de corrección en función del número de horas trabajadas y He son las horas trabajadas (dentro de la jornada laboral). Según el COITT, el coeficiente mencionado con anterioridad, Ct, tiene un valor variable en función del número de horas empleadas de acuerdo con la siguiente tabla:

Horas empleadas	Factor de corrección Ct
Hasta 36 horas	1,00
De 36 a 72 horas	0,90
De 72 a 108 horas	0,80
De 108 a 144 horas	0,70
De 144 a 180 horas	0,65
De 180 a 360 horas	0,60
De 360 a 540 horas	0,55
De 540 a 720 horas	0,50
De 720 a 1080 horas	0,45
Más de 1080 horas	0,40

Tabla 14. Factor de corrección en función del número de horas invertidas.

En base a la tabla anterior, observamos que en función de la duración del trabajo final de grado deberíamos de estar en la región de 180 a 360 horas con un factor de 0,60. Al aplicar el factor de corrección y las horas empleadas en la expresión anterior, esta quedarías de la siguiente forma:

$$H = 0.6 * 74.88 * 300 + 0.6 * 96.72 * 0 = 13.478.40 \in (16)$$

Finalmente, los honorarios totales por tiempo dedicado libres de impuestos, asciende a la cantidad de trece mil cuatrocientos setenta y ocho euros con cuarenta céntimos (13.478,40 €).

El importe de la redacción del proyecto se calcula de acuerdo con la siguiente expresión:

$$R = 0.07 * P * C_n \tag{17}$$

Donde, P es el presupuesto del proyecto y Cn es el coeficiente de ponderación en función del presupuesto. El coeficiente Cn está determinado por el presupuesto del proyecto en el que, hasta el momento, incluye los siguientes recursos:

Tabla 15. Presupuesto de ejecución material.

Recurso	Coste (€)
Recursos hardware	84,05
Recursos software	0
Trabajo tarifado por tiempo empleado	13.478,40
Total	13.562,45

El coeficiente de ponderación para presupuestos menores de 30050,00 € es de 1,00, por lo que el coste derivado de la redacción del presente informe es de:

$$R = 0,07 * 13.562,45 * 1,00 = 949,37 \in$$
(18)

Por lo tanto, el coste libre de impuestos derivados de la redacción del proyecto asciende a los novecientos cuarenta y nueve con treinta y siete céntimos (949,37 €).

9.3. Derechos del visado del COITT

Los gastos de visado del COITT se tarifan mediante la siguiente expresión:

$$V = 0,006 * P * C_{\nu} \tag{19}$$

Donde, P es el presupuesto del trabajo y Cv es el coeficiente reductor en función del presupuesto del trabajo. De esta forma, el presupuesto calculado hasta el momento asciende hasta la suma de los costes de ejecución material y de redacción, al no existir costes de material consumible, por no ser necesaria la impresión del documento para su evaluación.

$$P = 13.562,45 + 949,37 = 14.511,82 \in$$
 (20)

El coeficiente de ponderación para presupuestos menores de $30.050,00 \in$ definido por el COITT tiene un valor de 1,00, por lo que, el coste de los derechos de visado del trabajo asciende finalmente a:

$$V = 0,006 * 14.511,82 * 1,00 = 87,07 \in$$
(21)

Por lo tanto, el coste de los derechos de visado del proyecto suma ochenta y siete euros con siete céntimos (87,07 \in).

9.4. Gastos de tramitación y envío

Los gastos de tramitación y envío están estipulados en seis euros y un céntimo (6.01 €).

9.5. Aplicación de impuestos y presupuesto final

El coste total del proyecto con la aplicación de los correspondientes impuestos asciende a la cuantía de 14.604,9 €. A esta cantidad hay que sumarle el Impuesto General Indirecto Canario, que es de un 7%.

Recurso	Coste (€)
Recursos materiales	84,05
Trabajo tarifado por tiempo empleado	13.478,40
Costes asociados a la redacción del documento	949,37
Derechos de visado del COITT	87,07

Gastos de tramitación y envío	6,01
Subtotal	14.604,9
Aplicación de impuestos (IGIC 7%)	1022,34
Total	15.627,24

El presupuesto tota de este Trabajo de Fin de Grado asciende a la cuantía de quince mil seiscientos veintisiete euros con veinticuatro céntimos (15.627,24 €).

Las Palmas de Gran Canaria a 1 de julio de 2020

Firma:

Alberto Ramos Monagas

- 1. Cisco Corporation, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018, 2014.
- Parth H. Pathak, Xiaotao Feng, Pengfei Hu, y Prasant Mohapatra. (2015). "Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges". IEEE Communications surveys & Tutorials, Vol. 17, No. 4, Fourth Quarte, 2047-2077.
- The Institute of Electrical and Electronics Engineers, Inc. (4 de marzo de 2009). IEEE 802.15 WPANTM Task Group 7 (TG7) Visible Light Communication. Recurso digital: http://www.ieee802.org/15/pub/TG7.html.
- Md. Shahjalal, Moh. Khalid Hasan, Mostada Zaman Chowdhury, y Yeong Min Jang, (2019).
 "Future Optical Camera Communication Based Applications and Opportunities for 5G and Beyond". Department of Electronics Engineering, Kookmin University, Scoul, Korea. IEEE ICAIIC 2019.
- 5. 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).
- 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.
- Tianxing Li, Chuankai An, Xinran Xiao, Andrew T. Campbell, and Xia Zhou. 2015. Real-Time Screen-Camera Communication Behind Any Scene. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (MobiSys '15). ACM, New York, NY, USA, 197-211.
- 8. Porter, T., and Duff, T. Compositing digital images. In ACM SIGGRAPH *Computer Graphics* (1948).
- M. Yuen, H. Wu, A survey of hybrid MC/DPCM/DCT video coding distortions, Signal Process. 70 (3) (1998) 247–278.

- Mannos J., Sakrison D.: The effects of a visual fidelity criterion of the encoding of images. IEEE Transactions on Information Theory 20, 4 (1974), 525-536.
- Kandel RE, Schwarts J, Jessell TM. Principles of neural science. 3 ed. New York: Elsevier Science Publishing; 1991. Chapter 28. Phototransduction and information processing in the retina.
- 12. Lamme AFV, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci.
- 13. Super H. Figure-ground activity in V1 and guidance of saccadic eye movements. J. Physiol.
- 14. Super H. Working memory in the primary visual cortex. Arch. Neurol.
- 15. Perez D, Travieso D, Magano A, Lopez A, Marquez MT. Precepción visual. Tema 1.
- 16. Glassner A.: Principles of digital image synthesis. Morgan Kaufmann Pub, 1995.
- 17. Webb P.: Bioastronautics data book. Scientific and Technical Information Division, National Aeronautics and Space Administration, 1964.
- 18. D. A. Atchison, G. Smith, and G. Smith. *Optics of the human eye*. Butterworth-Heinemann Oxford, UK: 2000.
- 19. R. D. Valois and K. D. Valois. Spatial Vision. Oxford University Press, 1988.
- 20. D. Kelly. Flicker. In Visual psychophysics, pages 273–302, 1972.
- 21. E. Simonson and J. Brožek. Flicker fusion frequency: background and applications. Physiological reviews, 1952.
- 22. G. Brindley, J. Du Croz, and W. Rushton. The flicker fusion frequency of the blue-sensitive mechanism of colour vision. The Journal of physiology, 183(2):497–500, 1966.
- 23. D. G. Green. Sinusoidal flicker characteristics of the color-sensitive mechanisms of the eye. *Vision research*, 9(5):591–601, 1969.

- 24. W. A. Hershberger and J. S. Jordan. The phantom array: a perisaccadic illusion of visual direction. The Psychological Record, 48(1):2, 2012.
- 25. J. Roberts and A. Wilkins. Flicker can be perceived during saccades at frequencies in excess of 1 khz. *Lighting Research and Technology*, 45(1):124–132, 2013.
- 26. I. Vogels and I. Hernando. Effect of eye movements on perception of temporally modulated light. Recurso digital: <<u>http://2012.experiencinglight.nl/doc/28.pdf</u>>.
- 27. Michelson A.: A 1927 Studies in Optics. Scientific American (1927).
- 28. Li R., Polat U., Makous W., Bavelier D.: Enhancing the contrast sensitivity function through action video game training. Nature Neuroscience 12,5 (2009), 549-551.
- 29. Q. P. Wenjun Hu, Hao Gu. Lightsync: Unsynchronized visual communication over screencamera links. In *MobiCom*, 2013.
- 30. Campbell FW, Green DG. Optical and retinal factors affecting visual resolution. J Physiol.
- 31. Artal P, Navarro R. Monochromatic modulation transfer function of the human eye for different pupil diameters: an analytical expression. J Opt Soc Am A.
- 32. Navarro R, Artal P, Williams DR. Modulation transfer of the human eye as a function of retinal eccentricity. J Opt Soc A.
- 33. Williams DR, Artal P, McMahon MJ, Navarro R, Brainard DH. Off- axis optical quality and retinal sampling in the human eye. Vision Res.
- 34. Wang YZ, Thibos LN, Bradley A. Effects of refractive error on detection acuity and resolution acuity in peripheral vision. Invest Ophthalmol Vis Sci.
- 35. Murphy B.: Pattern thresholds for moving and stationary gratings during smooth eye movement. Vision Research 18,5 (1978), 521 530.
- 36. Tyler C.: Analysis of visual modulation sensitivity. II Peripheral retina and the role of photoreceptor dimensions. Journal of the Optical Society of America A 2,3 (1985), 393 398.

- 37. Nakayama K.: Properties of early motion processing: Implications for the sensing of egomotion. The Perception and Control of Self Motion (1990), 69 80.
- Dally S.: Engineering observations from spatiovelocity and spatiotemporal visual models.
 Vision Models and Applications to Image and Video Processing (1998), 179 200.
- 39. Reddy M.: Perceptually modulated level of detail for virtual environments. University of Edinburgh (1997).
- 40. HADDAD HERNÁNDEZ, Juan Carlos. Pantallas de televisión: Funcionamiento, características y diferencias. Sistemas y difución de Televisión. Escuelo de Inegeniería de Telecomunicación y electrónica.
- H. J. Jang, J. Y. Lee, J. Kwak, D. Lee, J.-H. Park, B. Lee, and Y. Y. Noh, "Progress of display performances: AR, VR, QLED, OLED, and TFT", *Journal of Information Display*, vol. 20, no. 1. Journal of Information Display, pp. 1–8, 2019.
- 42. L. He, M. Fei, J. Chen, Y. Tian, Y. Jiang, Y. Huang, K. Xu, J. Hu, Z. Zhao, Q. Zhang, H. Ni, and L. Chen, "Graphitic C3N4 quantum dots for next-generation QLED displays", *Materials Today*, vol. 22. Materials Today, pp. 76–84, 2019.
- 43. G. J. Supran, Y. Shirasaki, K. W. Song, J.-M. Caruge, P. T. Kazlas, S. Coe-Sullivan, T. L. Andrew, M. G. Bawendi, and V. Bulović, "QLEDs for displays and solid-state lighting", *MRS Bulletin*, vol. 38, no. 9. MRS Bulletin, pp. 703–711, 2013.
- 44. Tilano TV. ¿Conoces los barridos de línea en HD? Recurso digital: https://tilanotv.es/sincategoria/conoces-los-barridos-linea-hd/.
- 45. PINEDA, Jorge. Vídeo Progresivo y Entrelazado. Intercoud. Recurso digital: https://www.intercoud.com/blog/video-progresivo-y-entrelazado/.
- Grois, D., Marpe, D., Mulayoff, A., Itzhaky, B. & Hadar, O. (2013). Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders.
- 47. Layek, M. A. et al. (2017). Performance analysis of H.264, H.265, VP9 and AV1 video encoders.

- 48. Lin, T., Jiang, H., Huang, J. et al. Fast intra coding unit partition decision in H.266/FVC based on spatial features. J Real-Time Image Proc 17, 493–510 (2020).
- 49. CASADO MIRAZ, Fátima María. Cámaras profesionales de Televisión, Producción de Televisión. Dpto. de Señales y Comunicaciones. Universidad de Las Palmas de Gran Canaria.
- 50. JULIÁN, Pedro. Circuitos Integrados Digitales CMOS-Análisis y Diseño. Alfaomega Grupo Editor, 2016.
- 51. MENDÍAS CUADROS, José Manuel. Diseño Físico de Circuitos CMOS. Dpto. Arquitectura de Computadores y Automática. Universidad Complutense de Madrid.
- 52. https://www.xatakafoto.com/camaras/sensores-con-tecnologia-ccd-vs-cmos
- 53. INGELS, Mark; STEYAERT, Michiel. *Integrated CMOS circuits for optical communications*. Springer Science & Business Media, 2013.
- 54. GONZÁLEZ JIMÉNEZ, José Luis. Ténicas de reducción del ruido de conmutación en circuitos integrados CMOS. Universitat Politècnica de Catalunya.
- 55. WEBER, Klaus. CMOS: Listos para el Broadcast de hoy. Manager de Marketin de productos de cámara. Febrebro 2012.
- 56. R. G. Cárdenas, "Esteganografía lingüística". 2007.
- 57. K.M.Sullivan, "Image steganalysis: hunting & escaping". 2005. Recurso digital: http://vision.ece.ucsb.edu/publications/05ThesisSullivan.pdf.
- 58. F. Diatel, "Esteganografía y marcas de agua". 2002.
- 59. D. Master, "Introducción a la esteganografía". 2004.
- L.K. Choi, L.K. Cormack, A.C. Bovik, Eccentricity effect of motion silencing on naturalistic videos, in: Proc. IEEE 3rd Global Conf. Sig. and Inf. Process., GlobalSIP, 2015, pp. 1190 1194.
- 61. R. Blake, R. Sekuler, Perception, fifth ed., McGraw-Hill, New York, NY, USA, 2006.

- M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant, N.C. Rust, Do we know what the early visual system does?, J. Neurosci. 25 (46) (2005) 10577–10597.
- 63. E.P. Simoncelli, D.J. Heeger, A model of neuronal responses in visual area MT, Vis. Res. 38 (5) (1998) 743–761.
- L. M. Marvel & Charles T. Retter, "A methodology for data hiding using images", IEEE 1998. Recurso digital: http://www.argreenhouse.com/society/TacCom/papers98/30_07i.pdf.
- 65. Cisco Visual Networking Index: Forecast and Methodology, 2013–2018. Recurso digital: http://tinyurl.com/mev32z8.
- 66. A. Wang, S. Ma, C. Hu, J. Huai, C. Peng, and G. Shen. Enhancing reliability to boost the throughput over screen-camera links. In *MobiCom*, 2014.
- 67. Ericsson, Ericsson mobility report, Tech. rep., Ericsson (2018).
- 68. S. Wu, H. Wang, C. H. Youn, Visible light communications for 5Gwireless networking systems: From fixed to mobile communications, IEEE Netw. 28 (6) (2014) 41–45.
- P. H. Pathak, X. Feng, P. Hu, P. Mohapatra, Visible light communication, networking, and sensing: A survey, potential and challenges, IEEE Commun. Surveys Tuts. 17 (4) (2015) 2047–2077.
- Z. Ghassemlooy, P. Luo, S. Zvanovec, Optical camera communications, in: Optical Wireless Communications, Springer, 2016, pp. 547–568.
- 71. M. Z. Chowdhury, M. T. Hossan, A. Islam, Y. M. Jang, A comparative survey of optical wireless technologies: Architectures and applications, IEEE Access 6 (2018) 9819–9840.
- J. Hao, Y. Yang, J. Luo, CeilingCast: Energy efficient and location-bound broadcast through LED-camera communication, in: Proc. Info.Commun. Conf., (INFOCOM), IEEE, 2016, pp. 1–9.
- R. Boubezari, H. Le Minh, Z. Ghassemlooy, A. Bouridane, Smartphonecamera based visible light communication, J. Lightw. Technol. 34 (17) (2016) 4121–4127.

- 74. T. Nguyen, A. Islam, T. Yamazato, Y. M. Jang, Technical issues onIEEE 802.15. 7m image sensor communication standardization, IEEECommun. Mag. 56 (2) (2018) 213–218.
- A. Ashok, M. Gruteser, N. Mandayam, J. Silva, M. Varga, K. Dana, Challenge: Mobile optical networks through visual MIMO, in: Proc. ofMobiCom/MobiHoc, ACM, 2010, pp. 105–112.
- E. Wengrowski, W. Yuan, K. J. Dana, A. Ashok, M. Gruteser, N. Man-dayam, Optimal radiometric calibration for camera-display communi-cation, in: IEEE Winter Conf. on App. of Computer Vision, 2016, pp.1–10.
- 77. N.-T. Le, Y. M. Jang, MIMO architecture for optical camera commu-nications, The J. of the Korean Institute of Commun. Sci. 42 (1) (2017)8–13.
- 78. A. Goldsmith, Wireless Communications. Cambridge, 2005.
- 79. T.S. Lomheim, G.C. Holst, CMOS/CCD Sensors and Camera Systems, second ed., The International Society for Optical Engineering (SPIE), 2011.
- A.Tang, J.Kahn, K.-P.Ho, Wireless infrared communication links using multi-beam transmitters and imaging receivers, in:ICC96, Conference Record, Converging Technologies for Tomorrow's Applications, vol. 1, June 1996, pp. 180–186.
- 81. G. Woo, A. Lippman, and R. Raskar, "Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter," in Proc. IEEE Int. Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2012, pp. 59–64.
- 82. T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou, "Real-time screen- camera communication behind any scene," in Proc. of the 13th Annual Int. Conf. on Mobile Systems, Applications, and Services (MobiSys). ACM, 2015, pp. 197–211.
- 83. V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, E. Wen- growski, and N. Mandayam, "High-rate flicker-free screen-camera com- munication with spatially adaptive embedding," in Proc. IEEE Int. Conf. on Computer Communications (INFOCOM). IEEE, 2016.

- 84. M. Izz, Z. Li, H. Liu, Y. Chen, and F. Li, "Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel," in Proc. IEEE Int. Conf. on Computer Communications (INFOCOM). IEEE, 2016.
- 85. HAO, T., ZHOU, R., AND XING, G. COBRA: Color barcode streaming for smartphone systems. In Proc. of MobiSys (2012).
- PERLI, S. D., AHMED, N., AND KATABI, D. PixNet: Interference-free wireless links using LCD-camera pairs. In Proc. of MobiCom (2010).
- 87. HU, W., GU, H., AND PU, Q. LightSync: Unsynchronized visual communication over screen-camera links. In Proc. of MobiCom (2013).
- 88. HU, W., MAO, J., HUANG, Z., XUE, Y., SHE, J., BIAN, K., AND SHEN, G. Strata: Layered coding for scalable visual communication. In Proc. of MobiCom (2014).
- RAJAGOPAL, N., LAZIK, P., AND ROWE, A. Visual light landmarks for mobile devices. In Proc. of IPSN (2014).
- 90. WANG, A., MA, S., HU, C., HUAI, J., PENG, C., AND SHEN, G. Enhancing reliability to boost the throughput over screen-camera links. In Proc. of MobiCom (2014).
- 91. WANG, A., PENG, C., ZHANG, O., SHEN, G., AND ZENG, B. InFrame: Multiflexing fullframe visible communication channel for humans and devices. In *Proc. of HotNets* (2014).
- 92. WOO, G., LIPPMAN, A., AND RASKAR, R. VRCodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In Proc. of ISMAR (2012).
- 93. YUAN, W., DANA, K., VARGA, M., ASHOK, A., GRUTESER, M., AND MANDAYAM, N. Computer vision methods for visual mimo optical systems. *Proceedings of the IEEE International Workshop on Projector-Camera Systems (held with CVPR)* (2011), 37–43.
- 94. YUAN, W., ET AL. Dynamic and invisible messaging for visual mimo. In *IEEE Workshop* on Applications of Computer Vision (WACV) (2012).
- 95. S.D. Perli, N. Ahmed, D. Katabi, PixNet: Interference-free wireless links using LCD-camera pairs, in: Proceedings of MobiCom'10, ACM, New York, NY, USA, 2010, pp. 137–148.

- 96. S. Hranilovic, F. Kschischang, A pixelated-MIMO wireless optical communication system, IEEE J. Sel. Top. Quantum Electron. 12 (4) (2006) 859–874.
- 97. T. Hao, R. Zhou, G. Xing, COBRA: color barcode streaming for smartphone systems, in: Proceedings of MobiSys'12, ACM, New York, NY, USA, 2012, pp. 85–98.
- 98. B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, USA, 1986.
- 99. PORTER, T., AND DUFF, T. Compositing digital images. In ACM Siggraph Computer Graphics (1984).
- 100. TAN, K. W., ET AL. FOCUS: a usable & effective approach to OLED display power management. In UbiComp (2013).
- 101. Li, T., An, C., Campbell, A., Zhou, X., 2014. HiLight.
- 102. Li Ping, W. K. Leung, and K. Y. Wu, "Low-rate turbo-Hadamard codes", *IEEE Transactions on Information Theory*, vol. 49, no. 12. IEEE Transactions on Information Theory, pp. 3213–3224, 2003
- P. Benabes, A. Gauthier, and J. Oksman, "A Manchester code generator running at 1 GHz".
- Sánchez Ponz, Jorge Luis. CDMA: comunicaciones de espectro ensanchado. "Buran", 2000, núm. 16, p. 25-32.
- 105. Ying Li, Hoang Trinh, N. Haas, C. Otto, and S. Pankanti, "Rail Component Detection, Optimization, and Assessment for Automatic Rail Track Inspection", *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2. IEEE Transactions on Intelligent Transportation Systems, pp. 760–770, 2014.
- 106. Lin, H.; Si, J.; Abousleman, G.P. Region-of-interest detection and its application to image segmentation and compression. In Proceedings of the 2007 International Conference on Integration of Knowledge Intensive Multi-Agent Systems, Waltham, MA, USA, 30 April–3 May 2007.





Figura 63. Métrica con longitud del símbolo 7, alpha 3 y SIR al 10%.



Figura 64. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%.



Figura 65. Métrica con longitud del símbolo 7, alpha 3 y SIR al 90%.



Figura 66. Métrica con longitud del símbolo 7, alpha 5 y SIR al 10%.



Figura 67. Métrica con longitud del símbolo 7, alpha 5 y SIR al 50%.



Figura 68. Métrica con longitud del símbolo 7, alpha 5 y SIR al 90%.



Figura 70. Métrica con longitud del símbolo 7, alpha 10 y SIR al 10%.



Figura 69. Métrica con longitud del símbolo 7, alpha 10 y SIR al 50%.



Figura 71. Métrica con longitud del símbolo 7, alpha 10 y SIR al 90%.



Figura 72. Métrica con longitud del símbolo 14, alpha 3 y SIR al 10%.



Figura 74. Métrica con longitud del símbolo 14, alpha 3 y SIR al 50%.



Figura 73. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%.



Figura 76. Métrica con longitud del símbolo 14, alpha 5 y SIR al 10%.



Figura 75. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%.


Figura 77. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%.



Figura 78. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%.



Figura 80. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%.



Figura 79. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%.





Figura 82. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%.



Figura 83. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%.



Figura 84. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%.



Figura 86. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%.



Figura 85. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%.



Figura 88. Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%.



Figura 87. Métrica con longitud del símbolo 27, alpha 10 y SIR al 50%.



Figura 89. Métrica con longitud del símbolo 27, alpha 10 y SIR al 90%.

A.2. Métricas obtenidas del vídeo 2 ("flower.mp4").



Figura 90. Métrica con longitud del símbolo 7, alpha 3 y SIR al 10%.



Figura 91. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%.



Figura 92. Métrica con longitud del símbolo 7, alpha 3 y SIR al 90%.



Figura 93. Métrica con longitud del símbolo 7, alpha 5 y SIR al 10%.



Figura 94. Métrica con longitud del símbolo 7, alpha 5 y SIR al 50%.



Figura 96. Métrica con longitud del símbolo 7, alpha 5 y SIR al 90%.



Figura 95. Métrica con longitud del símbolo 7, alpha 10 y SIR al 10%.



Figura 98. Métrica con longitud del símbolo 7, alpha 10 y SIR al 50%.



Figura 97. Métrica con longitud del símbolo 7, alpha 10 y SIR al 90%.



Figura 100. Métrica con longitud del símbolo 14, alpha 3 y SIR al 10%.



Figura 99. Métrica con longitud del símbolo 14, alpha 3 y SIR al 50%.



Figura 102. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%.



Figura 101. Métrica con longitud del símbolo 14, alpha 5 y SIR al 10%.



Figura 104. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%.



Figura 103. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%.



Figura 106. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%.



Figura 105. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%.



Figura 108. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%.



Figura 107. Métrica con longitud del símbolo 27, alpha 3 y SIR al 10%.



Figura 110. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%.



Figura 109. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%.



Figura 112. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%.



Figura 111. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%.



Figura 114. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%.



Figura 113. Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%.



Figura 116. Métrica con longitud del símbolo 27, alpha 10 y SIR al 50%.



Figura 115. Métrica con longitud del símbolo 27, alpha 10 y SIR al 90%.

A.3. Métricas obtenidas del vídeo 3 ("walk.mp4").



Figura 118. Métrica con longitud del símbolo 7, alpha 3 y SIR al 10%.



Figura 117. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%.



Figura 120. Métrica con longitud del símbolo 7, alpha 3 y SIR al 90%.



Figura 119. Métrica con longitud del símbolo 7, alpha 5 y SIR al 10%.



Figura 122. Métrica con longitud del símbolo 7, alpha 5 y SIR al 50%.



Figura 121. Métrica con longitud del símbolo 7, alpha 5 y SIR al 90%.



Figura 124. Métrica con longitud del símbolo 7, alpha 10 y SIR al 10%.



Figura 123. Métrica con longitud del símbolo 7, alpha 10 y SIR al 50%.



Figura 126. Métrica con longitud del símbolo 7, alpha 10 y SIR al 90%.



Figura 125. Métrica con longitud del símbolo 14, alpha 3 y SIR al 10%.



Figura 128. Métrica con longitud del símbolo 14, alpha 3 y SIR al 50%.



Figura 127. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%.



Figura 130. Métrica con longitud del símbolo 14, alpha 5 y SIR al 10%.



Figura 129. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%.



Figura 132. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%.



Figura 131. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%.



Figura 134. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%.



Figura 133. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%.



Figura 136. Métrica con longitud del símbolo 27, alpha 3 y SIR al 10%.



Figura 135. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%.



Figura 138. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%.



Figura 137. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%.



Figura 140. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%.



Figura 139. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%.



Figura 142. Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%.



Figura 141. Métrica con longitud del símbolo 27, alpha 10 y SIR al 50%.



Figura 143. Métrica con longitud del símbolo 27, alpha 10 y SIR al 90%.



A.4. Métricas obtenidas del vídeo 4 ("birds.mp4").





Figura 146. Métrica con longitud del símbolo 7, alpha 3 y SIR al 50%.



Figura 145. Métrica con longitud del símbolo 7, alpha 3 y SIR al 90%.


Figura 148. Métrica con longitud del símbolo 7, alpha 5 y SIR al 10%.



Figura 147. Métrica con longitud del símbolo 7, alpha 5 y SIR al 50%.



Figura 150. Métrica con longitud del símbolo 7, alpha 5 y SIR al 90%.



Figura 149. Métrica con longitud del símbolo 7, alpha 10 y SIR al 10%.



Figura 152. Métrica con longitud del símbolo 7, alpha 10 y SIR al 50%.



Figura 151. Métrica con longitud del símbolo 7, alpha 10 y SIR al 90%.



Figura 154. Métrica con longitud del símbolo 14, alpha 3 y SIR al 10%.



Figura 153. Métrica con longitud del símbolo 14, alpha 3 y SIR al 50%.



Figura 156. Métrica con longitud del símbolo 14, alpha 3 y SIR al 90%.



Figura 155. Métrica con longitud del símbolo 14, alpha 5 y SIR al 10%.



Figura 158. Métrica con longitud del símbolo 14, alpha 5 y SIR al 50%.



Figura 157. Métrica con longitud del símbolo 14, alpha 5 y SIR al 90%.



Figura 160. Métrica con longitud del símbolo 14, alpha 10 y SIR al 10%.



Figura 159. Métrica con longitud del símbolo 14, alpha 10 y SIR al 50%.



Figura 162. Métrica con longitud del símbolo 14, alpha 10 y SIR al 90%.



Figura 161. Métrica con longitud del símbolo 27, alpha 3 y SIR al 10%.



Figura 164. Métrica con longitud del símbolo 27, alpha 3 y SIR al 50%.



Figura 163. Métrica con longitud del símbolo 27, alpha 3 y SIR al 90%.



Figura 166. Métrica con longitud del símbolo 27, alpha 5 y SIR al 10%.



Figura 165. Métrica con longitud del símbolo 27, alpha 5 y SIR al 50%.



Figura 168. Métrica con longitud del símbolo 27, alpha 5 y SIR al 90%.



Figura 167. Métrica con longitud del símbolo 27, alpha 10 y SIR al 10%.



Figura 170. Métrica con longitud del símbolo 27, alpha 10 y SIR al 50%.



Figura 169. Métrica con longitud del símbolo 27, alpha 10 y SIR al 90%.