Research Article | Vol. 31, No. 8 / 10 Apr 2023 / Optics Express 12261
Optics EXPRESS

# Blur-specific image quality assessment of microscopic hyperspectral images

LAURA QUINTANA-QUINTANA,[1] SAMUEL ORTEGA,[2,1,*] HIMAR FABELO,[1,3] FRANCISCO J. BALEA-FERNÁNDEZ,[1,4] AND GUSTAVO M. CALLICO[1]

[1]*Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain*
[2]*Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway*
[3]*Fundación Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, Spain*
[4]*Department of Psychology, Sociology and Social Work, University of Las Palmas de Gran Canaria, Spain*
[*]*sortega@iuma.ulpgc.es*

**Abstract:** Hyperspectral (HS) imaging (HSI) expands the number of channels captured within the electromagnetic spectrum with respect to regular imaging. Thus, microscopic HSI can improve cancer diagnosis by automatic classification of cells. However, homogeneous focus is difficult to achieve in such images, being the aim of this work to automatically quantify their focus for further image correction. A HS image database for focus assessment was captured. Subjective scores of image focus were obtained from 24 subjects and then correlated to state-of-the-art methods. Maximum Local Variation, Fast Image Sharpness block-based Method and Local Phase Coherence algorithms provided the best correlation results. With respect to execution time, LPC was the fastest.



## 1. Introduction

Traditionally, the study of histology slides in medicine is considered as the gold standard for the clinical diagnosis of cancer and performing differential diagnosis with tissue alterations (e.g., infection or inflammation) [1]. In classical histology analysis, pathologists, using standard bright-field (BF) microscopy, visually examine the sample (regularities of cell shapes, distributions, etc.). Then, they study different parameters to determine if the tissue is cancerous and if positive, evaluate its malignancy level. This procedure is broadly used in hospitals, including prostate [2] and, cancer diagnosis [3].

Nowadays, the histopathology research trend is to digitize histology slides into whole RGB (red-green-blue) slides for further computational image analysis. This approach can enhance diagnosis accuracy, making it more objective in a shorter time [4]. Conventional RGB imaging differs from hyperspectral (HS) imaging (HSI) in the big number of spectral channels (also called bands or wavelengths) which HSI can acquire in the electromagnetic spectrum (within and beyond the visual range). This technology enables the highly precise differentiation of the materials captured, and so, it can help pathologists to identify and diagnose tissue histology samples with higher precision compared to traditional RGB microscopy [5].

In previous works of this research group, HS histological images were captured for tumor cell classification [6] or to synthesize HS images from standard RGB images of normal and cancer cells using conditional generative adversarial networks (GANs) [7]. However, images may not have high quality due to several distortions, being the focus one of the most relevant parameters in HS microscopic images [8]. BF microscopy is a good technique for capturing thin materials (one layer cell cultures or thin tissue portions) on glass slides but has limitations imaging thick samples [9,10].

#476949 https://doi.org/10.1364/OE.476949
Journal © 2023 Received 4 Oct 2022; revised 26 Dec 2022; accepted 30 Jan 2023; published 29 Mar 2023

Optics EXPRESS

Z-stacking techniques are a solution to blurred images produced by imaging non-flat samples. This approach uses images captured at several working distances (having different parts of the image focused) and stacks them together into one clearly focused image [11]. For this process to be automatic, image quality assessment (IQA) must firstly be evaluated [12]. Until now, the most precise quality evaluation of an image is made by the human eye since it is the end-user of the multimedia devices. However, obtaining subjective rating by users is a high costly and inefficient task [13]. Thus, effective, and objective IQA methods are needed for automatically predict the quality of images before further processing methods are applied.

In this work, we aim to quantify the focus of microscopic images captured at different working distances. First, a background framework will be presented to deeply explain the different IQA methods for blur quantification. Then, the materials and methods employed in the experiments will be described. A microscope with HS capabilities will be used to capture a hyperspectral image dataset for focus analysis (HIDFA) [14]. Focus of the images will be evaluated using the human visual system (HVS) and state-of-the-art algorithms [13]. Concluding, both type of evaluations will be correlated to analyze and discuss the performance of the studied algorithms applied to different configurations of the HS data, such as synthetic RGB (generated from the HS images) or monochromatic images.

## 2. Background framework

### 2.1. Limitations in HS microscopic data capturing

Blurring in HS microscopic images is highly influenced by the depth of field (*DOF*) [15]. is the length between the closest and the farthest points where an object can be captured in acceptably sharp focus (Fig. 1(a)). Eq. (1) shows the parameters involved to compute the *DOF*: focal length ($f$), which is the distance from the center of the magnifying lens to the focal point of the sensor; working distance ($WD$), which is the distance between the magnifying lens and the surface of the specimen; f-number ($N$), which is the ratio of the system's focal length to the diameter of the lens aperture; and diameter of circle of confusion ($c$), which is the light ring formed by the light beams when not focusing on the same sport. Since $c$ and $N$ are intrinsic to the optical system, the blurring (Fig. 1(c)) occurs when there are changes in $WD$ and/or $f$ (Fig. 1(b) and (d)).

$$DOF = \frac{2 \cdot WD^2 \cdot N \cdot c}{f^2} \tag{1}$$

First, changes in the $WD$ may occur because of the irregular nature of microscopic samples. Since the samples are not flat, different areas of the image can be focused on several *WDs* (Fig. 1(b)). Figure 1(c) shows the final image when capturing at 20× magnification a non-flat rat intestine sample using a custom designed HS microscope. Second, variations in $f$ could happen when several wavelengths are captured. A typical optical issue, chromatic aberration, takes place when a lens cannot guide all wavelengths' rays to the same focal plane (Fig. 1(d)).

### 2.2. Image quality assessment (IQA)

IQA is defined as a process that systematically evaluates the quality of an image based on human quality judgments (sharpness, graininess, tone scale, and color rendition) [16]. Focused captures are crucial in most of his applications (e.g., cancer diagnosis in histological samples [17]), therefore, IQA becomes a key technique that must be investigated. IQA algorithms can be divided into:

- **Full-Reference IQA (FR-IQA)** calculates the image quality index by comparing it with a ground truth capture (e.g., image versus its compressed version). The Peak Signal to Noise Ratio (PSNR) is a technique of this type [18].

Optics EXPRESS

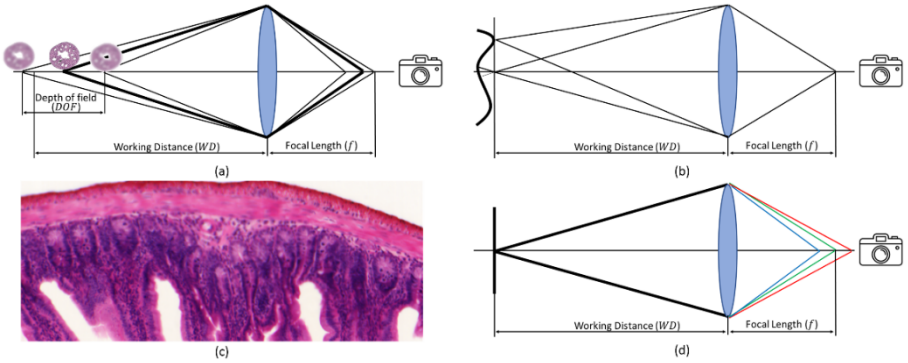


**Fig. 1.** Graphical representation of (a) DOF, (b) WD shift and, (d) chromatic aberration problems. (c) Example of an unfocused rat intestine (20×).

- **Reduced-Reference IQA (RR-IQA)** extracts features from an image and its reference for further calculation of the image quality (e.g., extraction of image features for easier recompositing after compression). The free-energy-based distortion metric is an algorithm of this type [19].
- **No-reference IQA (NR-IQA)** does not need any reference image for providing the quality of an image (e.g., focus of blurred image). The perceptual sharpness index (PSI) algorithm is classified in this type [20].

In this work, the interest is to enhance the quality of microscopic HS images. Sometimes, specimens are not perfectly flat, and the high gradient of the sample is bigger than the DOF offered by the imaging system. The aim of this work is to quantify the blurriness of the captured image, for a later selection of the most focused image regions. However, due to the lack of ground truth of blur images, NR-IQA is the best option, but it comes with its challenges. State-of-the-art studies indicate that the most relevant strategy for judging image quality when no reference is available, is HVS [13]. Thus, effectiveness of NR-IQA methods can be calculated by correlating their results (called Objective Scores (OS)) to the values given to the images by the HVS (called Subjective Scores (SS)). NR-IQA algorithms applied to RGB captures are further classified into learning-free (edge-free and edge-based), and learning-based (wavelet and Fourier like) methods. A general summary of the different techniques can be found in Fig. 2 [13]. Next Sections 2.4 and 2.5, will show a more detailed explanation of the state-of-the-art methods to finally select the ones employed in this work at Fig. 7.

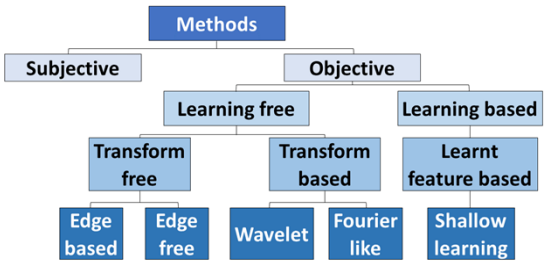


**Fig. 2.** Summary of the main NR-IQA methods found in the literature.

Optics EXPRESS

### 2.3. Database design

Before calculating image quality scores, a database (DB) showing different levels of blurriness must be created. The state-of-the-art shows that available DBs for NR-IQA are mainly based on RGB images, where blurriness is obtained synthetically by applying a filter over a reference image. These DBs employ several distortions, such as gaussian noise, chromatic aberrations, or image denoising to the reference images using several intensity levels.

Some of the most famous RGB DBs used by numerous researchers are the TID2013 [21], LIVE [22] and CSIQ [23]. Only a few multi/HS image DBs have been found, being the most relevant one the SIDQ DB [24]. A summary of the state-of-the-art in DBs for NR-IQA is shown in Table 1.

**Table 1. Summary of the Characteristics of the NR-IQA State-of-the-Art DBs**

| | No. reference images | Spectral range (nm) | Spatial range | Types of distortion | Levels of distortion | Total No. of images |
|---|---|---|---|---|---|---|
| **TID2013** [21] | 25 | RGB | Macro | 24 | 5 | 3000 |
| **LIVE** [22] | 29 | RGB | Macro | 5 | 5 | 779 |
| **CSIQ** [23] | 30 | RGB | Macro | 6 | 4 to 5 | ~810 |
| **SIDQ** [24] | 9 | 410 - 1000<br>160 bands | Macro | 5 | 1 | 45 |

### 2.4. Subjective scores (SSs)

Traditional methodologies for *SS* calculation are based on simple stimuli. Nowadays they are more complex, also registering user strategies and confidence in an opinion [25]. Some of the new parameters under study are measurement of consistency of the opinion (reliability), of correct feature (validity) and of specific empirical manipulation (sensitivity) [26]. Before planning a test to obtain subjective ratings, the panel of subjects must first be decided following the considerations shown in Table 2. Afterwards, a testing method should also be chosen. The main characteristics of the different methods can be summarized as follows in Table 3 [25]. Finally, a summary of the main characteristics of the most famous NR-IQA DBs is shown in Table 4.

**Table 2. Characteristics of Subject's Panel for NR-IQA**

| Parameter | Range |
|---|---|
| **No. of evaluators** | There are several ITU (International Telecommunication Union) recommendations (Rec.), such as ITU-T Rec. P.911 [27] or ITU-R Rec. BT.500-11 [28], which describe subjective assessment methods for evaluating one-way overall audiovisual quality at multimedia applications. ITU-T Rec. P.911 states: "The possible number of subjects in a viewing and listening test [. . . ] is from 6 to 40. Four is the absolute minimum for statistical reasons, while there is rarely any point in going beyond 40". Moreover, ITU-R Rec. BT.500-11 advocates at least 15 subjects. |
| **Representative evaluation panel for the target application** | Subject demographic data (e.g., sex, age, culture, and education level) are important and can affect the *SS* [29]. |
| **Expertise of the evaluation panel** | Expert users are usually in better agreement than regular users. This way, a smaller number of people is needed for the experiment. However, their critical opinions may bring them to lower ratings, creating a bias in the measurements [30,31]. |
| **Testing environment** | ITU recommendations remark that the experiment must be carried out under controlled lab conditions, although it may not perfectly reconstruct the target application conditions [32]. |

Optics EXPRESS

**Table 3. Characteristics of Subjective Tests for NR-IQA**

| Parameter | Range |
|---|---|
| **No. of images compared** | It can be single, double, or multi-stimulus. Multiple stimuli may be presented simultaneously or sequentially. |
| **No. of times an image is presented** | It could be from one to multiple times. |
| **Presence of a reference image** | It can be shown or hidden, depending on whether the subjects know which one are them. |
| **Reference image rating** | Users may value the quality of the image alone, both the image and the reference, or the contrast between both. |
| **Interactivity** | One or more subjects may be rating the images at the same time. |
| **Rating collection** | Ratings may be collected per image (time-discretely) or per time interval (continuously). Recommendations given in [33] indicates that the total time a user takes to fill a subjective test should not exceed 30 minutes. |
| **Range of test images** | Usually, rating scales have five discrete levels (e.g., from "bad" to "excellent"), but language dependent scales are non-linear due to different understanding of the words. Thus, bigger ranges are also employed (e.g., 7, 9 or 11 points). Continuous scales could be employed too. However, this option may incorporate noise to the ratings due to limitation of humans to distinguish infinite levels of a certain quantity. Moreover, different users can have quite several opinions about the quality range of the images under study, so high and low anchor captures could be shown for better scale understanding [34]. |

**Table 4. Subjective Test Features for Dataset Labeling**

| | | TID2013 [21] | LIVE [22] | CSIQ [23] | SIDQ [24] |
|---|---|---|---|---|---|
| **Subject's Panel** | **No. Observers** | 971 | 161 | 25 | 14 |
| | **Expertise in evaluation panel** | Students, tutors, and researchers | Male college students | NF | 22-66 years old |
| | **Testing environment** | Controlled lab setup or Internet, flexible viewing distances | Controlled lab setup, flexible viewing distances | Controlled lab setup, fixed viewing distance | Dark room |
| **Subjective Test** | **No. of images compared** | 3 | 1 | 866 | 3 |
| | **No. of times an image is presented** | 9 | 1 | 1 | NF |
| | **Presence of a reference image** | Yes | No | No | Yes |
| | **Reference image rating** | No | No | No | No |
| | **Interactivity** | Indifferent | Indifferent | Indifferent | Indifferent |
| | **Rating collection** | ~17 mins | NF | NF | NF |
| | **Range of test images** | Pair-wise sorting with respect to the original<br>(0-9) | Excellent, Good, Fair, Poor, Bad<br>(1-5) | Linear displacement of the images across four calibrated LCD monitors<br>(0-1) | Pair-wise sorting with respect to the original.<br>Ties allowed.<br>(0.5 to majority hit rate) |

Optics EXPRESS

Before the testing, subjects are typically given training to familiarize them with the interface. Once the test has taken place, studying the results can help removing unreliable subject responses. However, we must be careful since it may also eliminate just valid opinions. After screening, the *SSs* are transformed to mean opinion scores (MOSs) [25], which inevitably represents the view of a majority. Over the years, MOS have become the most popular ratings for media quality. But we must consider that the arithmetic averaging of opinions assumes homogeneity among subjects. Thus, statistics should be applied over MOS values [25].

### 2.5. Objective scores (OSs)

Although we have seen that subjective quality tests are essential for image quality analysis, they involve a lot of disadvantages (e.g., they are time-consuming to perform, biased by the user, etc.). Thus, objective quality scores (*OSs*) should be employed. The following sections explain the objective models, how they work and some applications in which they are used.

#### 2.5.1. Edge based

Cumulative probability of blur detection (CPBD) [35] is based on the study of human blur perception for varying contrast values. Edge pixels are firstly counted, if no edges are detected, the area is classified as smooth block and, edge block otherwise. For the last ones, contrast and edge width are computed. Then, the probability of finding blur at each edge is calculated and the normalized histogram of these values give the final CPBD value.

In [36], a NR-IQA metric is presented based on edge model (EMBM). The edge model can estimate width and contrast at the same time for each pixel. Thus, probability of belonging to a salient edge pixel can be determined, simulating the blur assessment performed by the HVS. The final value is obtained by adding the probability of the different pixels.

Perceptual sharpness index (PSI) [20] is calculated based on the statistical analysis of local edge gradients. It selects the biggest edges in the image through an adaptive edge selection procedure. Then, the edge widths of the selected edges are computed and the ones above the Just Noticeable Blur (JNB) width are subtracted. These steps are repeated in a block-wise way to create the local sharpness map, where the highest qth percentile average of the local sharp scores provide the final PSI score.

#### 2.5.2. Edge free

ARISM (autoregressive-based image sharpness metric) is a blind edge free method for focus quantification. It calculates the sharpness score by examining the variance of the locally evaluated autoregressive values in a pointwise way. ARISM also takes into account the color information to assess the sharpness of the image. In [37], the authors propose ARISMc, after extending ARISM to the YIQ space.

Maximum local variation (MLV) is defined by Bahrami et al. [38] as the higher intensity contrast of a pixel with respect to its 8-neighbors. Sharpness is defined by high variations, thus the pixels' MLVs are subjected to a weighting strategy. Heavier weights are assigned to greater MLVs, making the tail end of the MLV distribution thicker, and becoming more discriminative for different blur degrees. At the end, the metric value is computed from the standard deviation of the weighted MLV distribution.

#### 2.5.3. Wavelet

The fast image sharpness (FISH) method estimates both local and global image sharpness [39]. First, it decomposes the image in three different discrete wavelet transformations (Cohen-daubechies-fauraue 9/7 filters [6]). Next, the log-energies of each sub-band are calculated for all the transform levels. Finally, the sharpness score is calculated by a weighted average of these log-energies.

Optics EXPRESS

Hassen et al. [40] followed the idea that local phase coherence (LPC) structures are just shown in sharp edges. That way, they proposed the LPC-based sharpness index (LPC-SI). The algorithms first pass an image through 3-scale 8-orientation log-Gabor filters and then, repeat for each orientation and spatial location. They do not use block-based computation, but an efficient algorithm that largely simplifies the LPC calculation, making it easily usable in multiple applications.

#### 2.5.4. Fourier like

Blind image blur evaluation (BIBLE) method [41] starts by calculating the gradient of an image to evaluate its shape. Afterwards, the gradient map is split into blocks and the Tchebichef moments are computed. The sum of squared non-DC moment values defines the energy of a block. At the end, the variance-normalized moment energy (simulating the HVS) provides the BIBLE index.

Vu et al. [42] proposed a spectral and spatial sharpness ($S3$) measure. The spectral measure $S1(x)$ is calculated by the reduction of high-frequency components in unfocused captures followed by a rectification using the sigmoid function to follow the HVS. The spatial measure $S2(x)$ is obtained based on the local total variation. The sharpness map $S3$ is calculated by obtaining the geometric average of $S1(x)$ and $S2(x)$ of each independent block. Finally, to consider the HVS, the final blur index is calculated as the mean of the largest 1% scores of $S3$.

#### 2.5.5. Shallow learning

SPARISH (NR Sparse representation-based Image Sharpness) is based on sparse representations, representing signals with as few as possible significant coefficients [43]. Since a focused capture is defined by the localization of edges, a dictionary (made of edge patterns) can be used to measure the blurriness level. The sharpness score is defined as the variance-normalized energy (relative strength of blur in an image) of a group of chosen high-variance blocks. This algorithm is beneficial for real-world applications since it is not sensitive to training images and so, can evaluate different images using a universal dictionary.

## 3. Materials and methods

### 3.1. Instrumentation and sample description

The instrumentation employed in this study consists of an HS camera coupled to a conventional BF microscope (Fig. 3). The BF microscope is an Olympus BX-53 (Olympus, Tokyo, Japan). The HS camera is a Hyperspec VNIR A-Series from HeadWall Photonics (Fitchburg, MA, USA), which is based on an imaging spectrometer coupled to a CCD (Charge-Coupled Device) sensor, the Adimec-1000 m (Adimec, Eindhoven, Netherlands). It has four magnification lenses: 5×, 10×, 20× and 50×. These numbers refer to the magnification achieved by the system to capture the sample (e.g., a 4× objective lens is magnifying the sample four times its size). This HS system works in the visual and near-infrared (VNIR) spectral range from 400 to 1000 nm with a spectral resolution of 2.8 nm, sampling 826 spectral channels and 1004 spatial pixels per line. The push-broom camera performs spatial scanning to acquire an HS cube with a mechanical stage (SCAN, Märzhäuser, Germany) attached to the BF microscope, which provides accurate movement (±3 µm accuracy) of the specimens in the 3 axes directions: $[x, y, z]$. The objective lenses are from the LMPLFLN (Long working distance Plan SemiApochromat) family (Olympus, Tokyo, Japan), which are optimized for infra-red (IR) observations. The light source is a 12 V, 100 W halogen lamp. This system was previously employed in histological HS analysis of brain cancer samples [17], breast tumor cell detection [6] and disease biomarker identification in plasma [44]. Furthermore, to create our DB, 125 plant and animal histology samples (rat histology, stems, leaf structure and blood smears) from the company Brunel microscopes Ltd

Optics EXPRESS

(England, UK) were employed (Fig. 3). These microscope slides are characterized by being theoretically flat, so no roughness is expected in the specimens.

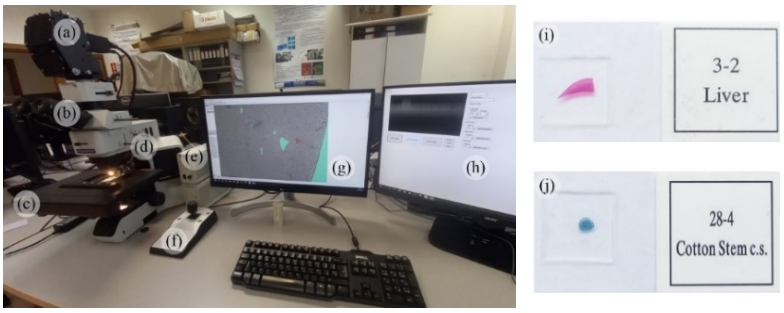


**Fig. 3.** Laboratory HS microscope employed in this work. (a) HS camera. (b) Eyepieces. (c) Motorized stage. (d) Objectives. (e) Light sources. (f) Joystick to move the stage. (g) RGB imaging software. (h) HS imaging software where a single HS frame can be visualized for focusing purpose. (i) Example of a histology rat liver tissue sample. (j) Example of histology cotton stem sample.

### 3.2. Database description

A dataset for focus analysis was obtained by capturing several images at different *WDs* to produce different levels of blurriness. Samples to be captured included rat intestine (Fig. 4(a)), bird blood (Fig. 4(b)), cotton leaf (Fig. 4(c)) and cladophores (Fig. 4(d)). However, the last ones were the least suitable for this experiment since they have a homogeneous spatial frequency and differences in blurriness are difficult to differentiate. Moreover, algae samples were also tested for this experiment, but it was discovered that they were non-flat and so, they cannot be included in these experiments (Fig. 4(d)).

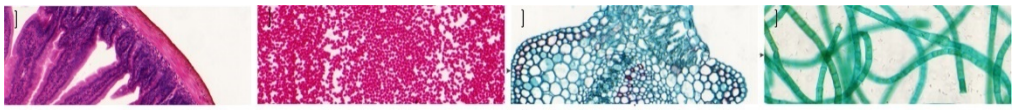


**Fig. 4.** Focused ROI images captured at 10× of different samples: (a) Rat intestine tissue; (b) Bird blood smear; (c) Cotton Leaf; and (d) Cladophores Algae. Scale bar: 50 µm.

Concerning the objectives lenses, in our HS acquisition system light attenuation is inversely proportional to the magnification. Thus, the higher the magnifying power, the higher the power of light it is needed to capture the sample. Using the default light system, it is not able to capture clear images using 50× objective lens. So, knowing that tissue structures captured at different magnifications can resolve different classification problems, three objective lenses (creating different pixel sizes) were selected (5×, 10×, 20×).

To select the number of images in the database, we had to consider that they would have to be reviewed by different users. Thus, following the ITU recommendations cited in Table 3, the total time of the test should not exceed 30 minutes. A pair of images is evaluated in around two seconds [45] so, it is possible to perform 900 rounds in 30 minutes. However, to be able to include some more images, we opted to design the subjective test in two sections of 20 minutes each, being able to perform 1200 pair comparisons in total. Following the swizz tournament, 24 pair comparisons (11 images, 2 images per round, 4 round per sample) are needed for each

Optics EXPRESS

sample. In conclusion, no more than 50 samples could be tested per magnification (considering that each magnification is performed by different subjects).

A summary of the number of reference images captured at each magnification and using each sample is shown in Table 5. The total number of images refers to the total number of reference images multiplied by the 11 levels of distortions selected in this work (i.e., number of images captured at different WDs for each reference image).

**Table 5. Summary of the number of images and samples included in the HIDFA database.**

| | No. of reference images | | | | Total No. of images |
|---|---|---|---|---|---|
| | 5× | 10× | 20× | Total | |
| **Rat histology** | 8 | 16 | 16 | 40 | 440 |
| **Leaf structure** | 5 | 10 | 6 | 21 | 231 |
| **Stems** | 16 | 10 | 19 | 45 | 495 |
| **Blood smears** | 0[a] | 7 | 5 | 12 | 132 |
| **Freshwater Algae** | 0[a] | 3 | 3 | 6 | 66 |
| **Total** | 29 | 46 | 49 | **124** | **1364** |

[a]No images were captured of blood smears and freshwater algae at 5× because they are mostly homogeneous and blur quantification algorithms could not perform properly.

The focusing process on HS data capturing using push-broom cameras is based on the subjective approach of the user. The HS image is focused by looking into one frame given by the push-broom camera ($Y\lambda$ frame shown in the grey image on the right at Fig. 5) and finding the sharpest spatial frequency along the different $WD$ from the sensor to the sample. For each specimen, the focus point, $WD_0$, was firstly found (Fig. 5). The procedure continued by moving upwards a distance of $\Delta z\frac{(N-1)}{2}$. From this point, $N$ HS images were captured distancing $\Delta z$ each time from the previous image. $N$ was always odd, producing the same number of slices upwards and downwards of $WD_0$. For each set of $N$ HS images captured, a white and a dark reference image were also obtained at $WD_0$.

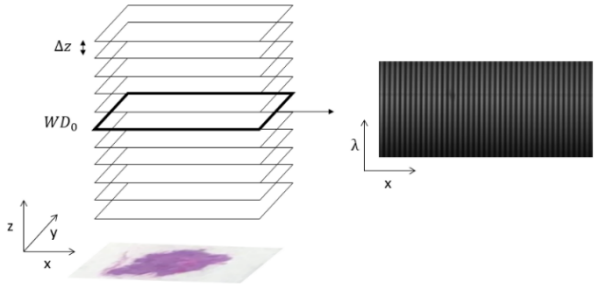


**Fig. 5.** Proposed HS DB capture procedure. Firstly, $WD_0$ was found and then $N$ HS images were captured distancing $\Delta z$ from each other.

Conventional HS calibration was performed over the HS cubes [8]. This is a standard procedure for flat field correction to transform captures to normalized transmittance. A push-broom frame is a capture of one spatial line containing the whole spectra of each pixel (1004 pixels × 826 spectral bands using the system of this work). During the capture process, a white reference image is recorded by capturing one push-broom frame of the light incident on the HS sensor (See Fig. S3 on Supplement 1). Similarly, the dark reference is obtained by turning off the light from

Optics EXPRESS

the system. Thus, an upper and lower light range are set. Then, Eq. (2) is performed over each push-broom frame (R) separately, to normalize the light.

$$I = \frac{R - DR}{WR - DR} \tag{2}$$

Then, preprocessing of the images were performed to reduce noise by removing the extreme bands as studied in [46], resulting HS cubes of 645 bands. After capturing and preprocessing the HS cubes, monochromatic images and synthetic RGB images were extracted. Monochromatic images were obtained by averaging the 645 bands of the preprocessed HS cubes. Synthetic RGB images were obtained by applying the normal probability density function for each RGB channel (R = 470 ± 0.04, G = 560 ± 0.06 and, B = 590 ± 0.08 nm), trying to imitate the HVS (see section S4.2. of Supplement 1).

Summarizing, 125 samples were selected for this specific application, captured at different *WD* from the focused plane and then, HS images were calibrated and preprocessed. Finally, a total of 1,375 HS images conformed the HIDFA and were used for testing different OSs.

### 3.3. Metrics for blur assessment

#### 3.3.1. Subjective WD-based score ($SS_{WD}$)

This score was given to each image according to their different *WDs* between the specimen and the objective lens. Thus, *WD* of each image can be calculated following Eq. (3), knowing that $WD_0$ is the point where the HS image is focused by visual inspection and, $\Delta z$ was the increment distance between captures (0.2, 0.1 and 0.05 mm for 5×, 10× and 20×, respectively).

$$SS_{WD} = \frac{WD_0 \pm WD}{\Delta z} \tag{3}$$

#### 3.3.2. Subjective MOS score ($SS_{MOS}$)

This score was obtained in the form of MOS. All-play-all blurriness comparison between these images is practically non-viable, thus, the Swiss system tournament was employed. The principle of a Swiss competition is that each player will be pitted against another player who has performed as well (or as poorly) as themself. The first round is either drawn at random or seeded according to rating. The advantages of this method are twofold: it can find a clear winner between many competitors (the one that reach to the end without a single bad result), and it has a relatively small number of total rounds [47]. In our case, tournaments were organized between 11 different blur versions of a reference image, where users had to select the image closest to the reference one (Fig. 6). Ties were allowed. The first round was set randomly. Tournaments were developed in the scope of Microsoft Access in Windows 10. Moreover, subjective tests took 45 minutes, making a 5-minute pause in the middle of the experiment.

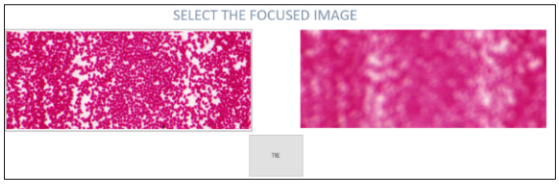


**Fig. 6.** Example of a round following the Swiss tournament system.

After all tests were completed MOS values for each single image were calculated following Eq. (4), where $R_n$ is the rating of a subject over such image and $N$ the total number of subjects

Optics EXPRESS

that went through the subjective test.

$$SS_{MOS} = \frac{\sum_{n=1}^{N} R_n}{N} \tag{4}$$

#### 3.3.3. Objective scores

The main OSs employed for blur assessment were explained in section 2.5. In our experiments, we are concentrated on studying learning free methods, which can be further divided as shown in Fig. 7. From each of the four learning-free NR-IQA algorithm categories, the best available algorithm for monochromatic images was chosen. First, these algorithms were evaluated in a set of monochromatic images (computed from the mean of all wavelengths of the preprocessed HS image) at different *WDs*, since there is dependency between *WD* and focus. Then, they were applied to each band of each image, independently, to show the dependency between the wavelength and the focus (to prove the existence of any chromatic aberration). Experiments were carried out in MATLAB R2020a in a Windows environment (Microsoft Windows 10 Pro) with an Intel i7-10700 K 3.80 GHz and 64 GB RAM.

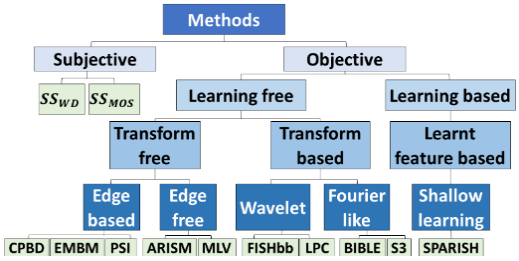


**Fig. 7.** Summary of all scores employed in this work.

#### 3.3.4. Results evaluation

Once all the scores employed in this project have been explained in detail, next step was to correlate their results to test the quality of the different objective methods. To evaluate the obtained results, the Video Quality Experts Group (VQEG) [48] proposes mapping the *SS* to the *OS*. Since two different *SSs* ($SS_{WD}$ and $SS_{MOS}$) were computed for each capture, three different correlations were performed: 1) $SS_{WD}$ to $SS_{MOS}$; 2) $SS_{WD}$ to *OS*; and, 3) $SS_{MOS}$ to *OS*.

VQEG performs the mapping following Eq. (5), where $\tau_1 = \max \text{F(OS)}$, $\tau_2 = \min \text{F(SS)}$, $\tau_3 = mean(SS)$, *and* $\tau_4 = \frac{std(SS)}{4}$.

$$(SS) = \frac{\tau_1 - \tau_2}{1 + e^{\frac{SS - \tau_3}{\tau_4}}} + \tau_2 \tag{5}$$

Afterwards, three evaluation metrics were calculated to evaluate the method's performance: 1) Spearman's Rank-Order Correlation Coefficient (SRCC), which employs a monotonic function; 2) Pearson's Linear Correlation Coefficient (PLCC), which is a measure of the linear correlation after nonlinear mapping; 3) and Root-Mean-Square Error (RMSE), which is used to measure the differences after the nonlinear mapping. In competitive NR-IQA algorithms, the values of SRCC and PLCC are close to one, while the value of RMSE is close to zero.

## 4. Experimental results

### 4.1. HIDFA creation

Following the common procedures to create a database for IQA, the HIDFA database was generated (section S3, Fig. S2, Fig. S3 and Fig. S4 from Supplement 1) and it is now publicly

Optics EXPRESS

available (more information in data availability). HIDFA is characterized by increasing the number of reference images to 125 and the number of bands to 826 (Table 6). At the end, a total of 1,375 HS images were captured. ITU recommendations, explained in Table 2, indicated that at least 15 subjects are needed but no more than 40. Thus, a number of 24 participants was selected. Since each magnification would need a different reviewer, only images at 10× were employed for $SS_{MOS}$ calculation. Concluding, twenty-four subjects evaluated a total of 46 reference images at 11 levels of distortion (506 images), resulting in 12,144 ratings.

**Table 6. Comparison between HIDFA and other NR-IQA DBs**

| | No. reference images | Spectral range (nm) | Spatial range | Types of distortion | Levels of distortion | Total No. of images |
|---|---|---|---|---|---|---|
| **TID2013** [21] | 25 | RGB | Macro | 24 | 5 | 3000 |
| **LIVE** [22] | 29 | RGB | Macro | 5 | 5 | 779 |
| **CSIQ** [23] | 30 | RGB | Macro | 6 | 4 o 5 | ~810 |
| **SIDQ** [24] | 9 | 410–1000 (160 bands) | Macro | 5 | 1 | 45 |
| | 29 at 5× | 400–1000 (826 bands) | Micro | 1 | 11 | 1364 |
| **HIDFA [14]** | 46 at 10× | | | | | |
| | 49 at 20× | | | | | |

Sociodemographic data of the subjects were studied to determine if $SS_{MOS}$ values were independent of the subject characteristics (Table 1 of Supplement 1). Subjects aged between 20 and 76 years (Fig. S1 of Supplement 1), being 15 men and 9 women. Two thirds of them were qualified workers (only two did not go to university) while the rest were students. Thus, the average intellectual activity presented medium-high values. Regarding to the clinical variables, except for two subjects, they did not have a psychiatric history. None of the subjects presented hypertension or type II mellitus diabetes. In terms of visual acuity, we found some variability: two thirds had normal vision, while the rest had some vision problem, and so, wore glasses (except for one person with astigmatism who did not wear glasses). One of the subjects with normal vision had an eye surgery (more information on section S1. of Supplement 1).

Moreover, to determine whether sociodemographic data could be related to $SS_{MOS}$ results, a statistical analysis was performed. To this aim, $SS_{MOS}$ scores were calculated for each categorical value (e.g., $SS_{MOS}$ for men and women). Standard deviation was computed over the 11 levels of distortions for a single reference image. The result were vectors of 46 values one per reference image. Finally, t-test and ANOVA [49] methods were performed between the standard deviation vectors of a feature with two or more categories, respectively. Statistical tests were employed to evaluate, at the 5% of significance level, if the null hypothesis is rejected (each variable is independent to the $SS_{MOS}$ results). None of the features presents a $p<0.05$, so, not statistically significance was found. According to this statistical analysis, we can argue that $SS_{MOS}$ results are independent to the sociodemographic data of the subjects.

### 4.2. *SS analysis*

$SS_{MOS}$ values of the images taken at 10× (46 set of images) were calculated. To this aim, synthetic RGB images were generated by imitating the human eye spectral response [50]. Figure 8 shows the distribution of $SS_{MOS}$ (from 1 to 6) along the $SS_{WD}$ for four sample z-stack groups. Same results were obtained for all set of images of HIDFA. Mean correlation between $SS_{MOS}$ and $SS_{WD}$ was measured, achieving $SRCC = 0.89$, $PLCC = 0.8$ and $RMSE = 0.77$.

Optics EXPRESS

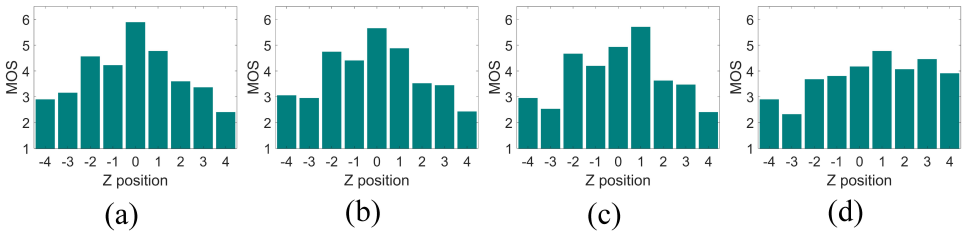


**Fig. 8.** Results of $SS_{MOS}$ values for sample z-stack groups: (a) rat intestine tissue; (b) bird blood smear; (c) cotton leaf; and (d) cladophores.

### 4.3. *OS score analysis*

OSs were calculated over the synthetic RGB images, and the results were correlated to the $SS_{WD}$ and the $SS_{MOS}$, as displayed in Table 7. Overall results are similar using both SSs. The best performance was achieved employing MLV, FISHbb, LPC and S3 algorithms. Same procedure was repeated for the monochromatic HS images obtained averaging all the bands of the HS cube (Table 8). In these results, ARISM, BIBLE and S3 algorithms were not computed since they only work for RGB images. The best performance was achieved employing MLV, FISHbb and LPC algorithms. Since there is a high correlation between the results obtained for $SS_{WD}$ and $SS_{MOS}$, the latter was used in the next experiments to simplify the presentation of the results.

**Table 7. Correlation Results for RGB Images**

| | Method | | CPBDM [35] | EMBM [36] | PSI [20] | ARISM [37] | MLV [38] | FISHbb [39] | LPC [40] | BIBLE [41] | S3 [42] | SPARISH [43] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $SS_{WD}$ | **SRCC↑** | **Mean** | 0.62 | 0.69 | 0.72 | 0.50 | 0.84 | **0.90** | 0.83 | 0.72 | **0.87** | 0.7 |
| | | **Std** | 0.26 | 0.24 | 0.25 | 0.26 | 0.15 | **0.06** | 0.18 | 0.23 | **0.08** | 0.22 |
| | **PLCC↑** | **Mean** | 0.63 | 0.64 | 0.70 | 0.51 | 0.81 | **0.84** | **0.82** | 0.76 | 0.81 | 0.76 |
| | | **Std** | 0.24 | 0.22 | 0.22 | 0.28 | 0.12 | **0.05** | **0.13** | 0.13 | 0.06 | 0.13 |
| | **RMSE↓** | **Mean** | 2.32 | 2.33 | 2.04 | 2.50 | **1.80** | 1.98 | **1.72** | 2.12 | 2.05 | 2.14 |
| | | **Std** | 0.41 | 0.34 | 0.18 | 0.52 | **0.18** | 0.09 | **0.21** | 0.18 | 0.10 | 0.18 |
| $SS_{MOS}$ | **SRCC↑** | **Mean** | 0.61 | 0.67 | 0.69 | 0.5 | 0.80 | **0.85** | 0.82 | 0.66 | **0.85** | 0.65 |
| | | **Std** | 0.27 | 0.27 | 0.28 | 0.26 | 0.25 | **0.21** | 0.23 | 0.28 | **0.19** | 0.28 |
| | **PLCC↑** | **Mean** | 0.62 | 0.59 | 0.67 | 0.5 | 0.77 | **0.80** | **0.80** | 0.70 | 0.77 | 0.7 |
| | | **Std** | 0.24 | 0.22 | 0.27 | 0.27 | 0.24 | **0.17** | **0.21** | 0.23 | 0.17 | 0.22 |
| | **RMSE↓** | **Mean** | 2.33 | 2.33 | 1.94 | 2.57 | **1.63** | 1.80 | **1.50** | 1.98 | 1.90 | 2 |
| | | **Std** | 0.63 | 0.47 | 0.44 | 0.71 | **0.40** | 0.31 | **0.45** | 0.30 | 0.22 | 0.3 |

### 4.4. *OS analysis over single bands*

After the evaluation of the focus *OS* on synthetic RGB and monochromatic images, we tested the focus measurement on the different bands of the HS cubes independently. The OS were computed over single band images (except for ARISM, BIBLE and S3 algorithms). Figure 9 shows the comparison of the results obtained for the monochromatic (mean over bands before metric computation) versus single band images (mean over bands after metric computation). Results show that methods can be divided in two: the ones with not similar means (CPBD, EMBM, PSI and, SPARISH) and the ones with similar means (MLV, FISHbb and, LPC).

For clarification, metric values were plotted against wavelength for focus visualization (Fig. 10). Some metrics seem to find differences between bands (CPBD, EMBM, PSI and, SPARISH),

Optics EXPRESS

**Table 8. Correlation Results for Monochromatic Images**

| | Method | | CPBDM [35] | EMBM [36] | PSI [20] | MLV [38] | FISHbb [39] | LPC [40] | SPARISH [43] |
|---|---|---|---|---|---|---|---|---|---|
| $SS_{WD}$ | **SRCC↑** | **Mean** | 0.6 | 0.73 | 0.47 | **0.89** | **0.9** | 0.83 | 0.61 |
| | | **Std** | 0.31 | 0.13 | 0.27 | **0.09** | **0.07** | 0.14 | 0.25 |
| | **PLCC↑** | **Mean** | 0.61 | 0.74 | 0.49 | **0.83** | **0.84** | 0.81 | 0.69 |
| | | **Std** | 0.29 | 0.06 | 0.27 | **0.07** | **0.06** | 0.12 | 0.22 |
| | **RMSE↓** | **Mean** | 2.44 | 2.18 | 2.34 | 1.99 | 1.96 | **1.78** | 2.17 |
| | | **Std** | 0.54 | 0.18 | 0.47 | 0.09 | 0.09 | **0.18** | 0.2 |
| $SS_{MOS}$ | **SRCC↑** | **Mean** | 0.6 | 0.72 | 0.45 | **0.85** | **0.85** | 0.8 | 0.56 |
| | | **Std** | 0.31 | 0.18 | 0.29 | **0.24** | **0.23** | 0.24 | 0.29 |
| | **PLCC↑** | **Mean** | 0.62 | 0.69 | 0.45 | 0.8 | **0.81** | 0.8 | 0.64 |
| | | **Std** | 0.28 | 0.16 | 0.3 | 0.17 | **0.18** | 0.19 | 0.26 |
| | **RMSE↓** | **Mean** | 2.46 | 2.11 | 2.45 | 1.8 | 1.76 | **1.57** | 2.06 |
| | | **Std** | 0.78 | 0.33 | 0.66 | 0.33 | 0.33 | **0.46** | 0.36 |

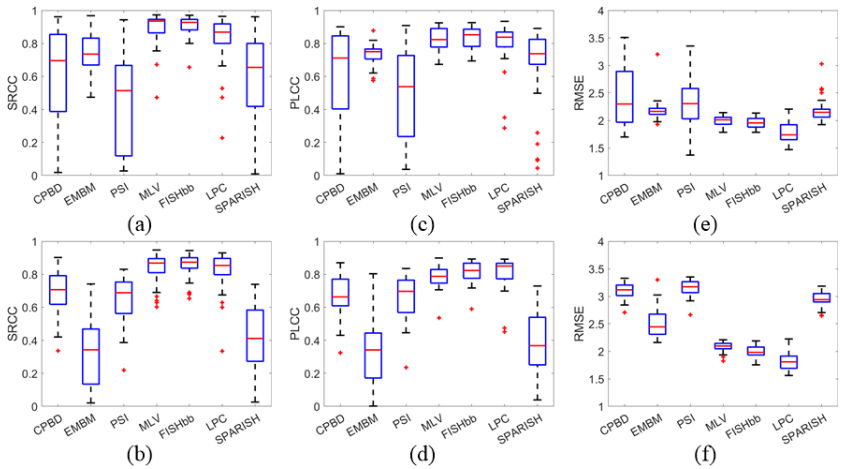


**Fig. 9.** Results of $SS_{MOS}$ values versus *OS* for monochromatic images (a) SRCC; (c) PLCC; and the RMSE. Results of $SS_{MOS}$ values vs *OS* for single band images (b) SRCC; (d) PLCC; and (f) RMSE.

Optics EXPRESS

while others keep a constant focus metric over the different wavelengths (MLV, FISHbb and, LPC). These results agree with the ones previously obtained.

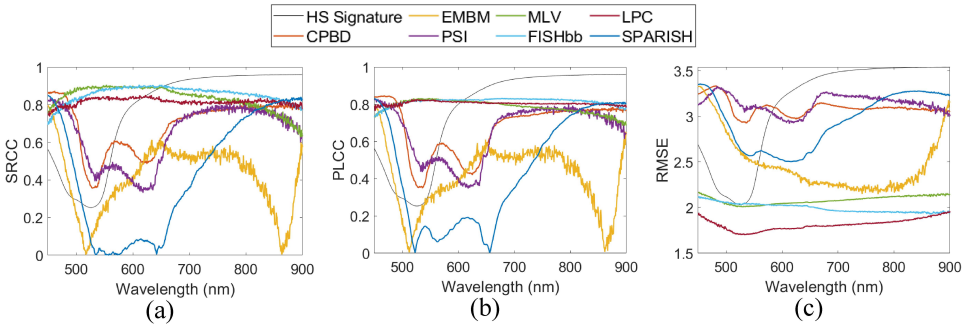


**Fig. 10.** Results of $SS_{MOS}$ values versus $OS$ for each single band image using the selected algorithms: (a) SRCC; (b) PLCC; and (c) RMSE.

### 4.5. *OS analysis over single sample groups*

To make a reliable wavelength comparison, samples must be divided into samples groups (blood smears, leaf structures, rat histology and stems). Figure 11 presents the results of the $OS$ for each sample group and their mean HS signature. Overall, it can be deduced that worst metric values are obtained at lower information bands (HS signature lower peaks). The low blur value may be due to the few edges found in images with almost no color in such wavelength.

### 4.6. *OS analysis over single frames*

As it is not clear which algorithms are behaving accordingly to reality, further analysis was performed. Specific cases were studied by displaying a single band image from HS cubes for visual examination. Figure 12 show the monochromatic picture and several single band images of a rat histology capture at their most focused point ($SS_{WD} = 6$). PSI was the method selected to represent the ones that seem to find differences between bands, while FISHbb was representing the ones that keep a constant focus metric over the different wavelengths. PSI values decrease along with the loss of information in the image, while FISHbb values keep constant values. Thus, for an image with the same $SS_{WD}$ for all wavelengths, FISHbb better recognizes the focus level along all bands.

### 4.7. *Summary of the experimental results*

In this work, several results have been obtained. Firstly, in order to test different blur quantification algorithms, the HIDFA was generated. Then a subjective test was carried out over 24 participants in order to recollect the MOS values of each synthetic RGB image of the dataset. Using these data, several state-of-the-art algorithms have been employed to select the most appropriate one for HS microscopy samples. MLV, FISHbb, LPC and S3 were the algorithms that provided the best results on synthetic RGB images. Afterwards, since HSI is intended for data collection of data across the light spectrum, each wavelength image was analyzed independently (ARISM, BIBLE and S3 methods could not be used because they only work over RGB images). From the results obtained, algorithms can be divided in two groups: methods that obtain similar values for the entire spectrum (G1: MLV, FISHbb and, LPC) and the ones that obtain different values (G2: CPBD, EMBM, PSI and, SPARISH). Afterwards, the images were divided into their original sample groups and the main HS signature was calculated for each one of them. In Fig. 11 it can be appreciated that G2 algorithms behave according to the information found in each band.

Optics EXPRESS

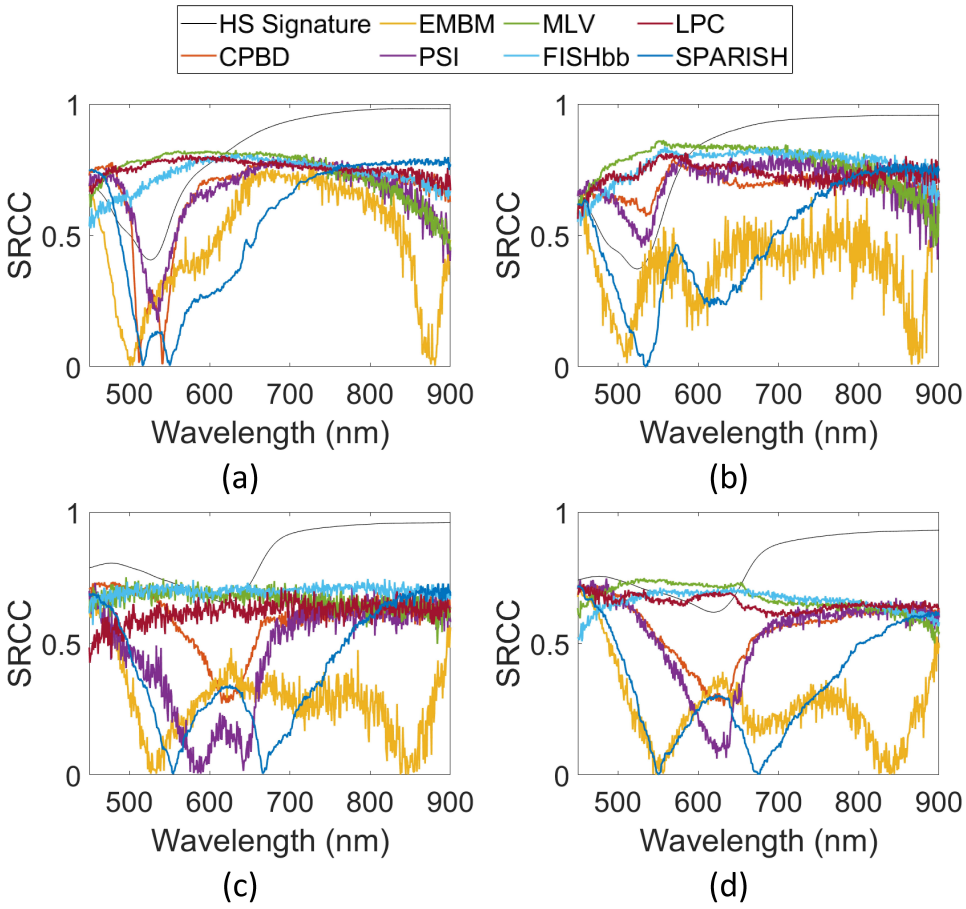


**Fig. 11.** SRCC correlation to $SS_{MOS}$ values for: (a) rat histology; (b) blood smears; and (c) leaf structure; and (d) stems.

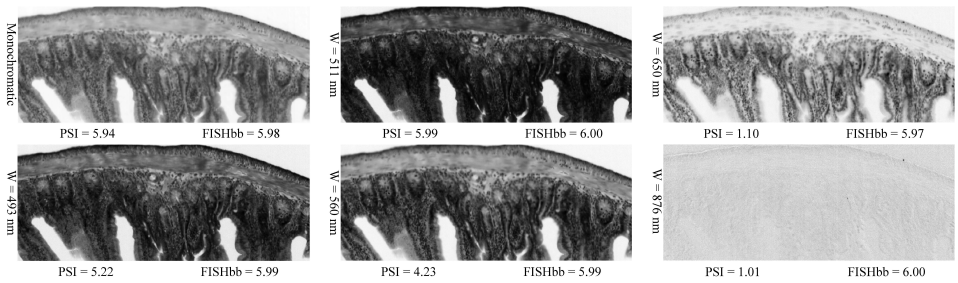


**Fig. 12.** Image of a rat intestine tissue at different wavelengths (W) and their corresponding *OS* values for the PSI and FISHbb algorithms, being $SS_{WD} = 6$.

Optics EXPRESS

Thus, we can conclude that G1 algorithms are sample independent while G2 methods are not. An example is found on Fig. 12, where different wavelength frames are shown next to their scores for one algorithm of each group. It can be appreciated that, although for the chromatic image PSI and FISHbb scores are similar, they vary along wavelengths. It is remarkable how FISHbb remains almost constant, although the spatial information is very different in each of the frames. Since all the spectrum of a HS cube is focused at the same WD, it would be coherent for all the bands to have the same score, making G1 algorithms more reliable.

## 5. Conclusion

Histopathology is the gold standard for cancer diagnosis. HSI is a novel technique that is showing high potential to improve the identification of tumor tissue and other diseases. High quality images are required to obtain an accurate automatic tissue classification. However, optimal focus is not always achieved due to the irregular surfaces of the samples. Thus, focus assessment becomes a relevant step before classifying histology images. In this paper, the state-of-the-art related to the existent *OSs* that quantify blurriness has been studied. Existing NR-IQA databases have also been reviewed.

One relevant contribution of this work is the creation of a public HS database for focus assessment based on 125 reference frames taken at 11 different WDs (a total of 1,375 HS cubes). $SS_{MOS}$ scores were obtained by asking 24 subjects to complete a subjective test, where images were rated following the Swiss tournament system. Afterwards, such *SSs* were correlated to the previously studied *OS*s. The evaluation of these techniques, applied to synthetic RGB and monochromatic images generated from the HS cubes, agreed that the algorithms with better performance are MLV, FISHbb and LPC ($SRCC>0.85$, $PLCC>0.8$ *and* $RMSE<1.8$). Then, OSs were also calculated over single band images, creating two groups of algorithms: the ones that obtain similar values for all the spectrum (G1: MLV, FISHbb and, LPC) and the ones that obtain different values (G2: CPBD, EMBM, PSI and, SPARISH). To select one group, the database was divided in the four origin groups (blood smears, leaf structures, rat histology and stems) and the HS signatures were compared to the values given by the algorithms. Since the entire spectrum of a HS cube is focused to the same WD, it would be coherent for all bands to have the same score. Thus, G1 algorithms should be selected since their score remain constant to information change along bands. Concluding, the most important contribution of this work is the finding of G1 algorithms to be more reliable for this application. Future work includes the use of these algorithms in z-stacking applications. The idea is to correct the blurriness, caused by different factors (Fig. 1) over the HS cubes. To this aim, the selected algorithms will be further studied to select the most focused areas and spectral bands over a range of HS cubes captured within the system's DOF.

**Funding.** Ministerio de Ciencia e Innovación (PID2020-116417RB-C42, TALENT-HExPERIA); Agencia Canaria de Investigación, Innovación y Sociedad de la Información (Conocimiento y Empleo, Consejería de Economía, Eje 3 Tema Prioritario 74 (85%), POC 2014-2020); Agencia Estatal de Investigación (European Union "NextGenerationEU/PRTR", FJC2020-043474-I, MCIN/AEI/10.13039/501100011033).

**Acknowledgments.** The authors would like to thank all twenty-four volunteers who participated in filling the MOS questionnaires for the subjective evaluation of the blurriness of the images of the dataset. This work was supported by the Spanish Government and European Union as part of the TALENT-HExPERIA (HypErsPEctRal Imaging for Artificial intelligence applications) project (PID2020-116417RB-C42). Moreover, this work was completed while Laura Quintana was beneficiary of the pre-doctoral grant given by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Economía, Conocimiento y Empleo", which is part-financed by the European Social Fund (FSE) (POC 2014-2020, Eje 3 Tema Prioritario 74 (85%)) and, Himar Fabelo was beneficiary of the FJC2020-043474-I funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** The datasets generated during the current study are available under reasonable request [14].

Optics EXPRESS

**Supplemental document.** See Supplement 1 for supporting content.

## References


1. L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," Comput. Methods Programs Biomed. **107**(3), 538–556 (2012).
2. G. Moallem, A. A. Pore, A. Gangadhar, H. Sari-Sarraf, and S. A. Vanapalli, "Detection of live breast cancer cells in bright-field microscopy images containing white blood cells by image analysis and deep learning," J. Biomed. Opt. **27**(07), 1 (2022).
3. A. Mora-Zuniga, S. Quiros-Barrantes, and F. Siles, "M-Phase Feature Extraction Algorithm for Phenotype Classification from Cancer Brightfield Microscopy," in *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)* (IEEE, 2018), pp. 1–8.
4. M. D. Zarella, D. Bowman, F. Aeffner, N. Farahani, A. Xthona, S. F. Absar, A. Parwani, M. Bui, and D. J. Hartman, "A Practical Guide to Whole Slide Imaging: A White Paper From the Digital Pathology Association," Arch Pathol Lab Med **143**(2), 222–234 (2019).
5. S. Ortega, M. Halicek, H. Fabelo, G. M. Callico, and B. Fei, "Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review," Biomed. Opt. Express **11**(6), 3195 (2020).
6. S. Ortega, M. Halicek, H. Fabelo, R. Guerra, C. Lopez, M. Lejeune, F. Godtliebsen, G. M. Callico, and B. Fei, "Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images," in *Medical Imaging 2020: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, eds., (SPIE, 2020), p. 30.
7. M. Halicek, S. Ortega, H. Fabelo, C. Lopez, M. Lejaune, G. M. Callico, and B. Fei, "Conditional generative adversarial network for synthesizing hyperspectral images of breast cancer cells from digitized histology," in *Medical Imaging 2020: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, eds., (SPIE, 2020), p. 29.
8. S. Ortega, R. Guerra, M. Diaz, H. Fabelo, S. Lopez, G. M. Callico, and R. Sarmiento, "Hyperspectral Push-Broom Microscope Development and Characterization," IEEE Access **7**, 122473–122491 (2019).
9. G. Wang and N. Fang, "Detecting and Tracking Nonfluorescent Nanoparticle Probes in Live Cells," in (2012), pp. 83–108.
10. L. Palmieri, G. Scrofani, N. Incardona, G. Saavedra, M. Martínez-Corral, and R. Koch, "Robust Depth Estimation for Light Field Microscopy," Sensors **19**(3), 500 (2019).
11. E. A. El-Gabry, A. V. Parwani, and L. Pantanowitz, "Whole-slide imaging: widening the scope of cytopathology," Diagn Histopathol **20**(12), 456–461 (2014).
12. C. Meng, P. An, X. Huang, C. Yang, L. Shen, and B. Wang, "Objective Quality Assessment of Lenslet Light Field Image Based on Focus Stack," IEEE Trans. Multimedia **24**, 3193–3207 (2021).
13. D. Li and T. Jiang, "Blur-Specific No-Reference Image Quality Assessment: A Classification and Review of Representative Methods," in (2019), pp. 45–68.
14. L. Quintana-Quintana, S. Ortega, H. Fabelo, F. J. Balea-Fernández, and G. M. Callico, "Hyperspectral Image Dataset for Focus Analysis (HIDFA)," https://hsidatabase.iuma.ulpgc.es/.
15. Elizabeth Allen and Sophie Triantaphillidou, *The Manual of Photography* (2011).
16. N. Burningham, Z. Pizlo, and J. P. Allebach, "Image Quality Metrics," in *Encyclopedia of Imaging Science and Technology* (John Wiley & Sons, Inc., 2002).
17. S. Ortega, M. Halicek, H. Fabelo, R. Camacho, M. de la, L. Plaza, F. Godtliebsen, G. M. Callicó, and B. Fei, "Hyperspectral Imaging for the Detection of Glioblastoma Tumor Cells in H&E Slides Using Convolutional Neural Networks," Sensors **20**(7), 1911 (2020).
18. A. Hore and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition* (IEEE, 2010), pp. 2366–2369.
19. Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang, "A Psychovisual Quality Metric in Free-Energy Principle," IEEE Trans. on Image Process. **21**(1), 41–52 (2012).
20. C. Feichtenhofer, H. Fassold, and P. Schallauer, "A Perceptual Image Sharpness Metric Based on Local Edge Gradient Analysis," IEEE Signal Process. Lett. **20**(4), 379–382 (2013).
21. N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Signal Process Image Commun **30**, 57–77 (2015).
22. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).
23. D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," J. Electron. Imaging **19**(1), 011006 (2010).
24. S. le Moan, S. George, M. Pedersen, J. Blahová, and J. Y. Hardeberg, "A database for spectral image quality," in M.-C. Larabi and S. Triantaphillidou, eds. (2015), p. 93960P.
25. R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," Multimed Syst **22**(2), 213–227 (2016).
26. J. R. Lewis, "Psychometric Properties of the Mean Opinion Scale," Proceedings of HCI International **1**, 149–153 (2001).
27. ITU-T Recommendation P.911, *Subjective Audiovisual Quality Assessment Methods for Multimedia Applications* (1998).

Optics EXPRESS


28. Rec. ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," (2002).
29. S. Jumisko-Pyykkö and J. Häkkinen, "Profiles of the Evaluators – Impact of Psychographic Variables on the Consumer-oriented Quality Assessment of Mobile Television," Multimedia on Mobile Devices **6821**, 68210L (2008).
30. F. Speranza, F. Poulin, R. Renaud, M. Caron, and J. Dupras, "Objective and subjective quality assessment with expert and non-expert viewers," in *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)* (IEEE, 2010), pp. 46–51.
31. O. Köster, M. Jessen, F. Khairi, and H. Eckert, "Auditory-perceptual identification of voice quality by expert and non-expert listeners," Proceedings of the XVI International Congress of the Phonetic Sciences (ICPhS) **1**, 1845–1848 (2007).
32. H. Choi, "Subjective video quality comparison using various displays," Opt. Eng. **48**(3), 037002 (2009).
33. "Methodologies for the subjective assessment of the quality of television images," 103 (2019).
34. G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychol Rev **63**(2), 81–97 (1956).
35. N. D. Narvekar and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," IEEE Transactions on Image Processing **20**(9), 2678–2683 (2011).
36. J. Guan, W. Zhang, J. Gu, and H. Ren, "No-reference blur assessment based on edge modeling," J Vis Commun Image Represent **29**, 1–7 (2015).
37. Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang, "No-Reference Image Sharpness Assessment in Autoregressive Parameter Space," IEEE Trans. on Image Process. **24**(10), 3218–3231 (2015).
38. K. Bahrami and A. C. Kot, "A Fast Approach for No-Reference Image Sharpness Assessment Based on Maximum Local Variation," IEEE Signal Process. Lett. **21**(6), 751–755 (2014).
39. P. V. Vu and D. M. Chandler, "A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation," IEEE Signal Process. Lett. **19**(7), 423–426 (2012).
40. R. Hassen, Zhou Wang, and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence," IEEE Trans. on Image Process. **22**(7), 2798–2810 (2013).
41. L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments," IEEE Trans. Cybern. **46**(1), 39–50 (2016).
42. C. T. Vu, T. D. Phan, and D. M. Chandler, "S3: A Spectral and Spatial Measure of Local Perceived Sharpness in Natural Images," IEEE Trans. on Image Process. **21**(3), 934–945 (2012).
43. L. Li, D. Wu, J. Wu, H. Li, W. Lin, and A. C. Kot, "Image Sharpness Assessment by Sparse Representation," IEEE Trans. Multimedia **18**(6), 1085–1097 (2016).
44. H. Fabelo, R. Leon, S. Ortega, F. Balea-Fernandez, C. Bilbao, G. M. Callico, and A. Wagner, "Novel Methodology for Alzheimer's Disease Biomarker Identification in Plasma using Hyperspectral Microscopy," in *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)* (2020).
45. M. Perez-Ortiz, A. Mikhailiuk, E. Zerman, V. Hulusic, G. Valenzise, and R. K. Mantiuk, "From Pairwise Comparisons and Rating to a Unified Quality Scale," IEEE Trans. on Image Process. **29**, 1139–1151 (2020).
46. H. Fabelo, S. Ortega, and A. Szolna, *et al.*, "In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection," IEEE Access **7**, 39098–39116 (2019).
47. L. Csató, "On the ranking of a Swiss system chess team tournament," Ann Oper Res **254**(1-2), 17–36 (2017).
48. VQEG, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment (2000).
49. P. Mishra, U. Singh, C. Pandey, P. Mishra, and G. Pandey, "Application of student's t-test, analysis of variance, and covariance," Ann Card Anaesth **22**(4), 407 (2019).
50. L. Giannoni, F. Lange, and I. Tachtsidis, "Hyperspectral imaging solutions for brain tissue metabolic and hemodynamic monitoring: past, current and future developments," J. Opt. **20**(4), 044009 (2018).