

# Multispectral Image Restoration of Historical Documents Based on LAAMs and Mathematical Morphology

Edwin Lechuga-S.,<sup>a</sup> Juan C. Valdiviezo-N.,<sup>a\*</sup> and Gonzalo Urcid<sup>b</sup>

<sup>a</sup>Engineering Division, Polytechnic University of Tulancingo, Tulancingo, Hgo. 43629, Mexico.

<sup>b</sup>Optics Department, INAOE, Tonantzintla, Pue. 72000, Mexico.

## ABSTRACT

This research introduces an automatic technique designed for the digital restoration of the damaged parts in historical documents. For this purpose an imaging spectrometer is used to acquire a set of images in the wavelength interval from 400 to 1000 nm. Assuming the presence of linearly mixed spectral pixels registered from the multispectral image, our technique uses two lattice autoassociative memories to extract the set of pure pigments conforming a given document. Through an spectral unmixing analysis, our method produces fractional abundance maps indicating the distributions of each pigment in the scene. These maps are then used to locate cracks and holes in the document under study. The restoration process is performed by the application of a region filling algorithm, based on morphological dilation, followed by a color interpolation to restore the original appearance of the filled areas. This procedure has been successfully applied to the analysis and restoration of three multispectral data sets: two corresponding to artificially superimposed scripts and a real data acquired from a Mexican pre-Hispanic codex, whose restoration results are presented.

**Keywords:** Historical documents, image restoration, lattice autoassociative memories, multispectral imaging

## 1. INTRODUCTION

The innovation of remote sensing systems has extended the range of application of imaging spectroscopy to many different fields, including astrophysics, medicine, and military research, as well as food industry. Recently, such systems have shown a great potential for the analysis and restoration of cultural heritage.<sup>1</sup> Imaging spectrometers, also known as *hyperspectral sensors*, are capable to register hundreds of images of the same area, from visible to near infrared (VIS-NIR) wavelength intervals of the electromagnetic spectrum. The data collected by such sensors contain the spectral information of the materials comprising the scene, which can be used for further identification of any component. Compared to other technologies, such as RGB color cameras, hyperspectral systems represent non-invasive techniques, suitable for the restoration of ancient documents and artworks because they provide additional information that is not visible to the naked eye. Because many of the existing ancient documents are damaged by the pass of time, hyperspectral imaging together with digital image processing serve for two main purposes. First, they can be used as a guide for the actual restoration of the document. Second, they can be employed as a tool to produce a virtually restored version of the document under study.<sup>2</sup>

During the last years, different techniques based on multi- and hyperspectral imaging have been used for the identification of pigments and inks in artworks, as well as the enhancement of illegible scripts in palimpsests.<sup>3</sup> For example, multispectral imaging was employed for the examination of the painting Holy Trinity Predella;<sup>4</sup> 21 images were acquired in the VIS and NIR portions. Ware *et al.* performed the characterization of pigments with similar coloration and different chemical composition in Maya murals found at the Naj Tunich cave in Guatemala.<sup>5</sup> By collecting 13 spectral images at wavelengths from 400 to 1000 nm, in increments of 50 nm, the authors could differentiate between the original pigments and over-paintings. Furthermore, multispectral imaging techniques have successfully revealed the erased or occluded text in palimpsests.<sup>1</sup> Particularly the acquisition of 12 images from 385 to 880 nm, in combination with digital image processing algorithms, let to the retrieval of most of the original text in the Archimedes palimpsest.<sup>6</sup> Similarly, a hyperspectral imager system used to enhance the

---

\*Corresponding author: carlos.valdiviezo@upt.edu.mx, Tel.: +52 775 755 8202, ext. 1509

contrast between the under and overwritten texts in the Patmos palimpsest has been proposed in Ref. [1]. For a survey of past and current applications of spectral imaging techniques in the restoration of artworks, the reader may consult Ref. [3].

On the other hand, once a set of spectral images has been collected from a document under study, two main steps are required to perform the digital restoration of the manuscript: 1) the analysis of the pigments to locate their spatial distribution along the image and 2) the digital restoration of the damaged areas. The first step can be realized based on *spectral unmixing* analysis, assuming that the spectral information in the recorded image is represented by mixtures of the constituent materials spectra;<sup>7</sup> thus, the goal in the analysis is to identify those pure pixels that best separate the spectral mixtures. For the autonomous identification of pure spectral pixels in hyperspectral imagery, different techniques can be employed as those described in Ref. [8]. Moreover, the digital restoration process contemplates the use of tools for cracks and holes removal. An algorithm for crack removal is usually a two-step procedure: first the cracks must be detected and then the selected areas must be filled in.<sup>9</sup> The automatic identification of cracks can be realized by the use of morphological filtering, such as the top-hat transformation;<sup>2</sup> cracks and holes removal can be realized by the application of a region filling algorithm. Another important step in the restoration process is the application of color interpolation algorithms to restore the original appearance of the filled areas. Therefore, digital image processing has a wide range of opportunities and challenges for virtual restoration of historical documents and artworks.<sup>10</sup>

This paper introduces an autonomous technique to perform the virtual restoration of the damaged areas in historical documents. For this purpose, the technique is divided into three main steps: (1) multispectral image acquisition of the document, (2) identification of constituent materials spectra (pure pigments) that best separate the spectral mixtures, and (3) the virtual restoration of the damaged parts in the image. The organization of this document is as follows. Section II introduces the mathematical preliminaries required to understand the concept of lattice autoassociative memories. Section III states the problem of determining a set of constituent materials spectra from multispectral data and presents the algorithm used to solve this task. In Section IV we describe the algorithms that we use for virtual restoration of the document, including cracks and holes removal, as well as color interpolation. Section V describes the spectral imaging systems used to collect the data used in our experiments; this section ends with experimental results obtained when our technique is applied to some multispectral data. Finally, Section VI states our conclusions and directions of future work.

## 2. MATHEMATICAL PRELIMINARIES

### 2.1 Lattice based operations

Lattice algebra operations have given place to new applications in pattern recognition problems, image processing, as well as the development of a novel class of neural networks named as *lattice associative memories*, also known as *morphological associative memories*.<sup>11,12</sup> From the point of view of computational intelligence, an associative memory is a device able to store  $k$  vector pairs  $(x^\xi, y^\xi)$ , with the property that the memory recalls each  $x^\xi$  when its corresponding associated vector  $y^\xi$  is presented as an input. Lattice associative memories are the cornerstone of our technique since they are able to extract the set of pure pigments conforming a given historical document. In order to understand their mathematical definitions, in the next lines we describe some lattice algebra operations based on the algebraic structure  $(\mathbb{R}_{\pm\infty}, \vee, \wedge, +)$ , with underlying set the extended real numbers together with the binary operations of taking the maximum, minimum, or adding two elements of  $\mathbb{R}_{\pm\infty}$ .

Taking the maximum or minimum of two numbers  $x, y$ , generally denoted as  $\max(x, y)$  and  $\min(x, y)$ , will be rewritten as binary operators, such that,  $x \vee y = \max(x, y)$  and  $x \wedge y = \min(x, y)$ . In particular, given two  $m \times n$  matrices  $X, Y$ , the maximum is defined as  $(X \vee Y)_{ij} = x_{ij} \vee y_{ij}$ , for all  $i = 1, \dots, m$  and  $j = 1, \dots, n$ . Inequalities between matrices are also verified elementwise, e. g.,  $X \leq Y$  if and only if  $x_{ij} \leq y_{ij}$ . Furthermore, the *conjugate matrix*  $X^*$  is defined as  $-X^T$  where  $X^T$  denotes usual matrix transposition, or equivalently,  $(X^*)_{ij} = -x_{ji}$ , hence  $X \vee Y = (X^* \wedge Y^*)^*$ . In lattice matrix algebra, the usual matrix operations of addition and multiplication are replaced by corresponding lattice operations, providing a faster computation. For a more complete treatment of applied lattice algebra the reader may consult Ref. [12–14].

## 2.2 Lattice associative memories

Let  $\mathbf{x}^\xi = (x_1^\xi, \dots, x_n^\xi)^T \in \mathbb{R}^n$  and  $\mathbf{y}^\xi = (y_1^\xi, \dots, y_m^\xi)^T \in \mathbb{R}^m$  for  $\xi = 1, \dots, k$ , be input and output vector sets, respectively. Given a set of  $k$  vectors associations, such that  $(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^k, \mathbf{y}^k)$ , we introduce a set of associated matrices  $(X, Y)$ , where  $X = (\mathbf{x}^1, \dots, \mathbf{x}^k)$  and  $Y = (\mathbf{y}^1, \dots, \mathbf{y}^k)$ , with an association expressed by  $(\mathbf{x}^\xi, \mathbf{y}^\xi)$  for  $\xi \in k$ . Two  $m \times n$  lattice associative memories (LAMs) able to store the set of vector associations  $(X, Y)$  are defined as the *min-memory*  $W_{XY}$  and the *max-memory*  $M_{XY}$ , whose entries are computed for  $i, j = 1, \dots, n$  by,

$$(W_{XY})_{ij} = w_{ij} = \bigwedge_{\xi=1}^k (y_i^\xi - x_j^\xi), \quad ; \quad (M_{XY})_{ij} = m_{ij} = \bigvee_{\xi=1}^k (y_i^\xi - x_j^\xi). \quad (1)$$

Notice from the above equations that the size of  $W_{XY}$  and  $M_{XY}$  depends on space dimensionality but not on the number of vectors stored in  $X$ . Moreover, if the input and output sets are equal, i.e.,  $X = Y$ , we speak of a *lattice autoassociative memory* (LAAM), otherwise we have a *lattice heteroassociative memory* (LHAM).<sup>11,12</sup> We emphasize that the set conformed by the column vectors of the autoassociative memories  $W_{XX}$  and  $M_{XX}$  can be used to determine the set of constituent materials spectra conforming a multispectral image.

## 3. AUTONOMOUS IDENTIFICATION OF PURE PIGMENTS

In order to perform the virtual restoration of a historical document, the identification of pure pigments spectra from the multispectral image is of crucial importance since such spectra can be used to compute maps of the distribution of each pigment along the scene. These maps can serve for two main purposes: 1) to determine the absence or presence of pigments in areas that are not visible to the naked eye and 2) to locate cracks and holes that are suitable for digital restoration. In the next few lines we state the approach used for the identification of pure pigments from multidimensional data.

### 3.1 Spectral mixtures

Many of the existing historical documents are conformed by mixtures of various pigments and inks. Hence, when a multispectral image is registered from them, the spectral pixels in the image will probably contain the spectral mixtures of such pigments. Therefore, we can assume that a single multispectral pixel has registered the combined spectral information of the constituent materials in the scene; in mathematical terms a single pixel can be modeled as a linear combination of such spectra, following the *constrained linear mixing model* (CLMM), expressed by,

$$\mathbf{x} = \sum_{i=1}^k a_i \mathbf{s}^i + \mathbf{n} = \mathbf{S}\mathbf{a} + \mathbf{n} \quad (2)$$

$$a_i \geq 0 \quad \forall i \quad \text{and} \quad \sum_{i=1}^k a_i = 1, \quad (3)$$

where  $\mathbf{x}$  is the  $n$ -dimensional spectrum of a multispectral pixel,  $\mathbf{S} = (\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^k)$  is an  $n \times k$  matrix whose columns are the  $k$  purest spectral pixels that are also associated with constituent materials,  $\mathbf{a} = (a_1, a_2, \dots, a_k)^T$  is a  $k$ -dimensional column vector whose entries are the corresponding fractional abundances or, equivalently, the percentages of purest spectra present in  $\mathbf{x}$ , and  $\mathbf{n}$  is an additive noise vector.<sup>7</sup>

The aim of the CLMM is to estimate the fractions of pure pigments present in a single pixel, assuming that we know their spectral signatures. The min- and max- lattice autoassociative memories provide special properties that can be used to identify sets of pure spectral pixels from a multispectral image,<sup>15</sup> as can be seen in the following subsection. Furthermore, the estimation of fractional abundances can be performed through the inversion of Eq. (2) subject to the imposed restrictions specified by Eq. (3). This process, known as *spectral*

*unmixing*, allows to quantify the proportion of each pure pigment in every image pixels. A simple solution is provided, making  $\mathbf{n} = 0$ , by the least square estimation method expressed by,

$$\mathbf{a} = (S^T S)^{-1} S^T \mathbf{x}. \quad (4)$$

This estimation exists when  $S$  is of full rank. The abundances that result from this estimation do not necessarily satisfy the constraints imposed in Eq.(3). Full additivity can be satisfied using the method of *Lagrange Multipliers*, while the non-negativity condition can be enforced by applying the *non-negative least squares numerical method*.<sup>16</sup> However, satisfying both constraints simultaneously is not an easy task.

### 3.2 Determination of pure pigments using lattice associative memories

For a given multispectral image  $X \in \mathbb{R}^n$ , conformed by  $k$  spectral vectors of dimension  $n$ , i.e.  $X = (\mathbf{x}^1, \dots, \mathbf{x}^k) \in \mathbb{R}^n$ , a fundamental result regarding the computation of the  $W_{XX}$  and  $M_{XX}$  matrices is that a subset of their column vectors can be used as a selection of pure pixels related to the given image  $X$ . Furthermore, since the relationship of lattice memories with the set of original data  $X$  is not direct, for example  $W_{XX}$  has negative entries, an *additive scaling* of both memories is required. For this purpose two scaled matrices, denoted respectively as  $\overline{W}$  and  $\overline{M}$ , are defined for all  $i = 1, \dots, n$  by the following expressions,

$$\overline{\mathbf{w}}^i = u_i + \mathbf{w}^i \quad ; \quad u_i = \bigvee_{\xi=1}^k x_i^\xi, \quad (5)$$

$$\overline{\mathbf{m}}^i = v_i + \mathbf{m}^i \quad ; \quad v_i = \bigwedge_{\xi=1}^k x_i^\xi. \quad (6)$$

where  $\mathbf{u} = \bigvee_{\xi=1}^k \mathbf{x}^\xi$  and  $\mathbf{v} = \bigwedge_{\xi=1}^k \mathbf{x}^\xi$  denotes, respectively, the *maximum* and *minimum vector bounds* of  $X$ , and whose entries are defined for all  $i = 1, \dots, n$ . According to Eq. (1), the size of both  $\overline{W}_{XX}$  and  $\overline{M}_{XX}$  is  $n \times n$ , where  $n$  corresponds to the dimensionality of the input data.

In order to perform the spectral unmixing process, it is necessary to select a subset  $\mathcal{C}$  of representative columns from the set  $\{\overline{W} \cup \overline{M}\}$ , such that  $S = \mathcal{C}$  in Eq. (2). This selection is mandatory for the two following reasons. First, contiguous columns of  $\overline{W}$  (independently  $\overline{M}$ ) are highly correlated forming  $\lfloor \sqrt{n+1} \rfloor$  groups, each one  $\lfloor \sqrt{n+1} \rfloor$  vectors. Second, the number of columns in the  $S$  matrix must be much less than  $n$ . In fact, the selection of the number of vectors in  $\mathcal{C}$  should be in accordance with the number of materials to be unmixed in the image. For this purpose, the columns of  $\overline{W} \cup \overline{M}$  can be group into a finite number of classes using the  $k$ -means clustering. As it was shown in Ref. [15] the centroids of this clustering process can be effectively used as the columns of  $S$ , providing a good estimation for  $\mathbf{a}$ .

## 4. TOOLS FOR VIRTUAL RESTORATION OF HISTORICAL DOCUMENTS

As we stated before, many historical documents show damages, such as cracks and holes, caused by inadequate environmental conditions and handling through the passage of time. In order to have a better understanding about the original appearance of a given historical document, the development of algorithms oriented toward its digital restoration becomes necessary. This section introduces the algorithm we are applying for cracks and holes removal, which is based on mathematical morphology. Once the algorithm has filled in the damaged areas, the next step consists of restoring their original color by application of an interpolation algorithm.

### 4.1 Cracks and holes removal

From the perspective of digital image restoration, a hole may be defined as a background region surrounded by a connected border of foreground pixels. Hence, the aim of a digital restoration would then be to remove such holes in the image. In this paper, the restoration process is conducted by means of a morphological algorithm, known as *region filling*; this algorithm is based on set operations together with morphological dilations. In order to clarify how our region filling algorithm works, we first introduce the basic morphological operations.

The basic operations of *erosion* and *dilation* works with small shapes known as *structuring elements* in order to shrink or grow objects in a given image. Given an image  $A$  and a structuring element  $B$ , both binary sets, erosion and dilation are defined, respectively, by either the left and right parts of the following expression,

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad ; \quad A \oplus B = \{z | [(\hat{B})_z] \cap A \neq \emptyset\}. \quad (7)$$

where  $z$  denotes the locations in  $A$  and  $\hat{B}$  denotes the reflected structuring element. Notice that the erosion of  $A$  by  $B$  refers to the set consisting of all locations  $z$  such that  $B$ , translated by  $z$ , is contained in  $A$ . Similarly, the dilation of  $A$  by  $\hat{B}$  is the set of all displacements,  $z$ , such that  $\hat{B}$  overlaps at least some portion of  $A$ .<sup>9</sup>

On the other hand, the region filling algorithm is described as follows. Let  $A$  denote a binary image with a region whose elements are 8-connected, enclosing a hole. Since the objective is to fill the hole region with 1's, the algorithm starts by taking a point  $p$  inside the boundary to form an initial set  $X_0$ ; a value of 1 is assigned to  $p$ , being 0 all the remaining elements. The set will be iteratively dilated until no more dilations be necessary. The corresponding algorithm is computed as,

$$X_q = (X_{q-1} \oplus B) \cap A^c \quad q = 1, 2, 3, \dots, \quad (8)$$

where  $B$  is the symmetric structuring element and  $A^c$  is the complement of  $A$ . The algorithm is repeated until  $X_q = X_{q-1}$  for iteration  $q$ . At the end of the algorithm, the set  $X_q \cup A$  contains the filled hole and its boundary. Notice in the above expression that the intersection of  $(X_{q-1} \oplus B)$  at each step with  $A^c$  limits the result to inside the region of interest. Once we have obtained the filled holes with this procedure, it is possible to know their corresponding pixel positions in the original image; this positions will then be used for an interpolation process. We turn now to the description of the color interpolation technique that we are using.

## 4.2 Color interpolation

We sometimes know the value of a function at a set of points, but we do not know an analytic expression for a function that lets us calculate its value at an arbitrary point.<sup>17</sup> In digital images, the goal of color interpolation algorithms is to restore the original appearance in those parts where color has been affected. There exist a number of interpolation algorithms commonly used in the literature, such as nearest neighbor, bilinear, cubic, B-spline interpolations, among others.

In this research paper we use bilinear interpolation to approximate the color of a pixel inside a crack or hole. For this purpose, we interpolate two nearby pixel values, belonging to the boundary of the hole. To ensure a good interpolation process, we first measure the Euclidean distance in eight straight line directions (up, down, left, right and four diagonals) between each boundary pixel and the pixel to be interpolated. The pixel values used for interpolation are those whose positions are along the shortest distance. Mathematically, if  $I_1(x_1, y_1)$  and  $I_2(x_2, y_2)$  denote two pixels belonging to the boundaries of an image hole and  $I(x, y)$  corresponds to the pixel we want to interpolate, at locations  $(x, y)$ , we define the Euclidean distance ( $d_E$ ) between  $I_1$  and  $I$  by the expression,<sup>18</sup>

$$d_E(I, I_1) = \sqrt{(x_1 - x)^2 + (y_1 - y)^2}. \quad (9)$$

The interpolated value for a pixel  $I$  at  $(x, y)$  coordinates is obtained according to the following equation,

$$I(x, y) = \frac{I_2(x_2, y_2) \times d_E(I, I_1) + I_1(x_1, y_1) \times d_E(I, I_2)}{d_E(I_1, I_2)}. \quad (10)$$

Figure 1 displays an RGB color image, representing a  $5 \times 5$  neighborhood of a pixel  $I(x, y)$  to be interpolated. According to the algorithm, the black pixel will be interpolated by taking into account the two nearest neighbors along a straight line drawn from  $I(x, y)$ . In this example, it is easy to note that pixels labeled as  $I_1$  and  $I_2$  show the closest distance to  $I(x, y)$ . Therefore, these pixel values will be used to interpolate the value of  $I(x, y)$ , according to Eq.(10). Finally, once the interpolation has been accomplished, a noise or penumbra effect could be detected given that the boundaries of cracks and holes are formed by gray pixels. As a result, we could have an image showing a small visible trace of the regions where damaged areas existed. To overcome this border problem, a median filter could be applied to the resulting image to smooth the restored regions.

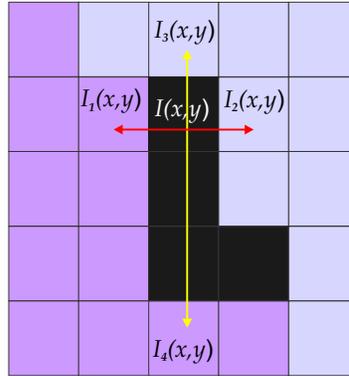


Figure 1. Bilinear interpolation by taking the two closest neighbor pixels to  $I(x, y)$ .

## 5. EXPERIMENTAL RESULTS

The techniques presented in previous sections have been applied to the restoration of three multispectral data sets. The first two examples were registered from artificially superimposed scripts drawn on conventional paper. Particularly, example 1 is composed by alcohol and pen inks, while example 2 is conformed by alcohol, fluorescent and pen inks, respectively. This last sample was scraped off to create holes for subsequent restoration. In both samples, inks were intentionally superimposed to test the capability of the described algorithms to separate the scripts. Example 3 corresponds to the multispectral data of a pre-Hispanic document known as *Colombino* codex. This codex was elaborated in the XII century to preserve the military, political and religious feats of the governors “8-Venado”, “Garra de Tigre” and “4 Viento” in the ancient Mixteca culture of Mesoamerica.<sup>19</sup> Most of these codices were painted on *amate* paper, a kind of paper elaborated from trees cortex. Moreover, in the following subsections we provide an explanation of the performed restoration process by the steps: 1) multispectral image acquisition, 2) autonomous determination of pure pigments, and 3) virtual restoration.

### 5.1 Multispectral image acquisition

#### 5.1.1 Superimposed scripts

Several images of superimposed scripts samples were acquired using a VIS-NIR hyperspectral camera (HeadWall Photonics, Massachusetts, USA), located at the Optics and Vision Laboratory of the Polytechnic University of Tulancingo. Table 1 presents the corresponding characteristics of the hyperspectral system. The instrument provides scan, capture and analysis capabilities of hyperspectral data by using mainly four elements: an imaging spectrometer, a fiber optics quartz tungsten halogen (QTH) lamp, a motorized displacement system, and a data processing unit (HDPU). The imaging spectrometer is conformed by diffraction gratings to decompose the incoming light spectrum into several images at wavelengths intervals from 400 to 1000 nm. The (QTH) lamp operates in conjunction with a base plate, attached to the motorized displacement system, to scan line by line a sample under study. Hence, the imaging spectrometer is conformed by a linear array of CCD sensors (*pushbroom* system), where the platen motion is used to acquire a whole image of the sample. Finally, the HDPU allows to quickly store and analyze the acquired data cube, since it is fitted out with a Graphics Processing Unit (GPU).<sup>20</sup>

Table 1. VIS-NIR Hyperspectral camera characteristics.

Specification	Value
Spectral coverage	400 - 1000 nm
Spectral resolution	0.7414800 nm
Spatial resolution	1004 × 2640 pxs
Acquisition speed	50 - 200 fps
Collected images	812

### 5.1.2 Colombino codex

For preservation of cultural legacy, the Colombino codex was digitalized by the National Library of Anthropology and History (Biblioteca Nacional de Antropología e Historia) in Mexico city.<sup>21</sup> For that purpose, a multispectral camera equipped with a CCD sensor of spatial resolution equal to  $2048 \times 2048$  was used. The camera was connected to a rotating wheel with 16 optical filters for recording 16 images, with a bandwidth of about 25 nm, in the interval between 380 to 780 nm. In order to minimize the radiation of light, an illumination coming from flush light synchronized with 16 exposures was used.<sup>22</sup>

## 5.2 Abundance maps of pigments distribution

The aim of this step is to identify the spatial distribution of each pigment in the image through the spectral unmixing analysis. For examples 1 and 2, twenty spectral bands were used as input to LAAMs, while for example 3 we used 16 bands. Then, for each case the distribution of pigments conforming the image was achieved according to the following procedure. Given a multispectral image  $X$  of size  $m \times p$  with  $n$  spectral bands, we first transformed it into a bidimensional array of the form  $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \in \mathbb{R}^n$ , where  $k = m \times p$ . Then,  $X$  served as the input set to compute the matrix memories  $W_{XX}$  and  $M_{XX}$ . In addition, using the vector bounds  $\mathbf{u}$  and  $\mathbf{v}$ , we obtained the corresponding scaled versions  $\bar{W}_{XX}$  and  $\bar{M}_{XX}$ , respectively. According to the discussion presented in Section 3.2, a set of pure pigments spectra representing the spectral mixtures in the scene can be extracted from the set  $\{\mathcal{C} = \bar{W} \cup \bar{M}\}$ . For this purpose, a clustering process based on the  $k$ -means algorithm was conducted from  $\mathcal{C}$ , and the resulting cluster centroids were employed as the columns of  $S$  in Eq.(2).

Moreover, the determination of fractional abundances for each pigment spectrum, satisfying the non-negativity condition and relaxing the full additivity, was realized by means of the *non-negative least square numerical method* (available in Matlab). This algorithm provides a good estimation of vector  $\mathbf{a}$  for each spectral pixel  $\mathbf{x}$  in the image. At the end of the procedure, the values of the abundance vector were mapped into a gray-scale image to build abundance maps, where brighter areas corresponds to maximum abundance of the corresponding material. Figure 2 presents the abundance maps obtained for three pigments in example 1. For this example we used a multispectral data set of size  $383 \times 985$ , with 20 spectral bands. By considering three pigments and the background paper, the number of clusters specified in the  $k$ -means algorithm was 4. Hence,  $S = (\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4)$ , where  $\mathbf{c}^i$ , for  $i = 1, \dots, 4$ , corresponds to the centroid for each cluster. Notice that our approach has effectively separate the superimposed scripts, which can be seen in the three abundance maps presented. Although the fractional abundances for the white paper is not presented, its corresponding spectral signature was used in the  $S$  matrix to unmix the spectral mixtures in the image.



Figure 2. Abundance maps computed for pigments in example 1 using LAAMs and spectral unmixing. First row, from left to right: sample RGB color image and distribution of blue ink. Second row, from left to right: distribution of black and alcohol inks, respectively.

Figure 3 displays the abundance maps showing the distribution of pigments for our example 2. For this experiment we use a multispectral data set of size  $417 \times 837$ , with 20 spectral bands. This time the  $S$  matrix



Figure 3. Abundance maps computed for pigments in example 2 using LAAMs and spectral unmixing. First row, from left to right: sample RGB color image and distribution of fluorescent ink. Second row, from left to right: distribution of alcohol and water inks, respectively.

was conformed by vectors of each of the 4 clusters instead of their centroids in order to demonstrate that a representative of each group can also be used. Hence  $S = \{\bar{\mathbf{w}}^7, \bar{\mathbf{w}}^{11}, \bar{\mathbf{w}}^{15}, \bar{\mathbf{m}}^4\}$ . In this particular example maps 1 and 2 show some holes in those positions where inks were scraped off.



Figure 4. Virtual restoration performed on example 2 using region filling and color interpolation algorithms. Left: abundance map for background paper used to locate all holes. Right: digitally restored RGB color image.

### 5.3 Virtual restoration

Once the fractional abundances for each pigment have been computed, their corresponding maps may be used for subsequent restoration. Therefore, the last step in the restoration process is accomplished by the use of the tools for virtual restoration discussed in Section 4. Because some cracks and holes could not be easily detected, the binarization of the abundance maps at a given threshold can help to emphasize those damaged portions in the image. The region filling algorithm (see Eq.(8)) is applied on cracks and holes appearing in the binarized maps. When all regions are filled, their respective spatial coordinates will be used to restore the original color in the image.

On the other hand, before performing the color interpolation algorithm described in Eq. (10), it is necessary to compose an RGB color image from the multispectral bands to perform the color restoration. For this purpose, three spectral bands centered approximately at wavelengths of 605 nm, 555 nm, and 430 nm, respectively, can be used.<sup>9</sup> Then, the spatial positions of the previously filled holes will be used as a guide to interpolate the color on the RGB color image.

Getting back to the example 2, the upper right and lower left parts of Figure 3 show the abundance maps where holes were located. In order to perform the virtual restoration process, these holes were filled based on the abundance map corresponding to the background since it contains all the holes simultaneously (see the left

part of Fig. 4). After applying the region filling and color interpolation algorithms, the damaged parts can be effectively restored as shown in the right part of Fig. 4. By comparing the resulting image against the original one, we observe that the color was correctly restored in the regions of interest.

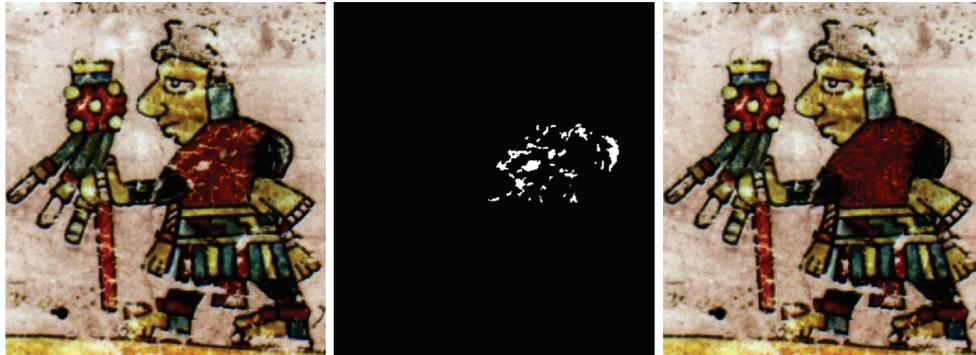


Figure 5. Virtual restoration performed on example 3 using region filling and color interpolation algorithms. From left to right: sample RGB color image, filled holes used for subsequent interpolation, and digitally restored RGB color image.

Finally, the previously described algorithms were tested on a real multispectral set acquired from the Colombino codex. For this purpose, we employed a subimage of  $376 \times 341$  pixels. First, the matrix of constituent materials spectra was formed with the resulting centroids of the clustering achieved on  $\overline{W} \cup \overline{M}$ , such that  $S = \{\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4, \mathbf{c}^5\}$ . Through the spectral unmixing process, five abundance maps were generated, which served as a reference for virtual restoration. In particular maps showing the distribution of red and black pigments were used to fill in the holes appearing in the body of the personage. As in the previous example, the spatial positions of such holes were used as a guide to interpolate the color in the image. The left part of Fig. 5 displays an RGB color image of the codex formed by the combination of bands 10, 8, and 3, respectively\*. The central and right parts of the same Figure shows, respectively, the filled holes and the resulting restored image obtained with the discussed procedure. As one can see, the restored image allows an improved visualization and, consequently, a better understanding of the original appearance of the document.

## 6. CONCLUSIONS

In this research we have described an unsupervised technique, based on multispectral imaging, that can be effectively used for digital restoration of historical documents. Our technique is divided into three main steps: (1) multispectral image acquisition, (2) identification of pure pigments spectra that best represent the spectral mixtures, and (3) the virtual restoration of the damaged parts in the image. Multispectral imaging in the VIS-NIR wavelength intervals has allowed to retrieve additional information related to the distribution of pigments that is not visible using conventional color imaging. Furthermore, lattice autoassociative memories are used to extract a set of pure pigments that are employed to separate the spectral mixtures at every image pixels. As a result, the calculus of the fractional abundance maps for each pigment let us determine the locations of holes and cracks in the image. Finally, the algorithms used for virtual restoration have helped to improve the visualization and understanding of the original appearance of a given document.

The results achieved with the application of our technique to the previous examples state that multispectral imaging, in conjunction with the spectral unmixing based on lattice autoassociative memories, can be used for digital restoration of historical documents. We remark that the procedure here described is an unsupervised technique that let us know the distribution of pigments even in those parts that are not visible to the naked eye. In fact, the pigments distribution are the basis to perform a virtual restoration. Future work contemplates the improvement of our color interpolation algorithm to restore those areas where boundaries have been partially erased.

\*recall that the codex was digitized by a 16-bands camera

## ACKNOWLEDGMENTS

Edwin Lechuga thanks to the National Council of Science and Technology for scholarship No. 556848. Juan C. Valdiviezo and Gonzalo Urcid are grateful with the National Research System (SNI-CONACyT) for partial support through grants No. 57564 and 22036, respectively. We also express our gratitude to Professor M. Yamaguchi for the pre-Hispanic data set used in our simulations.

## REFERENCES

- [1] Rapantzikos, k., and Balas, C., "Hyperspectral imaging: potential in non-destructive analysis of palimpsest," *Proc. IEEE: International Conference on Image Processing*, 2, 618-621 (2005).
- [2] Spagnolo, G.S., and Somma, F., "Virtual restoration of cracks in digitized image of paintings," *Journal of Physics: Conference Series*, 249(1), 012059 (2010).
- [3] Fisher, C., and Kakoulli, I., "Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications," *Review in Conservation*, 7, 3-16 (2006).
- [4] Baronti, S., Casini, A., Lotti, F., and Porcinai, S., "Multispectral imaging system for the mapping of pigments in works of art by use of principal component analysis," *Applied Optics*, 37, 1299-1309 (1998).
- [5] Ware, G.A., Chabries, D.M., Christiansen, R.W., Brady, J.E., and Martin, C.E., "Multispectral analysis of ancient Maya pigments: implications for the Naj Tunich Corpus," *Proc. IEEE: International Geoscience and Remote Sensing Symposium*, 2489-2491 (2000).
- [6] Easton, R.L., Knox, K.T., and Christens-Barry, W.A., "Multispectral imaging of the Archimedes palimpsest," *Proc. IEEE: 32nd Applied Imagery Pattern Recognition Workshop*, 111-116 (2003).
- [7] Keshava, N., and Mustard, J. F., "Spectral unmixing," *IEEE Signal Processing Magazine*, 19(1), 44-57 (2002).
- [8] Plaza, A., Martinez, P., Perez, R., Plaza, J., "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Transactions on Geoscience and Remote Sensing*, 40(9), 2025-2041, (2002)
- [9] Gonzalez, R.C., and Woods, R.E., [Digital image processing], Pearson Prentice Hall, 649-667 (2008).
- [10] Barni, M., Pelagotti, A., and Piva, A., "Image processing for the analysis and conservation of paintings: opportunities and challenges," *IEEE Signal Processing Magazine*, 22(5), 141-144 (2005).
- [11] Ritter, G.X., and Sussner, P. Díaz de León, J.L., "Morphological associative memories," *IEEE Trans. Neural Networks*, 9(2), 281-293 (1998).
- [12] Ritter G.X., and Gader P., "Fixed point of lattice transforms and lattice associative memories," in *Advances in Imaging and Electron Physics*, P. Hawkes ed., Elsevier, 144, 165-242 (2006).
- [13] Kaburlasos, V.G., and Ritter, G.X. (eds.), [Computational Intelligence based on Lattice Theory], Springer Verlag, (2007).
- [14] Urcid, G., and Valdiviezo-N, J.C., "Generation of lattice independent vector sets for pattern recognition applications," *Proc. SPIE*, 6700, 1 (2007).
- [15] Valdiviezo-N, J. C., and Urcid, G., "Lattice algebra approach to multispectral analysis of ancient documents," *Applied Optics*, 52(4), 674-682 (2013).
- [16] Lawson, C.L., Hanson, R.J., [Solving least squares problems], chap. 23, Prentice-Hall, Englewood Cliffs NJ, (1974).
- [17] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., [Numerical recipes in C], Cambridge university press, 2, (1996).
- [18] Lee, T., Ng, V., Gallagher, R., Coldman, A., and McLean, D., "Dullrazor®: A software approach to hair removal from images," *Computers in Biology and Medicine*, 27(6), 533-543 (1997).
- [19] B.D.M., "Biblioteca Digital Mexicana," available in <http://bdmx.mx>.
- [20] H.P., "Headwall Photonics," available in <http://headwallphotonics.com>.
- [21] Conde, J., Haneishi, H., Yamaguchi, M., Ohshima, N., Baez, J., "Spectral reflectance estimation of ancient mexican codices, multispectral images approach," *Revista Mexicana de Física*, 50(5), 484-489 (2004).
- [22] Yamaguchi, M., Teraji, T., Ohsawa, K., Uchiyama, T., Motomura, H., Murakami, Y., Ohshima, N., "Color image reproduction based on the multispectral and multiprimary imaging: experimental evaluation," *Proc. SPIE*: 4663, 15 (2002).