Lattice algebra approach to multispectral analysis of ancient documents

Juan C. Valdiviezo-N[1,*] and Gonzalo Urcid[2]

[1]División de Ingenierías, Universidad Politécnica de Tulancingo, Tulancingo, Hidalgo 43629, Mexico

[2]Coordinación de Óptica, Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla 72000, Mexico

*Corresponding author: carlos.valdiviezo@upt.edu.mx

Received 24 September 2012; revised 5 December 2012; accepted 20 December 2012; posted 20 December 2012 (Doc. ID 176752); published 30 January 2013

This paper introduces a lattice algebra procedure that can be used for the multispectral analysis of historical documents and artworks. Assuming the presence of linearly mixed spectral pixels captured in a multispectral scene, the proposed method computes the scaled min- and max-lattice associative memories to determine the purest pixels that best represent the spectra of single pigments. The estimation of fractional proportions of pure spectra at each image pixel is used to build pigment abundance maps that can be used for subsequent restoration of damaged parts. Application examples include multispectral images acquired from the Archimedes Palimpsest and a Mexican pre-Hispanic codex. © 2013 Optical Society of America

OCIS codes: 100.2000, 100.2960, 100.4996.

1. Introduction

Advances in imaging technologies and the development of efficient image processing methods have motivated their application to the analysis and restoration of artworks and historical documents. Many of the existing ancient documents and artworks are damaged by inadequate environmental conditions, storage, and handling through the passage of time. For example, reusing the parchment of a book to write a novel text was a common practice during the fourteenth century [1]. In recent years, different efforts have been directed toward the analysis, enhancement, and restoration of the original texts in historical documents. The importance behind the study of such manuscripts resides in the fact that valuable information can still be retrieved from them. For that purpose, noninvasive techniques, such as multispectral and hyperspectral imaging, have shown to be effective methods for the study of documents and paintings based on the discrimination of material spectra [2–4]. Analysis of pigments and other pictorial materials has been first realized using reflectance spectroscopy acquired by fiber-optic devices operating in the ultraviolet (UV), visible (VIS), and near-infrared (NIR) electromagnetic ranges [5]. However, the development of multispectral imaging instruments, able to register simultaneously the spatial and spectral information of a scene at different wavelength intervals, represents an important advantage over other technologies. Because the spectral reflectance of materials is a quantity directly related to their physical properties, the spectral information derived from a painting or a handwritten document provides an effective way to identify and discriminate among different inks and pigments present.

In recent years, several techniques using multispectral devices have been directed to the interpretation of paintings, the identification of pigments and inks in artworks, and the enhancement of important illegible scripts in palimpsests [5]. One of the first applications of multispectral imaging was reported in [6]. The authors performed the examination of the painting *Holy Trinity Predella* through the acquisition of 21 images in the VIS and NIR wavelength

1559-128X/13/040674-09$15.00/0

intervals. Furthermore, Ware *et al.* [7] performed the characterization of pigments with similar coloration and different chemical composition in Maya murals found at the Naj Tunich cave in Guatemala; a clustering analysis of the data led to differentiation between the original and later paintings. False color images generated by a combination of multispectral imaging and principal component analysis (PCA) have been employed for the examination of artworks from the Winnipeg Art Gallery: the drawing *Untitled* and the oil painting named *The Mocking of Christ* [8].

More recently, spectral imaging techniques have been successfully applied to reveal the unreadable text lines of the Dead Sea Scrolls [9]. Moreover, the use of multispectral imaging has been oriented toward the reconstruction of erased or occluded texts in palimpsests. One of the most important studies has been motivated by the Archimedes Palimpsest project [1,10]; the combination of imaging and digital analysis techniques, such as spectral mixing analysis and PCA, led to the retrieval of most of the original text. A recent method applied to the enhancement of the underwritten text in the palimpsest makes use of statistical processing techniques, such as *independent component analysis* and PCA [11]. Also, in [12] a character segmentation method of ancient documents that includes a Markov random field model using the spectral information and the spatial features of the characters is reported. For a survey of past and current applications of spectral techniques and digital image processing algorithms in the restoration of artworks, as well as the limitations of both technologies, the reader may consult references [5,13].

Although the combination of multispectral imaging and digital image processing methods has assisted in the enhancement and the recovery of most of the damaged parts of historical documents, there are still important limitations that require the investigation of new techniques. In particular, palimpsests conformed by scripts with similar spectra are difficult to discriminate. Therefore, the development of new methods for image analysis should be considered in order to perform a better enhancement of the underwritten texts.

This paper introduces an unsupervised multispectral image procedure that can be used to perform the analysis of ancient historical documents and artworks. Under the assumption that most spectral pixels registered in a multispectral scene are conformed by a linear combination of constituent material spectra, our technique enables one to find sets of the purest spectral signatures corresponding to the different pigments in the document or, equivalently, pigment spectra. The selected signatures are then used to segment the multispectral image into pigment maps. Segmentation results are given for a pair of multispectral data collections belonging to the Archimedes Palimpsest and a pre-Hispanic codex named *Matrícula de Tributos*. The main contribution of this paper is the proposal of a hybrid scheme that uses both lattice associative memories and a clustering classification method for the unsupervised determination of pigment distribution from multispectral images. This technique represents a noninvasive tool designed to facilitate the restoration process of artworks and occluded texts in palimpsests. The organization of the document is as follows. In Section 2 we describe the multispectral image acquisition process that was used for our simulations, and we discuss briefly the basic method for reflectance estimation. Section 3 presents the mathematical basis of our pigment analysis based on lattice algebra. In Section 4 we provide illustrative examples showing the distribution of pigments obtained by means of the proposed technique. Finally, in Section 5 we present the conclusions of this research and directions for future work.

2. Multispectral Imaging of Historical Documents

Multispectral devices, developed to sample the electromagnetic spectrum at different spectral bands, had been mainly employed for remote sensing applications. In recent years, the use of such devices to estimate the spectral reflectance of materials conforming artworks and manuscripts has allowed new applications focused on their analysis, preservation, and restoration. A pioneer research for these applications is presented in [14]; the authors found that at wavelengths close to 2 μm the common paints, such as iron gall ink and sepia, become invisible. In addition, given that the spectral reflectance is a quantity directly related to the physical properties of materials, an appropriate estimation is of fundamental importance in order to perform the characterization of materials conforming a scene.

For the analysis discussed here we have employed, as an initial test, a portion of the multispectral image that was collected from the Archimedes Palimpsest [15]. The manuscript was created on parchment in the tenth century to copy seven of the treatises of the Greek mathematician; two centuries later, the manuscript was washed out and the pages were cut in half to overwrite a Christian prayer book. In order to retrieve most of the original text, the whole manuscript was imaged using light from LED illumination at different wavelength intervals and a color Sinar 54H sensor as the capturing system [1,10]; 12 spectral bands were registered at different wavelength intervals, covering the spectral range from 395 to 870 nm of the electromagnetic spectrum.

For the second experiment, a multispectral image collected from a set of pre-Hispanic documents called *codices* was used. These documents represented a pictorial writing system used to keep testimonies in the cultures of Mesoamerica before the Spanish conquest. Most of these pictorial documents were drawn on *amate* paper, a kind of paper manufactured from tree cortex, whose content described the adventures of important people as well as their beliefs about their time and history. From the existing codices that are still preserved, we mention the codex known as *Colombino* that was created in the twelfth century to relate the feats of the governor "8-Venado"

in the ancient cultures of Mixtecas [16]. Another important codex, named *Matrícula de Tributos*, was created in the sixteenth century to register the tributes that the subordinates had to pay to Tenochtitlan, the center of the Aztec Empire. Specifically, we used this last codex to perform our simulations.

The codices were digitalized for conservation at the National Library of Anthropology and History (Biblioteca Nacional de Antropología e Historia) in Mexico city [17,18]. For the image acquisition, a multispectral camera conformed by a rotating wheel with 16 optical filters mounted to a charged couple device (CCD) was used [19]. Illumination coming from flush light synchronized with 16 exposures was employed to minimize light irradiation. Hence, 16 spectral bands in the region from 380 to 780 nm were collected, each one covering an area of 2048×2048 pixels, with a bandwidth of about 25 nm. Figure 1 displays the spectral sensitivities of the optical filters. Because of the registered data corresponding to the intensity values, it is necessary to estimate the spectral reflectance of the scene, as is explained in the next subsection.

A. Reflectance Estimation

For a multispectral capturing system conformed by a monochromatic CCD camera, a set of m optical filters, and a collimated white light source, the output of the camera c_i obtained with the ith filter (for $i = 1, \ldots, m$) is given by [20]

$$c_i = \int_{\lambda_{\min}}^{\lambda_{\max}} s_i(\lambda) r(\lambda) \mathrm{d}\lambda + n_i, \tag{1}$$

where $s_i(\lambda) = f_i(\lambda) E(\lambda)$ is the spectral sensitivity of the camera, computed as the multiplication of the ith filter response $f_i(\lambda)$ and the spectral distribution $E(\lambda)$ of the illumination source, $r(\lambda)$ is the spectral reflectance of the object, and n_i is a noise term added to the ith image. A simple solution to this equation can be obtained through discrete estimation techniques [20,21]. By taking q samples from $s_i(\lambda)$ and $r(\lambda)$, we can approximate Eq. (1) by a discrete matrix equation as

$$c_i = \mathbf{s}_i^T \mathbf{r} + n_i, \tag{2}$$

where T denotes vector transposition, $\mathbf{s}_i$ and $\mathbf{r}$ are $q \times 1$ vectors, and $\mathbf{s}_i^T = \mathbf{f}_i^T E$ is the product of the spectral sensitivity vector for the ith filter and the $q \times q$ diagonal matrix E that specifies the spectral distribution of the illumination. In addition, if $\mathbf{s}_i^T$ occupies the ith row of matrix S, then all c_i values can be represented as an $m \times 1$ vector, such that

$$\mathbf{c} = S\mathbf{r} + \mathbf{n}. \tag{3}$$

Assuming the absence of noise, a first approximation $\hat{\mathbf{r}}$ of the spectral reflectance can be obtained using the generalized inverse estimation given by

$$\hat{\mathbf{r}} = S^{+}\mathbf{c} = S^T(SS^T)^{-1}\mathbf{c}. \tag{4}$$

As was shown in [20], the generalized inverse estimate may give oscillatory responses. To diminish the effect of unwanted oscillations, smoothing or Wiener estimation methods are usually applied. More specifically, the smoothing estimation is given by

$$\hat{\mathbf{r}} = M^{-1}S^T(SM^{-1}S^T)^{-1}\mathbf{c}, \tag{5}$$

where M is a smoothing matrix. Another alternative consists in the application of Wiener estimation methods. In this case, $\mathbf{r}$ is considered to be a sample vector resulting from a random process with a known mean and a covariance matrix C. The Wiener estimation is expressed as

$$\hat{\mathbf{r}} = CS^T(SCS^T)^{-1}\mathbf{c}. \tag{6}$$

As an approximation, the covariance matrix can be modeled as a covariance matrix resulting from a first-order Markov process, for $i, j = 1, \ldots, q$, such as

$$C_{ij} = \rho^{(i-j)\mathrm{sgn}(i-j)}, \tag{7}$$

where $0 \le \rho \le 1$ is the correlation factor of the adjacent element and sgn is the signum function. Figure 2 shows the spectral reflectance curves computed from the generalized inversion and Wiener estimation techniques for intensity values representing a red pigment in the *Matrícula de Tributos* codex. Since the Wiener estimate provides smoothed results compared with the generalized inversion, we applied this technique with $\rho = 0.94$ for reflectance estimation of the codex.

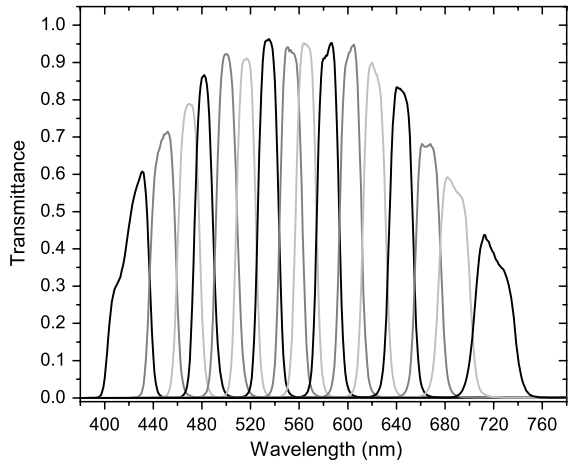

Fig. 1. Spectral sensitivity of the 16 CCD multispectral camera used to collect the multispectral image of the pre-Hispanic codices.

3. Lattice Algebra and Pigment Extraction

A. Spectral Mixtures

In many situations, art paintings and historical documents show mixtures of different pigments and inks. Thus, when a multispectral image is acquired from them, the spectral pixels in the image will probably contain the spectral mixture of such pigments. Hence, we can assume that a single pixel

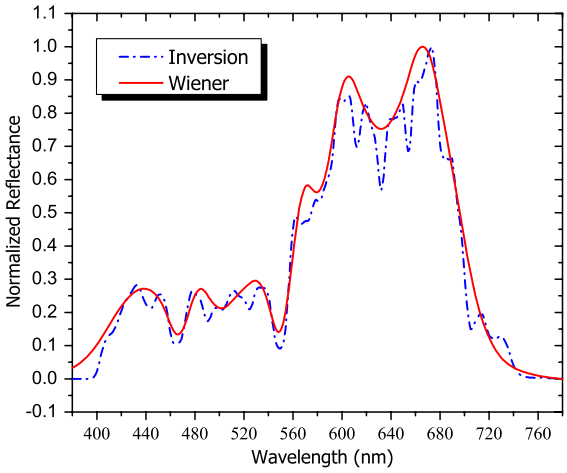

Fig. 2. (Color online) Reflectance estimation realized with the generalized inversion (dashed curve) and the Wiener estimate (solid curve) with $\rho = 0.94$, for a red pigment taken from the *Matrícula de Tributos* codex.

in a multispectral collection has registered the combined spectral reflectance of various materials. The spectral mixtures can be represented as a linear combination of the constituent material spectra, following the *constrained linear mixing model* (CLMM), expressed by

$$\mathbf{x} = \sum_{i=1}^{k} a_i \mathbf{p}^i + \mathbf{n} = P\mathbf{a} + \mathbf{n}, \tag{8}$$

$$a_i \geq 0 \forall i \quad \text{and} \quad \sum_{i=1}^{k} a_i = 1, \tag{9}$$

where $\mathbf{x}$ is the measured spectrum over m bands of a multispectral pixel, $P = (\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^k)$ is an $m \times k$ matrix whose columns are the k purest spectral pixels that are also associated with constituent materials, $\mathbf{a} = (a_1, a_2, \ldots, a_k)^T$ 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 [22].

The purpose of the CLMM is to estimate the fractions of constituent materials present in a single pixel, assuming that we know their spectral response (appearing in the columns of P). It is clear that a good election of purest spectral pixels will allow us to perform a better estimation of corresponding abundances. The min- and max-lattice auto-associative memories, a recent development in lattice-algebra-based neural networks [23,24], provide special properties that can be used to extract sets of the purest spectral pixels from a multichannel image.

B. Lattice Algebra and Associative Memories

In the next few lines we describe some lattice algebra operations based on the algebraic structure $(\mathbb{R}_{\pm\infty}, \vee, \wedge, +)$, with the underlying set the extended real number system together with the binary operations of taking the maximum, minimum, or adding two elements of $\mathbb{R}_{\pm\infty}$. It turns out that $(\mathbb{R}_{\pm\infty}, \vee, \wedge, +)$ is a *bounded lattice ordered group*. For a more complete treatment of applied lattice algebra the reader may consult [25–27]. The basic numerical operations of taking the maximum or minimum of two numbers, usually denoted as functions $\max(x,y)$ and $\min(x,y)$, will be written as binary operators using the "join" and "meet" symbols of lattice algebra, namely, $x \vee y = \max(x,y)$ and $x \wedge y = \min(x,y)$. Lattice matrix operations are defined componentwise; for example, let A, B be two $m \times n$ matrices such that $A = (a_{ij})$ and $B = (b_{ij})$, then the maximum, minimum, and addition of two matrices are defined respectively as $(A \vee B)_{ij} = (a_{ij} \vee b_{ij})$, $(A \wedge B)_{ij} = (a_{ij} \wedge b_{ij})$, and $(A + B)_{ij} = a_{ij} + b_{ij}$, for each $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Furthermore, given an $m \times p$ matrix A and a $p \times n$ matrix B with real entries, two new lattice matrix operations known as the *max-sum* and the *min-sum* of A and B, denoted respectively as $A \boxtimes B$ and $A \boxtimes B$ are given by

$$(A \boxtimes B)_{ij} = \bigvee_{k=1}^{p} (a_{ik} + b_{kj}); \quad (A \boxtimes B)_{ij} = \bigwedge_{k=1}^{p} (a_{ik} + b_{kj}). \tag{10}$$

If $p = 1$, then matrix A is reduced to a column vector and matrix B is reduced to a row vector; consequently, $(A \boxtimes B)_{ij} = (A \boxtimes B)_{ij} = a_i + b_j$. This particular case reduces to an $m \times n$ matrix that defines the *outer sum* of two vectors [24]. Note the similarity of Eq. (10) to the usual matrix product as used in linear algebra, where the summation operation is replaced by the generalized min or max operations and componentwise scalar multiplication is substituted by addition. We remark that lattice matrix operations are the mathematical basis for developments in diverse areas that include machine scheduling, pattern recognition, associative memories, and image processing [26,28]. In these applications, the usual matrix operations of addition and multiplication are replaced by corresponding lattice operations, providing a faster computation.

We turn to the concept of associative memories based on lattice matrix operations. Given two sets $X = (\mathbf{x}^1, \ldots, \mathbf{x}^k)$ and $Y = (\mathbf{y}^1, \ldots, \mathbf{y}^k)$ of n-dimensional vectors, we define the association of X and Y as the set $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1, \ldots, k\}$, consisting of *associated pairs* denoted by $(\mathbf{x}^\xi, \mathbf{y}^\xi)$ for each ξ. For $X = Y$, an *auto-associative memory* is a physical or computational device able to store k vectors, with the property that the memory recalls each $\mathbf{x}^\xi$ when it is presented as an input. The *lattice auto-associative min-* and *max-memories* (LAAMs), denoted respectively by W_{XX} and M_{XX}, are defined componentwise for $i,j = 1, \ldots, n$ by

$$(W_{XX})_{ij} = w_{ij} = \bigwedge_{\xi=1}^{k} (x_i^\xi - x_j^\xi);$$

$$(M_{XX})_{ij} = m_{ij} = \bigvee_{\xi=1}^{k} (x_i^\xi - x_j^\xi). \tag{11}$$

According to Eq. (11), the size of W_{XX} and M_{XX} depends on space dimensionality but not on the number of vectors stored in X. For more details about lattice associative memories the reader may consult [29]. An important fact of the column vectors of LAAMs is that their relationship with the set of original data X is not direct; for example, W_{XX} usually has negative entries. Hence, an *additive scaling* of both memories is required to relate the column vectors to the data set X. The *maximum* and *minimum vector bounds* of X, denoted respectively by $\mathbf{u}$ and $\mathbf{v}$, are used to realize the additive scaling and are defined for all $i = 1, \ldots, n$ by the following expressions:

$$\mathbf{u} = (u_i) = \bigvee_{\xi=1}^{k} x_i^{\xi}; \qquad \mathbf{v} = (v_i) = \bigwedge_{\xi=1}^{k} x_i^{\xi}. \tag{12}$$

Thus, the new shifted column vectors of two scaled matrices, denoted by $\overline{W}$ and $\overline{M}$, are defined for all $i = 1, \ldots, n$ as follows:

$$\overline{\mathbf{w}}^i = \mathbf{w}^i + u_i; \qquad \overline{\mathbf{m}}^i = \mathbf{m}^i + v_i. \tag{13}$$

A fundamental result regarding the W and M matrices states that the set of column vectors $\mathcal{C} = \overline{M} \cup \overline{W} \cup \{\mathbf{u}, \mathbf{v}\}$ forms a *convex polytope* with $2(n+1)$ vertices containing X [30,31]. Thus, any subset of $\mathcal{C}$ can be employed as a selection of the purest pixels related to the given image X. In other words, $\mathcal{C}$ allows us to determine the spectral pixels that best represent the material spectra comprising the scene.

Before introducing the application examples, we have achieved an initial simulation based on a multispectral data set acquired from a Macbeth color chart, conformed by 31 spectral bands in the range from 400 to 700 nm [32]. The objective of the experiment is to illustrate the use of LAAMs for the autonomous extraction of representative spectral signatures from the image. According to the described procedure, from the set X conformed by all the spectral vectors in the multispectral imagery, we compute the matrix memories W_{XX} and M_{XX}, as well as the corresponding scaled versions $\overline{W}$ and $\overline{M}$. Hence, a reduced number of columns obtained from $\overline{M} \cup \overline{W} \cup \{\mathbf{u}, \mathbf{v}\}$ could be selected as representative spectra of specific color charts in the image. For that purpose, three spectral vectors were selected from pixels in the image corresponding to the red, green, and blue charts, respectively, which were labeled as reference spectra. In order to determine the columns of LAAMs that best match the corresponding reference spectra, the spectral angle computed between each reference signature and the column vectors of $\overline{W} \cup \overline{M}$ allowed us to associate the columns $\overline{\mathbf{w}}^{31}$, $\overline{\mathbf{w}}^{14}$, $\overline{\mathbf{w}}^{6}$, with the red, green, and blue spectra, respectively. The left part of Fig. 3 displays the positions of these spectral pixels selected from the image (shown with white squares); their associated spectral signatures and those approximated by $\overline{W}$ are displayed in the right part of the same figure, which are drawn, respectively, in solid and dashed-dotted curves. The computed spectral angle values in each case are 0.13, 0.15, and 0.3 rad, respectively. It is important to remark that LAAMs enable us to approximate the constituent material spectra from a particular scene.

4. Multispectral Image Analysis

The mathematical procedure previously described has been applied for the analysis of two real multispectral image collections. The first application example corresponds to the Archimedes Palimpsest multispectral image, whose spectral signatures are already given as reflectance values. The second example was performed using the multispectral image of the *Matrícula de Tributos* codex, for which in Section 2 we have provided an explanation of the image acquisition process. Since collected raw data are intensity values, it was necessary to calculate the spectral reflectance with the Wiener estimation given in Eq. (6). In the following subsections we present the segmentation results obtained with the proposed lattice-algebra-based method for both collections.

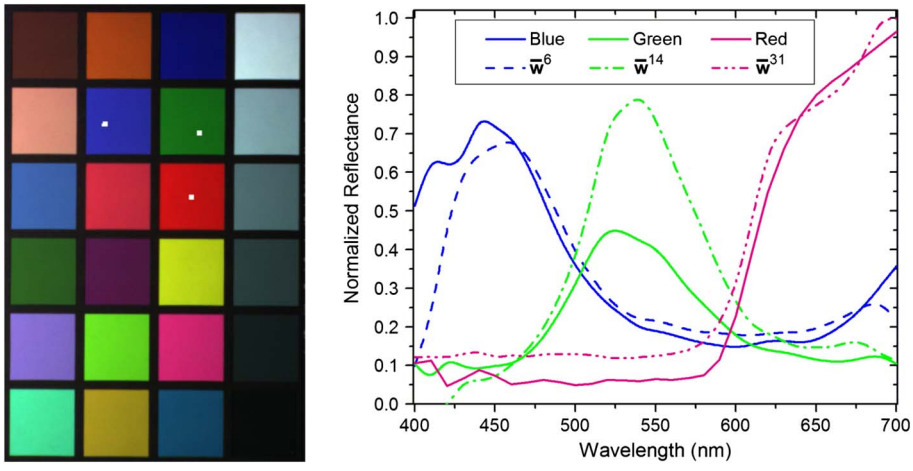

Fig. 3. (Color online) Left: three spectral pixels selected from the Macbeth chart multispectral image. Right: spectral curves associated with these selected pixels (solid curves) and corresponding spectral curves approximated by $\overline{\mathbf{w}}^{31}$, $\overline{\mathbf{w}}^{14}$, $\overline{\mathbf{w}}^{6}$ (dashed-dotted curves).

A. Archimedes Palimpsest

The proposed method was tested using a subimage of size 270×360 pixels, taken from the 11 multispectral bands of the left page in the folium $014v$–$019r$ of the Archimedes Palimpsest. We first formed the set $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^k\} \in \mathbb{R}^n$, conformed by all spectral vectors in the subimage, where $k = 270 \times 360 = 97200$ and $n = 11$. Next, we computed W_{XX} and M_{XX}, the vector bounds $\mathbf{u}$ and $\mathbf{v}$, and the scaled matrices $\overline{W}$ and $\overline{M}$, both of size 11×11. According to our results, adjacent columns of $\overline{W}$ (independently $\overline{M}$) are highly correlated, forming $\lfloor \sqrt{n+1} \rfloor$ groups, each one with $\lfloor \sqrt{n+1} \rfloor$ vectors. To eliminate highly correlated column vectors from each matrix memory that could produce erroneous results, we computed the matrix of correlation coefficients and then selected those vectors with the lowest correlation values. In this case only three vectors were selected as the "purest" pixels from the set $\mathcal{C}$ to form the columns of P in Eq. (8); in particular, $P = \{\overline{\mathbf{m}}^1, \overline{\mathbf{w}}^1, \overline{\mathbf{w}}^{11}\}$.

Furthermore, the estimation of fractional proportions for each constituent material can be realized by inversion of Eq. (8), subjected to the full additivity and non-negativity constraints of abundance coefficients. Although the simple inversion provides a unique solution, some coefficients result as negative for many pixels in the image. To overcome this problem, an alternative procedure consists in satisfying the non-negative condition but relaxing full additivity. For the examples discussed here we made use of the *non-negative least square* (NNLS) estimation algorithm [33] (available in MATLAB). The numerical estimation provides the proportions of each constituent material appearing at spectral pixels. Figure 4 shows the abundance maps of two pigments and the parchment present in the Archimedes Palimpsest, whose values were scaled to the interval [0,255] for visualization purposes. Thus, brighter areas mean maximum abundance of the corresponding material. Notice that the use of LAAMs followed by linear unmixing allows us to segment the original script in the manuscript.

B. Matrícula de Tributos Codex

As was shown in the previous example, the lattice-algebra-based analysis provides significant results that can be applied for the interpretation and subsequent restoration of historical documents. In this second example, a subimage of the *Matrícula de Tributos* multispectral collection was employed. Similarly, from the subimage of size 481×671 pixels, we formed the set X containing all the spectral pixels; thus, the set $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^k\} \in \mathbb{R}^n$, with $k = 322751$ and $n = 16$, was used to extract the constituent material spectra. By computing the matrix memories W_{XX} and M_{XX} and their respective scaled versions, $\overline{W}$ and $\overline{M}$, two matrices of size 16×16 were obtained. The lowest values in the matrix of correlation coefficients computed from $\mathcal{C}$ were used to select four uncorrelated column vectors; as a result, $P = \{\overline{\mathbf{w}}^1, \overline{\mathbf{w}}^7, \overline{\mathbf{m}}^2, \overline{\mathbf{m}}^7\}$. Finally, the estimation of the abundance coefficients at image pixels by means of the NNLS numerical method allowed us to know the distribution of each pigment along the image. Figure 5 displays the colored distribution of two pigments and the amate paper present in the codex (grayscale image for print version). The image was composed by combining the abundance maps, which were colored with representative colors in order to improve visualization.

In many applications it is necessary to perform the digital enhancement or even the restoration of the damaged parts of a document. For that purpose, digital image processing algorithms can provide adequate results. Given that we have determined the distribution of pigments, even in those parts of the document that are not visible, we can employ the abundance maps to enhance the image. For example, by thresholding a particular abundance map, we can emphasize some important parts of the manuscript. In addition, the

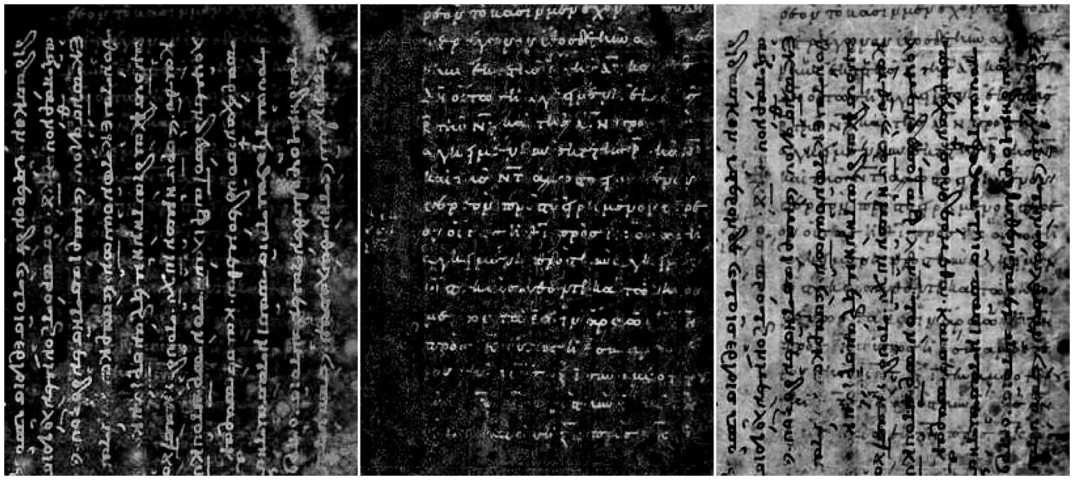

Fig. 4. Different scripts extracted from the Archimedes Palimpsest using the lattice-based method and linear unmixing. From left to right: Medieval twelfth century overlaid text (vertical), ancient Greek original text (horizontal), and background parchment. Brighter areas correspond to maximum percentage of the corresponding material.

application of smoothing or morphological filters can help to fill in the holes of the damaged parts. Some of these techniques have been applied to digitally restore the color of the codex. First, a thresholding procedure performed on the abundance maps allowed us to highlight the distribution of pigments in the regions that are not visible. Once each abundance map was modified, the mean color for each pigment was computed from the image itself, whose values were used to assign an RGB color to the corresponding map. Finally, the abundance maps were combined in order to form a complete image. The upper part of Fig. 6 shows a true color image of the codex formed by the combination of bands 13 (red), 8 (green), and 3 (blue), while the lower part presents a digitally restored image obtained with the discussed procedure applied to the codex. According to these images, it is clear that color restoration can be performed using basic digital image processing algorithms.

C. Selection of Columns from $\overline{W} \cup \overline{M}$

A remarkable property derived from the calculus of $\overline{W}$ or $\overline{M}$ is the high correlation between their corresponding contiguous columns. As was shown in [31,34], a reduced number of them can be used as representative spectra of constituent materials (also known as *endmembers*). Therefore, a similarity measure should be applied to select a reduced number of uncorrelated vectors. In the current research we have used different techniques for column selection in order to demonstrate that any of them can provide similar results. As was stated before, a selection based on the matrix of linear correlation computed from $\overline{W} \cup \overline{M}$ provides good results for pigment extraction, which is in accordance with the previously discussed results. Furthermore, clustering techniques computed from $\overline{W}$ or $\overline{M}$ may be applied to subdivide

Fig. 6. (Color online) Digital restoration of the *Matrícula de Tributos* codex achieved by means of the pigment distribution maps. Top: RGB color image representing true color. Bottom: digitally restored image.

Fig. 5. (Color online) Pigment distribution in the *Matrícula de Tributos* codex determined from the lattice-based method and linear unmixing. Blue color (dark gray in print version) represents red pigment, pink color (light gray) represents yellow pigment, and gray area (background) represents the amate paper.

their respective column vectors in a finite number of classes, say $\lfloor \sqrt{n} \rfloor$, from which a representative of each group can be selected. Particularly, the centroids approximated by K-means and fuzzy C-means (FC-means) clustering techniques, denoted respectively as K^{γ} and C^{γ}, for $\gamma = 1, \ldots, \lfloor \sqrt{n} \rfloor$, are related to some columns of $\overline{W}$ or $\overline{M}$, being suitable candidates of pigment spectra. Table 1 shows a selection of endmembers from the set $\overline{W} \cup \{\mathbf{u}\}$ that was accomplished for both the Archimedes Palimpsest and the *Matrícula de Tributos* codex. Column vectors similar to K^{γ} and C^{γ} were chosen based on the correlation coefficient for values above 0.92 and 0.94, respectively, and whose relative position in the row specifies the correspondence among them. Similarly, Table 2 presents a selection of endmembers from the set $\overline{M} \cup \{\mathbf{v}\}$ that was computed for both multispectral data sets. Column vectors similar to the centroids of K-means and FC-means were selected based on the correlation coefficient for values above 0.91 and 0.93, respectively. Notice that instead of using these classical clustering techniques for column selection,

Table 1. Techniques Used for Endmember Selection from the Set $\overline{W} \cup \{\mathbf{u}\}$[a]

Image	Classes	K-Means	Endmembers	FC-Means
Palimpsest	3	K^2, K^3, K^1	$\{\overline{\mathbf{w}}^3, \overline{\mathbf{w}}^9, \overline{\mathbf{w}}^{11}\}$	C^3, C^1, C^2
Codex	4	K^3, K^1, K^2, K^4	$\{\overline{\mathbf{w}}^1, \overline{\mathbf{w}}^6, \overline{\mathbf{w}}^{14}, \overline{\mathbf{w}}^{15}\}$	C^4, C^2, C^1, C^3

[a]This set was computed from the Archimedes Palimpsest and the *Matrícula de Tributos* codex.

Table 2. Techniques Used for Endmember Selection from the Set $\overline{M} \cup \{\mathbf{v}\}$[a]

Image	Classes	K-Means	Endmembers	FC-Means
Palimpsest	3	K^1, K^3, K^2	$\{\overline{\mathbf{m}}^1, \overline{\mathbf{m}}^9, \overline{\mathbf{m}}^{12}\}$	C^2, C^1, C^3
Codex	4	K^4, K^2, K^1, K^3	$\{\overline{\mathbf{m}}^2, \overline{\mathbf{m}}^6, \overline{\mathbf{m}}^{12}, \overline{\mathbf{m}}^{13}\}$	C^1, C^2, C^4, C^3

[a]This set was computed from the Archimedes Palimpsest and the *Matrícula de Tributos* codex.

any other classification method, such as the novel algorithm based on sub-band partitioning and fusion for hyperspectral classification [35], could give similar results, or even better than those previously presented.

Finally, given an $M \times N$ multispectral data set, conformed by n spectral bands whose values are already specified as reflectance quantities, the sequence of steps required for the unsupervised determination of pigment abundances in our hybrid approach is summarized as follows.

1. Transform the multispectral image into a bidimensional array of spectral vectors $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^k\} \in \mathbb{R}^n$, with $k = MN$.
2. Compute W_{XX} and M_{XX} from X.
3. Using $\mathbf{u}$ and $\mathbf{v}$, scale both memories to obtain $\overline{W}_{XX}$ and $\overline{M}_{XX}$.
4. Determine a subset of $\sqrt{n}$ vectors from the set $\mathcal{C} = \overline{M} \cup \overline{W} \cup \{\mathbf{u}, \mathbf{v}\}$ by means of the correlation coefficients or K-means.
5. For each of the $\sqrt{n}$ constituent spectra build the corresponding abundance map through the NNLS numerical method.

5. Conclusions

This paper presents a fast and reliable procedure based on lattice algebra that can be used for pigments analysis and subsequent restoration of ancient and historical documents. The multispectral imaging of historical documents in the VIS and IR portions allows us to extract additional information that is not available with common techniques, such as broadband color photography. The combination of multispectral imaging and digital image processing algorithms provides new tools to enhance the appearance of ancient scripts as well as to increase our understanding of their content. Basically, the proposed hybrid method computes two scaled lattice autoassociative matrix memories to extract the set of purest spectral pixels that represents the constituent materials in the document. Assuming linear spectral mixtures at image pixels, we used the NNLS numerical method to determine the pigment abundances present in the multispectral image. The method was applied to a portion of the Archimedes Palimpsest and the *Matrícula de Tributos* codex. In the first case, the results obtained clearly show the original script in the manuscript, whereas in the latter case, the determination of pigment abundances was used to enhance the color of the codex. Future work will consider the application of the technique to the analysis of other documents as well as the development of algorithms for restoration purposes.

The authors thank to the owner of the Archimedes Palimpsest and Professor M. Yamaguchi for the data sets used in our simulations. J. C. Valdiviezo thanks the National Council of Science and Technology for doctoral scholarship no. 175027 and G. Urcid is grateful to the National Research System (SNI-CONACYT) for partial support through grant no. 22036. We express our gratitude to the anonymous reviewer whose appropriate comments helped to improve this manuscript.

References

1. R. L. Easton and W. Noel, "Infinite possibilities: ten years of study of the Archimedes Palimpsest," in Proc. Am. Philos. Soc. **154**, 50–76 (2010).
2. M. Lettner and R. Sablatnig, "Multispectral imaging for analyzing ancient manuscripts," in *Proceedings of the 17th European Signal Processing Conference* (EURASIP, 2009), pp. 1200–1204.
3. K. Rapantzikos and C. Balas, "Hyperspectral imaging: potential in nondestructive analysis of paintings," in *Proceedings of the IEEE International Conference on Image Processing* (IEEE, 2005), pp. 618–621.
4. S. Tominaga and N. Tanaka, "Spectral image acquisition, analysis, and rendering for art paintings," J. Electron. Imaging **17**, 043022 (2008).
5. C. Fisher and I. Kakoulli, "Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications," Rev. Conserv. **7**, 3–16 (2006).
6. S. Baronti, A. Casini, F. Lotti, and S. Porcinai, "Multispectral imaging system for the mapping of pigments in works of art by use of principal component analysis," Appl. Opt. **37**, 1299–1309 (1998).
7. G. A. Ware, D. M. Chabries, R. W. Christiansen, J. E. Brady, and C. E. Martin, "Multispectral analysis of ancient Maya pigments: implications for the Naj Tunich corpus," in *Proceedings of the IEEE 2000 International Geoscience and Remote Sensing Symposium* (IEEE, 2000), pp. 2489–2491.
8. M. Attas, E. Cloutis, C. Collins, D. Goltz, C. Majzels, J. R. Mansfield, and H. H. Mantsch, "Near-infrared spectroscopy imaging in art conservation: investigation of drawing constituents," J. Cult. Herit. **4**, 127–136 (2003).
9. K. Knox, R. Johnston, and R. L. Easton, Jr., "Imaging the Dead Sea Scrolls," Opt. Photonics News **8**(8), 30–34 (1997).

10. R. L. Easton, K. T. Knox, and W. A. Christens-Barry, "Multispectral imaging of the Archimedes Palimpsest," in *Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop* (IEEE, 2003), pp. 111–116.
11. E. Salerno, A. Tonazzini, and L. Bedini, "Digital image analysis to enhance underwritten text in the Archimedes Palimpsest," Int. J. Document Anal. **9**, 79–87 (2007).
12. M. Lettner and R. Sablatnig, "Spatial and spectral based segmentation of text in multispectral images of ancient documents," in *Proceedings of the 10th International Conference on Document Analysis and Recognition* (IEEE, 2009), pp. 813–817.
13. M. Barni, A. Pelagotti, and A. Piva, "Image processing for the analysis and conservation of paintings: opportunities and challenges," IEEE Signal Process. Mag. **22**(5), 141–144 (2005).
14. J. R. J. Van Asperen de Boer, "Reflectography of paintings using an infrared vidicon television system," Stud. Conserv. **14**, 96–118 (1969).
15. "Archimedes Palimpsest project," http://archimedespalimpsest.net/Data/0000-100v/.
16. "Biblioteca Digital Mexicana," http://bdmx.mx.
17. J. Conde, H. Haneishi, M. Yamaguchi, N. Ohyama, and J. Baez, "Spectral reflectance estimation of ancient Mexican codices, multispectral images approach," Rev. Mexicana Fís. **50**, 484–489 (2004).
18. M. Yamaguchi, H. Haneishi, H. Fukuda, J. Kishimoto, H. Kanazawa, M. Tsuchida, R. Iwama, and N. Ohyama, "High-fidelity video and still-image communication based on spectral information: natural vision system and its applications," Proc. SPIE **6062**, 60620G (2006).
19. M. Yamaguchi, T. Teraji, K. Ohsawa, T. Uchiyama, H. Motomura, Y. Murakami, and N. Ohyama, "Color image reproduction based on the multispectral and multiprimary imaging: experimental evaluation," Proc. SPIE **4663**, 15–26 (2002).
20. W. Pratt and C. E. Mancill, "Spectral estimation techniques for the spectral calibration of a color image scanner," Appl. Opt. **15**, 73–75 (1976).
21. S. K. Park and F. O. Huck, "Estimation of spectral reflectance curves from multispectral image data," Appl. Opt. **16**, 3107–3114 (1977).
22. N. Keshava and J. F. Mustard, "Spectral unmixing," IEEE Signal Process. Mag. **19**(1), 44–57 (2002).
23. G. X. Ritter, P. Sussner, and J. L. Díaz de León, "Morphological associative memories," IEEE Trans. Neural Netw. **9**, 281–293 (1998).
24. G. X. Ritter, G. Urcid, and L. Iancu, "Reconstruction of patterns from noisy inputs using morphological associative memories," J. Math. Imaging Vision **19**, 95–111 (2003).
25. V. G. Kaburlasos and G. X. Ritter, eds., *Computational Intelligence based on Lattice Theory*, Vol. **67** (Springer-Verlag, 2007).
26. G. X. Ritter and P. Gader, "Fixed point of lattice transforms and lattice associative memories," in *Advances in Imaging and Electron Physics*, P. Hawkes, ed., Vol. **144** (Elsevier, 2006), pp. 165–242.
27. G. Urcid and J. C. Valdiviezo-N, "Generation of lattice independent vector sets for pattern recognition applications," Proc. SPIE **6700**, 67000C (2007).
28. R. Cuninghame-Green, "Minimax algebra and applications," in *Advances in Imaging and Electron Physics*, P. Hawkes, ed., Vol. **90** (Academic, 1995), pp. 1–121.
29. G. Urcid and J. C. Valdiviezo-N, "Lattice algebra approach to color image segmentation," J. Math. Imaging Vision **42**, 150–162 (2012).
30. G. X. Ritter, G. Urcid, and M. S. Schmalz, "Autonomous single-pass endmember approximation using lattice auto-associative memories," Neurocomputing **72**, 2101–2110 (2009).
31. G. X. Ritter and G. Urcid, "Lattice algebra approach to endmember determination in hyperspectral imagery," in *Advances in Imaging and Electron Physics*, P. W. Hawkes, ed., Vol. **160** (Academic, 2010), pp. 113–169.
32. S. Hordley, G. Finlayson, and P. Morovic, "A multi-spectral image database and an application to image rendering across illumination," in *Proceedings of the IEEE Third International Conference on Image and Graphics* (IEEE, 2004), pp. 394–397.
33. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, 1974).
34. J. C. Valdiviezo-N and G. Urcid, "Multispectral images segmentation of ancient documents with lattice memories," in *Digital Image Processing and Analysis Conference*, OSA Technical Digest Series (Optical Society of America, 2010), paper DMD6.
35. Y.-Q. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," IEEE Trans. Geosci. Remote Sens. **49**, 747–756 (2011).