https://doi.org/10.2352/issn.2168-3204.2024.21.1.19

Evaluation of Binarization Methods for Hyperspectral Samples of 16th and 17th Century Family Trees

Francisco Moronta-Montero[1,*], Ramón Fernández-Gualda[1], Ana B. López-Baldomero[1], Marco Buzzelli[2], Miguel A. Martínez-Domingo[1], Eva M. Valero[1].

[1]: Department of Optics; University of Granada; Granada, Spain.
[2]: Department of Informatics Systems and Communication; University of Milano - Bicocca; Milan, Italy.
****: Campus Fuentenueva, s/n 18071 Granada, Spain; fmoronta@ugr.es***

Abstract

The purpose of this work is to present a new dataset of hyperspectral images of historical documents consisting of 66 historical family tree samples from the 16th and 17th centuries in two spectral ranges: VNIR (400-1000 nm) and SWIR (900-1700 nm). In addition, we performed an evaluation of different binarization algorithms, both using a single spectral band and generating false RGB images from the hyperspectral cube.

Introduction

Historical documents are profound testimonies of the history of human civilization and contain the knowledge and cultural heritage of past times [1]. Their preservation and understanding are essential to understand the traditions and evolution of societies over the centuries. The study of historical documents goes beyond archival preservation, as it involves a multidisciplinary study between history, archaeology, art, and technology.

These documents often suffer the damage caused over time, like ink fading, paper degradation, environmental exposure, and physical damage, which affects their legibility and preservation [2]. To overcome these problems, new approaches have emerged with cutting-edge non-invasive technologies that are transforming the field of heritage science.

One of these tools is hyperspectral imaging, that unlike conventional imaging, which captures data in three broad spectral bands (such as red, green and blue), divides the electromagnetic spectrum into numerous narrow bands [3]. This technique facilitates the non-invasive acquisition of spectral reflectance from every fragment of a historical document, which depends on the component materials such as inks [4], pigments [5], binders and substrate, and the physical interaction between them.

In most documents, the majority of the pixels belong to the substrate. Then, it is interesting to consider, as data preprocessing, the separation between the background (substrate, such as parchment or paper) and the foreground (text or illustrations) [6], referred to as binarization [7]. In this sense, spatial image segmentation techniques play a key role in this step [7], allowing us to differentiate between background and foreground, which can be the first step towards identifying the materials present and digitizing historical documents for preservation.

Classic binarization techniques employ thresholding methods, such as global thresholding and local thresholding [8]. Global thresholding applies a single threshold value to the entire image, while local thresholding adapts threshold values to different regions within the image. These methods have their advantages and limitations, particularly when dealing with documents that exhibit varying ink shades, background complexities, or degradation due to aging [9].

In this regard, several competitions have been held in the field of document image binarization [10-18] to test different proposed algorithms, most of them with RGB image datasets and one with multispectral images [19], using 8 bands from 340 to 1100 nm.

Our objective is to generate a dataset of hyperspectral images and use it to evaluate different methods of binarization of historical documents. For this purpose, we captured different documents from the family tree collection that is preserved in the Archive of the Royal Chancery of Granada. All the information about the samples was obtained from an unpublished internal report that documented a previous restoration work [20].

Materials and Methods

Dataset capture and processing

The documents from which our dataset has been generated consist of a series of family trees from the 16th and 17th centuries. A large percentage of these documents are handwritten, although there are several stamped fragments [20].

All the samples have the same support: rag paper, and in previous analyses it has been possible to determine the presence of two different inks, one of them carbon black and the other consisting of a mixture of sepia and iron gall [20].

For each document, two hyperspectral images have been obtained. The first one was captured by a Resonon Pika L device [21] in the visible and near infrared range (VNIR, 121 bands from 400 to 1000 nm), with a resolution of 900 pixels per line. The second image was obtained using a Resonon Pika IR+ [22] in the short wavelength infrared (SWIR, 161 bands from 900 to 1700 nm), with a spatial resolution of 640 pixels per line. The lighting setup consisted of four halogen lamps positioned to prevent specular reflection from the samples. The spectral data was flat-field corrected, and dark signal noise was subtracted from them to obtain the spectral reflectance of the samples.

Once the hyperspectral cubes were obtained, 33 regions of interest were chosen to retrieve their spectral reflectance in both VNIR and SWIR ranges, finally obtaining the 66 samples that constitute our dataset. Every sample consists of the full hyperspectral cube of the fragment and its metadata, that contains information about the device and illumination used for capture, reference white and a description of the known present materials. It also contains a false RGB image generated by choosing the 605, 535 and 430 nm bands for the VNIR range and the 1600, 1200 and 1000 nm bands for the SWIR range. Finally, each fragment also contains its Ground Truth (GT) within the metadata as a logical matrix, separating every ink pixel (1) from every background pixel (0). These Ground Truths were generated by a semi-automatic procedure based on [23].

The fragments range in size from 67x52 pixels for the smallest to 180x195 pixels for the largest, resulting in a total of 570285 pixels in the entire dataset, 311563 for the VNIR spectral range and 258722 for SWIR. The difference between the two sets is due to the fact that two of the fragments could not be registered to obtain exactly the same sample in VNIR and SWIR due to the low contrast of the latter range, so two similar zones were chosen.

This dataset is a part of the Hyperdoc project's hyperspectral image database, which will be progressively released at [24] along 2024, and contains both real historical documents and mock-up samples.

Segmentation algorithms

Five different binarization methods (Otsu [25], Niblack [26], Wolf [27], Bradley [28] and a Deep-Learning based algorithm) have been evaluated with our family trees dataset. For this purpose, instead of binarizing a RGB image converted to grayscale, which could potentially be detrimental to the final results when using out-of-focus and noisy bands, a single band representing the best image quality of the whole hyperspectral cube has been selected. To achieve this, we performed preliminary tests with different approaches for the selection of the best band, and they suggested that the best method for our particular case was to employ a Signal-to-Noise Ratio (SNR) metric to identify the channel within the hyperspectral image that exhibits the lowest noise level. SNR is defined as [29]:

$$SNR(\lambda) = 10 \cdot log_{10}\left(\frac{\mu^2}{\sigma^2}\right)$$

where μ represents the average intensity of the whole image (signal), σ is the standard deviation (noise) and λ represents each wavelength band. We choose the band that minimizes the value of this Signal-to-Noise Ratio, since we are interested in obtaining the band with the best contrast for binarization and therefore we want the standard deviation of the whole image to be high.

Several binarization methods have been presented over the last decades [6], each with its distinct approach to threshold determination and image segmentation. However, for our specific analysis, we will focus on utilizing four of the most prominent techniques:

Otsu Method

Otsu's global thresholding algorithm [25] is a widely used technique for automatic image segmentation into foreground and background [6]. This method operates by analyzing the histogram of pixel intensities, which is divided into two clusters by a chosen threshold T. The optimal threshold T_{Otsu} is calculated by minimizing the within-cluster variance, given by:

$$T_{Otsu} = argmin\, w_0(T) \cdot \sigma_0^2(T) + w_1(T) \cdot \sigma_1^2(T)$$

where 0 and 1 refer to the first and second cluster, $w_n(T)$ is the number of pixels in the cluster n and $\sigma_n^2(T)$ is the variance of the cluster n. Once the threshold is calculated, binarization is performed by taking pixels with an intensity greater than T_{Otsu} as background and those with a lower intensity as foreground.

Niblack Method

Niblack introduced a local adaptive thresholding method [26] where each pixel's threshold is determined based on statistics for a local window centered on that particular pixel, which allows for potential handling of cases involving overlap in foreground and background intensity distributions. This per-pixel Niblack threshold T is calculated as:

$$T = \mu + k \cdot \sigma$$

where μ and σ are the local mean and standard deviation calculated for each window and k is a parameter set by the user that for our case has been chosen as -0.2, as recommended in [6]. The window size chosen for this method is 1/3 of the image width by 1/3 of the image height, with borders handled by using replicative padding.

Wolf Method

The Wolf algorithm [27] is an adaptation of Sauvola and Niblack methods designed to optimize its performance in scenarios involving low-contrast images. The algorithm employs a sliding window approach to dynamically adjust the threshold for each pixel, enhancing its adaptability to different images. The threshold T for a given pixel within the sliding window is determined by:

$$T = \mu - k \cdot \left(1 - \frac{\sigma}{S}\right) \cdot (\mu - M)$$

where μ is the local mean intensity within the window, k is an empirically chosen parameter (set to 0.5 for our study), σ is the local standard deviation of pixel intensities, S is the maximum value of σ within the window and M represents the minimum intensity value within the window. For window size, we use the same dimensions and edge treatment as utilized with Niblack's method.

Bradley Method

Bradley's Local Image Thresholding [28] is a simple and effective method for segmentation based on local adaptive thresholding. This method addresses variations in contrast across different regions of the image by setting a dynamic threshold based on the local mean intensity. This threshold is determined by:

$$T = \mu \cdot \left(1 - \frac{t}{100}\right)$$

where T is the threshold value for a pixel, μ is the local mean intensity within a specified window centered at the pixel and t is a

user-defined parameter representing the percentage of intensity values to be considered as foreground. This method introduces adaptability through the parameter t, allowing users to control the sensitivity of the threshold to local variations in pixel intensities. For this study, t is set to 10, and we use the same window size and border treatment than the previous methods.

Deep-Learning based algorithm

We will also compare the binarization results with a Deep-Learning algorithm based on DeepLabV3 [30], which we have chosen for its good performance in semantic segmentation tasks, making it a suitable candidate for historical documents binarization. As DeepLab is designed for RGB image inputs, we have adapted it to work with the full hyperspectral images. Detailed information about the algorithm and its evaluation with a different dataset than the one used in this work is in the process of being published.

Evaluation metrics

Three different evaluation metrics have been chosen to test the performance of each algorithm, Pseudo F-Measure, Peak Signal-to-Noise Ratio and Distance Reciprocal Distortion, as they have been widely used in binarization contests like DIBCO [10-18].

Pseudo F-Measure (pFM)

Pseudo F-Measure is an extension of F-Measure that was introduced in [31]. This metric goes from 0% to 100% in the case of two identical images, and it is given by:

$$pFM = 2 \cdot \frac{R_{ps} \cdot P_{ps}}{R_{ps} + P_{ps}}$$

where R_{ps} is the pseudo-Recall and P_{ps} the pseudo-Precision.

To define pseudo-Recall, the foreground GT image is weighted by assigning distance values to pixels based on their proximity to the contour, and pseudo-Precision introduces a weighting mechanism for the background of the GT, which is determined by the distance from the pixel to the contour of the text, mitigating the excessive penalization of false positive pixels located far from it.

Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio measures how close an image is to another one [32]. A greater PSNR value indicates a stronger similarity between them and tends to infinity in the case of two identical images. We can calculate this metric by the following expression:

$$PSNR = 10 \cdot log_{10}\left(\frac{C^2}{MSE^2}\right)$$

where

$$MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left(I(x,y) - I'(x,y)\right)^2}{M \cdot N}$$

C is the highest possible value of the signal, that in our case equals to 1. $I(x,y)$ and $I'(x,y)$ are the intensity values on the pixel (x,y) of the prediction and the Ground Truth. Finally, M and N are the width and height of the image respectively.

Distance Reciprocal Distortion Metric (DRD)

The Distance Reciprocal Distortion Metric was introduced to improve PSNR, since it does not fit subjective evaluations, as it does not consider inter-pixel dependencies [33]. Two predictions having the same number of flipped (wrong) pixels but in different positions would have the same PSNR, so DRD is proposed as a technique to determine which one looks better for the human eye. The lower the DRD, the more similar the images are, and it can be calculated as:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN}$$

where S is the total number of flipped pixels, $NUBN$ represents the count of non-uniform (containing pixels that are not all black or white) 8x8 blocks within the GT, and DRD_k is the k-th flipped pixel distortion, defined in [33].

Results

All 66 samples have been binarized using the five algorithms, separating the VNIR and SWIR ranges, so that for each metric we obtain the result by averaging those of the 33 samples from each subset. Apart from this quantitative evaluation, we performed a visual test to find out if by increasing the number of bands considered in the SWIR range we can also obtain as a foreground the bleed-through of the back page.

Single band image binarization

Table 1. Binarization results for each method in the VNIR range. The value for each metric corresponds to the average of the results of the entire VNIR subset and the standard deviation is used as a margin of error.

VNIR			
Method	pFM	PSNR	DRD
Bradley	97.5 ± 1.4	15.2 ± 2.6	2.4 ± 2.3
Wolf	97.2 ± 1.3	14.8 ± 2.2	2.2 ± 1.0
Otsu	96.9 ± 2.2	14.4 ± 2.4	2.6 ± 1.3
Niblack	96.6 ± 2.5	13.4 ± 2.6	2.9 ± 1.5
Deep Learning	93.6 ± 2.3	11.6 ± 1.8	3.8 ± 1.4

Table 2. Binarization results for each method in the SWIR range. The value for each metric corresponds to the average of the results of the entire SWIR subset and the standard deviation is used as a margin of error.

SWIR			
Method	pFM	PSNR	DRD
Bradley	92.9 ± 3.4	10.2 ± 2.3	5.5 ± 2.2
Wolf	91.7 ± 3.8	10.2 ± 2.3	5.7 ± 2.5
Otsu	90.6 ± 4.3	10.1 ± 2.4	7.5 ± 9.8
Niblack	91.1 ± 4.2	9.2 ± 2.6	7.2 ± 3.0
Deep Learning	91.0 ± 3.1	9.3 ± 1.8	6.6 ± 2.1

Table 1 shows the metrics obtained after binarization of our dataset with the different algorithms in the VNIR range, while Table 2 shows the results for the SWIR range. In general, all algorithms perform better in the VNIR range, which makes sense given that SWIR images tend to have lower contrast. The presence of iron gall ink also poses a problem for binarization in the SWIR range, since it has a high reflectance at infrared wavelengths and therefore fades from the image [34], being indistinguishable from the also highly reflective background.

For both spectral ranges, Bradley outperforms all other methods used, being the highest rated for all metrics except DRD in VNIR and being closely followed by Wolf. As for Otsu and Niblack, the former performs better in the VNIR range while the latter performs better in SWIR according to pFM and DRD. This can also be explained by the low contrast and high presence of noise in the SWIR range, since Otsu's method generates a threshold for the whole image while Niblack, Wolf and Bradley avoid this problem by averaging over different windows. Finally, the Deep Learning-based algorithm does not appear to be on par with the two top-scoring methods, falling behind Otsu and Niblack in the VNIR range. This issue may be due to the deep learning algorithm having problems with the low spatial resolution of the images. When using Transfer Learning, the algorithm is conditioned by the previous training, where there are not many figures with very fine details. Because of this, we are currently working on alternative methods to overcome this situation.

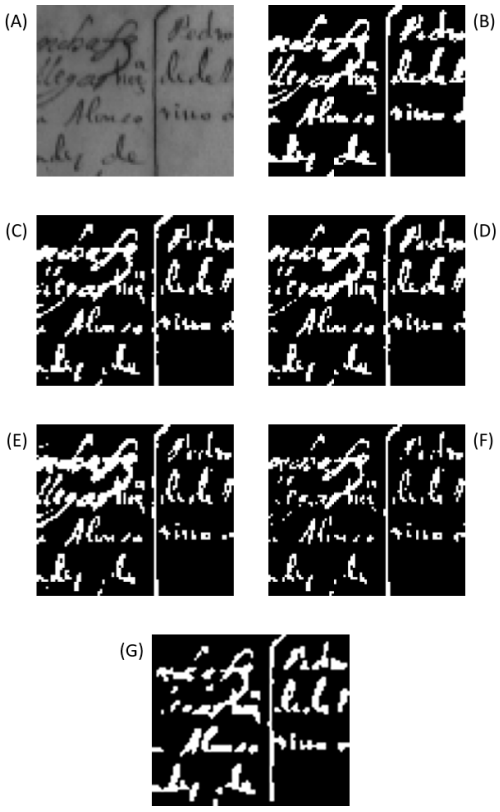

Figure 1. *Binarization results for a sample in the VNIR range. (A) Best channel image (440 nm). (B) Ground Truth. (C) Bradley method result. (D) Wolf method result. (E) Otsu method result. (F) Niblack method result. (G) Deep Learning algorithm result.*

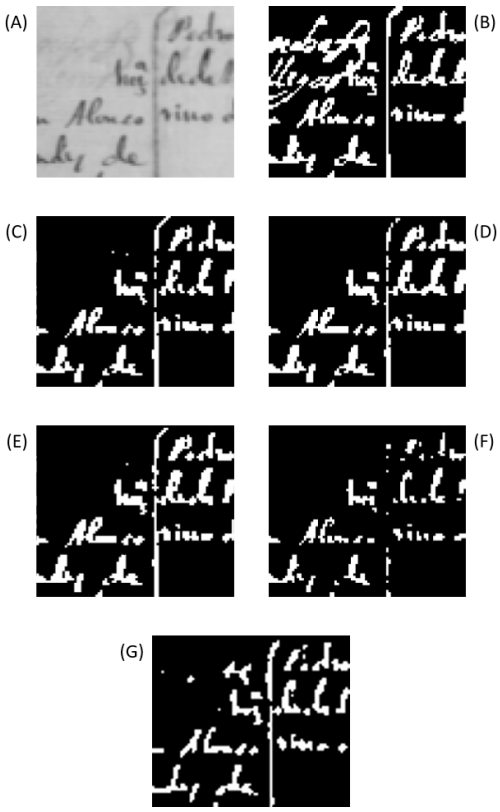

Figure 2. *Binarization results for a sample in the SWIR range. (A) Best channel image (975 nm). (B) Ground Truth. (C) Bradley method result. (D) Wolf method result. (E) Otsu method result. (F) Niblack method result. (G) Deep Learning algorithm result.*

Figures 1 and 2 show the same fragment of the family trees in both VNIR and SWIR ranges along with its Ground Truth and the results for each algorithm. The bands with the lowest signal-to-noise ratio that have been chosen for binarization clearly show the difference between the two spectral ranges. All algorithms offer very similar results, but we can see that the algorithms with the worst results for the metrics in Tables 1 and 2 show more irregularities and gaps in the letters.

While in VNIR there is no difference between the ink used in the upper left corner and the one used in the rest of the fragment, the SWIR range shows that they are actually two different inks, with the first one disappearing almost completely, which makes the binarization task much more difficult, but at the same time, can be very informative for material identification. Thus, if we want to differentiate which areas contain metallo-galate inks and which contain carbon-based inks, we only need to compare the results for both ranges and look for clear differences.

In the SWIR range, none of the algorithms was able to correctly identify the presence of the iron gall ink. The method that correctly identifies the most pixels in this area is the Deep Learning algorithm, but it is still far from identifying the full text as it appears in Ground Truth. All of them also worsen their results for other ink in comparison to VNIR range, with gaps appearing in the straight line in the center and missing pieces of some letters.

False RGB image binarization

There is also a problem that we have not considered so far: if we want to use this binarization to make a precise identification of all the inks present in the image, it would be necessary to also include the bleed-through as foreground, since it is also an ink that could be the same or different from those used in the page we are studying and it could also be interesting to know its composition. We have found that our employed technique for band selection does not give us an adequate wavelength to meet this objective, since in several fragments bleed-through areas are clearly observed in the false RGB image of the SWIR range that are not present in the band with the best signal-to-noise ratio. This is because, in the SWIR range, different inks may exhibit more transparency at some wavelengths than at others, as we have seen in the case of iron gall, so using only one band may not give us the full information of the inks present.

Therefore, the simplest way to take multiple bands into account is to generate false RGB images with three different wavelengths and combine them into a single grayscale image for binarization:

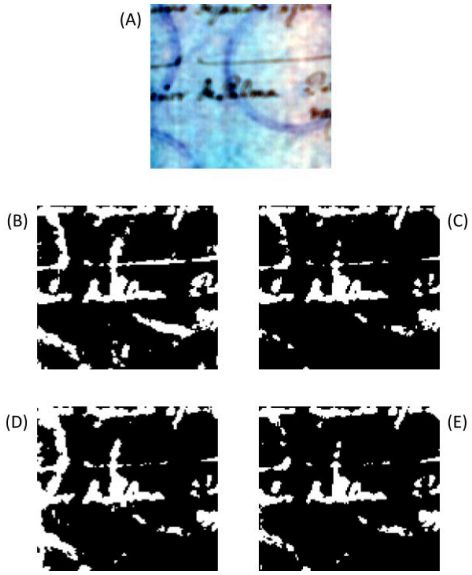

Figure 3. *Binarization results for a sample in the SWIR range. (A) False RGB image ([1600, 1200, 1000] nm). (B) Bradley method result. (C) Wolf method result. (D) Otsu method result. (E) Niblack method result.*

Figure 3 shows the false RGB image for a fragment of the dataset in the SWIR range, where we can clearly observe stamped circles coming from the back side of the page.

For the case shown in Figure 3, we consider that the SWIR range introduces too much noise by choosing three bands that are not optimal and are out of focus, so it would not be fair to evaluate the algorithms with the Ground Truth generated for the single-band binarizations. For the same reason, the Deep Learning based algorithm has not been included in this section, since it has been trained using the full hyperspectral cubes and the single-band Ground Truths, so it is biased to not identify bleed-through as ink.

None of the binarization algorithms used could select the complete circle as foreground, only fragments of it. Among all of them, Bradley seems to have obtained the best results for the bleed-through circles, but at the cost of selecting several noise pixels, especially in the lower part of the fragment. Wolf's algorithm loses certain areas of the circles but gets the fewest false positives in areas where there is no ink. In the case of Otsu, a large area on the left side is incorrectly identified as a foreground, while Niblack offers a very similar result to Wolf.

Conclusions

In this work we have presented a new dataset of 66 hyperspectral images of historical documents preserved in the Archive of the Royal Chancery of Granada, with 33 different samples captured in the SWIR and VNIR spectral ranges, all of them with metadata containing information on the wavelengths measured, the illumination and capture devices used to take the sample and the reference target with which the correction has been made, as well as the corresponding Ground Truth.

These spectral captures have been used with the aim of finding the best bands to achieve the binarization of the areas with ink, evaluating four different algorithms with which good results have been achieved for both ranges, being those of VNIR better than those of SWIR. This last conclusion is logical given that in the second range the image has less contrast due to the increased reflectance of the inks in the short wavelength infrared, becoming as reflective as paper in the case of the iron gall ink, which disappears almost completely, and the algorithms find it impossible to identify it as a foreground. In this aspect, the only advantage observed for the deep learning based algorithm is that it manages to identify slightly better the iron gall areas in the SWIR range, which makes sense because it uses all the spectral information of the fragment instead of only one band in which the ink may not appear.

We observe a similar problem in the bleed-through case, where we can see that if we want to correctly identify these areas as ink in the SWIR range, it is not enough to obtain the best band in terms of noise, and it is convenient to use different wavelengths in a single false RGB image to increase the contrast of these areas.

From these results, we can conclude that, for our data set, the Bradley method outperforms all others in the case of binarization of the lowest noise band, a case in which the Niblack algorithm performs the worst. When working with the SWIR images in false RGB, Wolf's algorithm seems to perform better than the others, although the same quantitative evaluation has not been performed as for the previous case. This information can be useful if we want to use these binarization results as a first step for other work, such as optical character recognition (OCR) or material identification.

It is important to address the limitations of our study, given that we are working with a very specific dataset in which all samples have the same substrate and contain only two different inks. This work is just a glimpse into the possibilities of hyperspectral imaging to digitize historical documents with all the information of their spectral reflectance, so that it can be used for future preservation tasks. The differentiation between both spectral ranges has allowed us to discriminate, only by using binarization, between metallo-gallate inks and carbon-based inks, so this opens the future possibility of using hyperspectral imaging for a complete identification of the materials present in each zone of the document in more complex cases.

The creation of the Hyperdoc project's hyperspectral database [24], in which this dataset is included, is the first step to achieve this goal, containing over a thousand hyperspectral samples from both historical and synthetic modern documents with multiple substrates, inks and pigments.

Acknowledgements

This work was partially supported by MCIN/AEI/10.13039/501100011033, by “ERDF A way of making Europe” and "ESF Investing in your future" [grant numbers PID2021-124446NB-I00 and PRE2022-101352], and by Ministry of Universities (Spain) [grant number FPU2020-05532].

This work was partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy

References

[1] I. Ciortan, H. Deborah, S. George and J.Y. Hardeberg, "Color and hyperspectral image segmentation for historical documents". In 2015 Digital Heritage, Vol. 1, pp. 199-206 (2015).

[2] E. Kavallieratou and H. Antonopoulou. "Cleaning and enhancing historical document images". In Advanced Concepts for Intelligent Vision Systems: 7th International Conference, ACIVS 2005. Proceedings 7, pages 681–688. Springer (2005).

[3] M. Diem, F. Hollaus and R. Sablatnig, "MSIO: MultiSpectral Document Image BinarizatIOn". In 2016 12th IAPR Workshop on Document Analysis Systems (DAS), pp. 84-89 (2016).

[4] Z. Khan, F. Shafait and A. Mian, "Hyperspectral imaging for ink mismatch detection." In 2013 12th International Conference on Document Analysis and Recognition, pp. 877–881, IEEE (2013).

[5] E.M. Valero, M.A Martínez-Domingo, A. B. López-Baldomero, A. López-Montes, D. Abad-Muñoz, and J.L. Vílchez-Quero, "Unmixing and pigment identification using visible and short-wavelength infrared: Reflectance vs logarithm reflectance hyperspaces". In Journal of Cultural Heritage, 64:290–300 (2023).

[6] C. Tensmeyer, T. Martinez, "Historical Document Image Binarization: A Review". In SN Computer Science, 1, 173 (2020).

[7] W. Xiong, L. Zhou, L. Yue, L. Li and S. Wang, "An enhanced binarization framework for degraded historical document images", In EURASIP Journal on Image and Video Processing, 1, 13 (2021).

[8] S. Chauhan, E. Sharma and A. Doegar, "Binarization techniques for degraded document images – A review". In 2016 5th international conference on reliability, infocom technologies and optimization (Trends and Future Directions) (ICRITO), pp. 163-166 (2016).

[9] C.A.B. Mello, R.D. Lins, "Image segmentation of historical documents". In Visual2000, Mexico City, 30 (2000).

[10] B. Gatos, K. Ntirogiannis and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)". In 2009 10th International Conference on Document Analysis and Recognition, pp. 1375-1382 (2009).

[11] I. Pratikakis, B. Gatos and K. Ntirogiannis, "H-DIBCO 2010 - Handwritten Document Image Binarization Competition". In 2010 12th International Conference on Frontiers in Handwriting Recognition, pp. 727-732 (2010).

[12] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)". In 2011 International Conference on Document Analysis and Recognition, pp. 1506-1510 (2011).

[13] I. Pratikakis, B. Gatos, K. Ntirogiannis, "ICFHR 2012 Competition on Handwritten Document Image Binarization (H-DIBCO 2012)". In 2012 International Conference on Frontiers in Handwriting Recognition, pp. 817–822 (2012).

[14] I. Pratikakis, B. Gatos, K. Ntirogiannis, "ICDAR 2013 Document Image Binarization Contest (DIBCO 2013) ". In 2013 12th International Conference on Document Analysis and Recognition, pp. 1471–1476 (2013).

[15] K. Ntirogiannis, B. Gatos, I. Pratikakis, "ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014) ". In 2014 14th International Conference on Frontiers in Handwriting Recognition, pp. 809–813 (2014).

[16] I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, "ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016)". In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 619–623 (2016).

[17] I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, "ICDAR2017 Competition on Document Image Binarization (DIBCO 2017)". In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), pp. 1395–1403 (2017).

[18] I. Pratikakis, K. Zagori, P. Kaddas, B. Gatos, "ICFHR 2018 Competition on Handwritten Document Image Binarization (H-

DIBCO 2018)". In 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 489–493 (2018).
[19] R. Hedjam, H. Z. Nafchi, R. F. Moghaddam, M. Kalacska and M. Cheriet, "ICDAR 2015 contest on MultiSpectral Text Extraction (MS-TEx 2015)". In 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, pp. 1181-1185 (2015).
[20] I. Galán, T. Espejo, "Proyecto de Restauración de la Colección de Árboles Genealógicos del Archivo de la Real Chancillería de Granada". Final Degree Project, University of Granada (2005).
[21] Resonon Inc. Resonon Pika L: https://resonon.com/Pika-L
[22] Resonon Inc. Resonon Pika IR+: https://resonon.com/Pika-IR-Plus
[23] K. Ntirogiannis, B. Gatos and I. Pratikakis, "An Objective Evaluation Methodology for Document Image Binarization Techniques". In The Eighth IAPR International Workshop on Document Analysis Systems, pp. 217-224 (2008).
[24] "HYPERDOC project", https://colorimaginglab.ugr.es/hyperdoc
[25] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms". In IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66 (1979).
[26] W. Niblack, "An Introduction to Digital Image Processing". Strandberg (1985).
[27] C. Wolf, J.-M. Jolion, F. Chassaing, "Text Localization, Enhancement and Binarization". In Multimedia Documents, 2 (2002).
[28] D. Bradley, G. Roth, "Adaptive Thresholding using the Integral Image". In J. Graphics Tools. 12, pp. 13-21 (2007).
[29] D. H. Johnson, "Signal-to-noise ratio". In Scholarpedia, 1(12), 2088 (2006).
[30] LC. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". In Proceedings of the European conference on computer vision (ECCV), pp. 801-818 (2018).
[31] K. Ntirogiannis, B. Gatos and I. Pratikakis, "Performance Evaluation Methodology for Historical Document Image Binarization". In IEEE Transactions on Image Processing, vol 22, no. 2, pp. 595-609 (2013).
[32] F.A. Fardo, V.H. Conforto, F.C. de Oliveira, P.S.S. Rodrigues, "A Formal Evaluation of PSNR as Quality Measurement Parameter for Image Segmentation Algorithms". In ArXiv (2016).
[33] Haiping Lu, A. C. Kot and Y. Q. Shi, "Distance-reciprocal distortion measure for binary document images". In IEEE Signal Processing Letters, vol. 11, no. 2, pp. 228-231 (2004).
[34] A. B. López-Baldomero, M. A. Martínez-Domingo, E. M. Valero, R. Fernández-Gualda, A. López-Montes, R. Blanc-García and T. Espejo, "Selection of optimal spectral metrics for classification of inks in historical documents using hyperspectral imaging data". In Optics for Arts, Architecture, and Archaeology (O3A) IX, volume 12620, pp. 99-111. SPIE (2023).

Author Biography

Francisco Moronta-Montero *is a Ph.D. student at the Color Imaging Lab (University of Granada). He graduated in Physics in 2021 at the University of Granada with a master's degree in Laser Physics and Technology in 2023 at the University of Salamanca. His research interests are spectral imaging and machine learning applications in Optics.*

Ramón Fernández Gualda *is a Ph.D. Student at the Color Imaging Lab (University of Granada). He is graduated in Telecommunications Engineering and a degree in Optics and Optometry in 2022, both at the University of Granada, where he also obtained the Official Master in Telecommunications Engineering. His research interests are Color Vision Deficiency, Spectral Image Processing, Deep Learning and Computer Vision.*

Ana B. López-Baldomero *is a Ph.D. candidate at the Color Imaging Lab (University of Granada). She obtained her B.Sc. in Optics and Optometry in 2019 at the University of Granada, where she also obtained both her M.Sc. in Clinical Optometry and Advanced Optics in 2020, and her M.Sc. in Physics in 2021. Currently her research focuses on the acquisition and analysis of hyperspectral images to identify materials in historical documents.*

Marco Buzzelli *obtained his PhD in Computer Science in 2019 at the University of Milano – Bicocca (Italy), where he currently serves as assistant professor at the Department of Informatics, Systems and Communication. He is actively engaged in conducting cutting-edge research in the field of signal/image/video processing and understanding, using machine learning techniques. He is particularly passionate about color imaging. Marco Buzzelli has collaborated with several European institutions, and he is an active ELLIS member.*

Miguel Ángel Martínez-Domingo *obtained his PhD. in Physics and Space Science in 2017. He is graduated in Telecommunications Engineering and Optics. He is currently Assistant professor at the Department of Optics of University of Granada and member of the Color Imaging Lab. His research interests are focused on spectral imaging applied to cultural heritage, color vision and food inspection among others.*

Eva M. Valero *obtained a B.D. in Physics in 1995, and a Ph.D. in 2000, both at the University of Granada. She has worked at the Department of Optics as Assistant Prof. from 2001 to 2007 and as Associate Prof. from 2007. She's a member of the Color Imaging Lab at the University of Granada. Her research interests were initially spatial color vision, and more recently multispectral imaging and color image processing.*