Int J Comput Vis (2011) 94: 136–151
DOI 10.1007/s11263-011-0419-1

Near-Infrared Ink Differentiation in Medieval Manuscripts

Alexandra Psarrou · Aaron Licata · Vasiliki Kokla · Agamemnon Tselikas

Received: 25 January 2010 / Accepted: 5 January 2011 / Published online: 27 January 2011

Abstract One of the tasks facing historians and preservationists is the authentication or dating of medieval manuscripts. To this end it is important to verify whether writings on the same or different manuscripts are concurrent. We propose a novel approach for the automated image-based differentiation of inks used in medieval manuscripts. We consider the problem of capturing images of manuscript pages in near-infrared (NIR) spectrum and compare the ink appearance and textural features of segmented text. We present feature descriptors that capture the variability of the visual properties of the inks in NIR based on intensity distributions of histograms and co-occurrence matrices. Our approach is novel as it is entirely image based and does not include the spectrum analysis of the inks. The method is validated by using model ink images manufactured based on known recipes and ink segmented from medieval manuscripts dated from the 11th to the 16th century. Model inks are classified by using both supervised and unsupervised clustering. Comparison of inks of unknown composition is achieved through unsupervised multi-dimensional clustering of the feature descriptors and similarity measures of derived probability density functions.

Keywords Document image analysis · Ink type modelling · Co-occurrence matrix analysis

A. Psarrou (✉) · A. Licata · V. Kokla
School of Electronics and Computer Science, University of Westminster, London, UK
e-mail: psarroa@westminster.ac.uk

A. Tselikas
National Bank of Greece Cultural Foundation, Athens, Greece
e-mail: agatselikas@gmail.com

1 Introduction

Researchers in the area of art preservation and historians are in need of authenticating and dating ancient or medieval manuscripts. Such authentication or dating is usually possible through the study of manuscripts and the recovery of historical information such as the year a manuscript was written or facts described in the manuscripts. However, often researchers are not certain of the concurrency of the writings on manuscripts, as some writings are added at a later date. In addition, often information about the date or place a manuscript was written is not available.

In order to extract more information researchers often resort to the study of the type of scripting found on manuscripts in order to determine whether certain writings are by the same scriber. In other cases researchers compare text from different manuscripts in order to establish whether they are of the same era. In addition it is often useful to researchers to know whether the script and the miniatures of manuscripts are from the same ink as this could give an indication of the workshop the manuscripts was written. To successfully address this problem scholars are in need of scientific information, such as the type of ink used on manuscripts, that can be reliably used in the historical examination of works of art. The availability of such information would allow researchers to determine whether the writings on the same or different manuscripts are concurrent.

The main objective of this research is to develop computational models and algorithms for automated image based differentiation of the types of inks used in medieval manuscripts. Manuscript inks are made through the combination of inorganic and organic pigments such as metals, salts and vegetable materials. Existing methods used for the examination of pigments can be applied in the analysis of manuscript inks, however, most are based on destructive testing

techniques that require the physical sampling of data. Such methods cannot be used widely due to their destructive nature and the historical value of the artifacts.

Non-destructive techniques such as spectroscopy and reflectography are more suited to the study and preservation of old manuscripts, where the optical properties of the pigments are studied under illumination beyond the visible spectrum (Janssens et al. 2000). In particular Raman spectroscopy has been applied successfully to the in situ analysis of pigments and some colored inks contained in illuminated manuscripts (Clarke 2001). However, analysis of historic iron gall ink samples using sources of 514 and 633 nm laser excitation has been regarded as difficult due to increased fluorescence of the inks (Brown and Clark 2004). Recently work using near-infrared excitation Raman spectroscopy has been applied but it is yet not clear whether it can be used for the analysis of iron gall inks (Lee et al. 2006). In addition techniques such as Raman spectroscopy even though they can provide quantitative information on ink constitutes, the expense of the equipment and the ability to only provide localized information poses some limits in their application. Examples of such limitations are when the localized samples are contaminated or when the appearance of inks in the visible spectrum does not suggest the usage of different types of inks on a page, and therefore does not probe localized investigation.

Computer vision techniques can be used as alternative or complementary diagnostic methods by computing models and interpreting the visual properties of the material used such as brown inks. In an early approach Kokla studied techniques for image-based ink classification of historical documents using statistical modelling of ink intensity using Gaussian mixtures (Kokla et al. 2000). In a later work, the same authors consider co-occurrence matrices of ink intensities as models of the joint probability of adjacent ink pixels in order to represent the spreading behavior of writing inks and classify eight specific ink compositions (Kokla et al. 2007). Dasari and Bhagvati used an 11-dimensional color and texture vector to derive within-class and between-class distance distributions for text written with ball and gell/roller pens (Dasari and Bhagvati 2007). Another approach is to capture the physical characteristics of liquid inks. In forensics analysis Franke employed Haralick texture features of co-occurrence matrices and support vector machines classifier to discriminate among three classes of ink traces, solid, viscous, and fluid (Franke et al. 2002).

Although we share some of the insights of these authors, we view the previous ink texture recognition classifiers as proof-of-concept. Instead in this paper we focus on two different tasks: (a) to find suitable features to describe ink of different chemical composition and (b) to use these features for comparing previously unseen inks. In the latter case our approach does not rely on any labelled training set as the ink composition of the manuscripts cannot be known. The remaining of this paper is organized as follows. Section 2 provides a background on the type of inks considered during our experiments. Section 3 describes the ink feature descriptors used. Section 4 describes the clustering of the inks based on these features and ink similarity measures applied to differentiate between inks of different composition. Section 5 describes our experiments based on model and manuscript images. We conclude in Section 6.

2 Background

Inks used in Medieval manuscripts are mainly of black-brown color and only occasionally color inks such as red or green have been used. The two common black-brown writing fluids were composed of either carbon or iron galls (Barrow 1972). The carbon inks were composed generally of either soot, lampblack, or some type of charcoal to which gum arabic and solvent such as water, wine, or vinegar were added. The basic ingredients of metal gall inks are copper, iron, galls, gum arabic, and a solvent such as water, wine, or vinegar (Flieder et al. 1975; Monique 1975). If compounded with the proper amount of gum arabic, this ink will flow easily from a quill pen and penetrate the fibres of the paper to form a black, insoluble compound.

The inks used in this study date from the 11th to the 16th century and are employed in manuscripts located in south-east Europe and the eastern Mediterranean areas, especially in areas where the Byzantine Empire and its influence spread, which means that all the writing employed in this study is Greek.

Our first aim is to derive models from standards of inks manufactured according to the recipes given in Zerdoun Bat-Yehouda (1983) for validating our method. We prepared eight inks with various known chemical compositions, in order to represent as many types of inks as possible. The inks we prepared are as follows:

- Carbon ink.
- Metal gall ink. This category contains the copper gall inks and iron gall inks.
- Incomplete inks. This group includes inks, that have a similar composition to that of metal gall inks, although their composition does not include one of the basic ingredients of metal gall inks. We treat them as subclasses of metal gall inks (type A, B and C).
- Mixed ink. This category contains inks that have ingredients of the last two categories.

A summary of the composition of inks is shown in Table 1.

The identifying term “brown ink” is commonly used in the cataloging of all these types of inks. This broad descriptive term does little to indicate the richness or variety of

Table 1 The composition of the model inks used during our experiments and prepared based on recipes from Zerdoun Bat-Yehouda (1983)

Ink types	Chemical components					
	Carbon	$CuSO_4$	$FeSO_4$	Gallic oxide	Alcohol	Arabic glue
Carbon	*X*					*X*
Coppergall		*X*		*X*	*X*	*X*
Fourna		*X*		*X*		*X*
Irrongall			*X*	*X*		*X*
Mixed	*X*	*X*		*X*		*X*
Type A			*X*	*X*	*X*	
Type B				*X*		*X*
Type C		*X*				*X*

tones of the inks which fall within this category. Direct observation and examination of inks under normal light can provide preliminary clues toward identification but mainly differentiate between inks with carbon and non-carbon composition.

Viewing of artifacts under near infrared radiation is widely adopted in the visual examination of works of art, as it provides further information on invisible inks or pigments due to fading or cover of more recent layers of paint. Reflectographical studies on the optical behaviors of inks under visible and near infrared radiation have shown that inks that have very similar photometric properties under visible light can be separated when viewed under infrared radiation (Alexopoulou and Kokla 1999). The differentiation in near infrared is mainly due to the different chemical composition of the inks and can be represented using histograms or mixture of Gaussian functions. A representation is shown in Fig. 1 with the intensity distribution of the eight model inks in the near-infrared spectrum. Figure 2 shows that while in the visible region of the electromagnetic spectrum, the differentiation of various types of inks is low, in the infrared region, this differentiation is increased. This papers shows that such differentiation allows for the statistical classification of the inks.

3 Ink Feature Descriptors

The results in Fig. 1 show clearly that even though there is a difference in the intensity distribution of the inks in near infrared spectrum, this alone is not sufficient to discriminate between the different inks. Therefore additional characteristics have to be taken into account such as the behavior of the inks during the scripting process. In addition materials imaged in NIR spectrum have the advantage of penetrating the ink outer surface without being excessively absorbed. This optical property provides valuable information to the image-based characterization of the spreading behavior of the inks.

One difficulty in defining the set of feature descriptors is that as inks are semi-transparent liquids the density and opacity of their texture are part of their characterizing properties. Frequency, and repeatability of patterns are irrelevant to characterizing ink type texture. As such, texture feature methods used in machine vision, and image recognition with the assumption that texture is defined by pattern frequency, texton, and other perceptual assumptions impinge upon ink texture characterization. For this reason, texture features based on Gabor filter banks, wavelets, Fourier phase, autocorrelation, edge masks, and textons are not well suited for our purpose (Coggins and Jain 1985; Farrokhnia 1990; Malik et al. 2001; Varma and Zisserman 2005). MRF and GRF were shown by Picard to be equivalent, and also to be related to the co-occurrence matrices (Picard et al. 1991). MRF has been used in old documents to separate out and remove ink-bleed from foreground ink intended for reading (Huang et al. 2008). However, our strategy differs in that it seeks to extract important ink information even from the areas of thin ink spread.

Given that grey-level intensity distributions of eight known classes of ink types are insufficiently discriminative we seek features that yield higher between-class separation. Following investigation of a large pool of features we propose to combine features from two families: 1-D intensity descriptors and joint co-occurrence descriptors.

3.1 Grey Level Intensity Statistics

The intensity descriptor is a set of features from grey-level intensity histograms statistics and joint intensity probability distributions encoded in co-occurrence matrices. Ink class attributes are assumed to be independent because although chemically mixed with similar substances they nevertheless exhibit different physical characteristics.

Figure 3 shows the pseudocolor intensity distributions of NIR images of the eight known ink compositions at ten different densities. The distributions provide global statistics of the NIR ink images and show that they vary with different ink types and to some extent with physical density.

Distinctive features of ink properties are derived through five statistical measurements of the histogram distributions as given in Table 2. Under the assumption that the amount of sampling is sufficient the smoothness is a statistical measure that relates to light scattering properties of an ink surface, whereas 1-D entropy is indicative of ink transparency and thickness. In addition mean, variance and skewness provide distinctive features of varying ink properties. Even though the latter statistics are notoriously variant to lighting conditions, as they shift with illuminant direction, after imposing

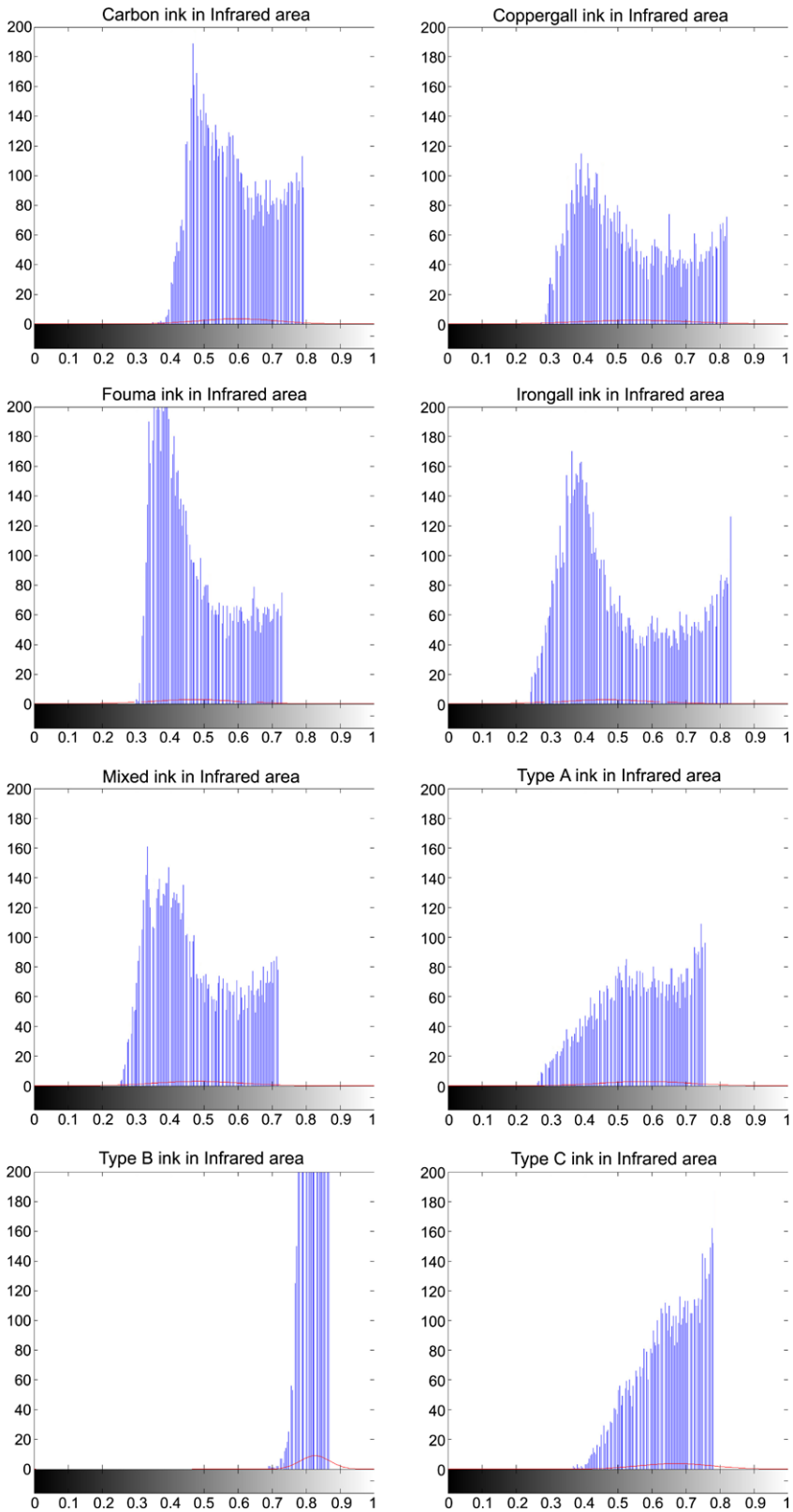

Fig. 1 Intensity histograms in the near infrared area of the spectrum that correspond to the eight model inks used during our experiments. From *left* to *right*: *First row*: Carbon, Coppergall. *Second row*: Fourna inks, Irongall. *Third row*: Mixed, Type A inks. *Bottom row*: Type B, Type C inks

some constraints on the capture conditions (such as camera, illuminant position, scale, and orientation) these features are rather rich in information.

In addition second-order statistics allow to capture the spreading structure of textures such as ink. Our hypothesis is that second-order statistics capture patterns invisible to the

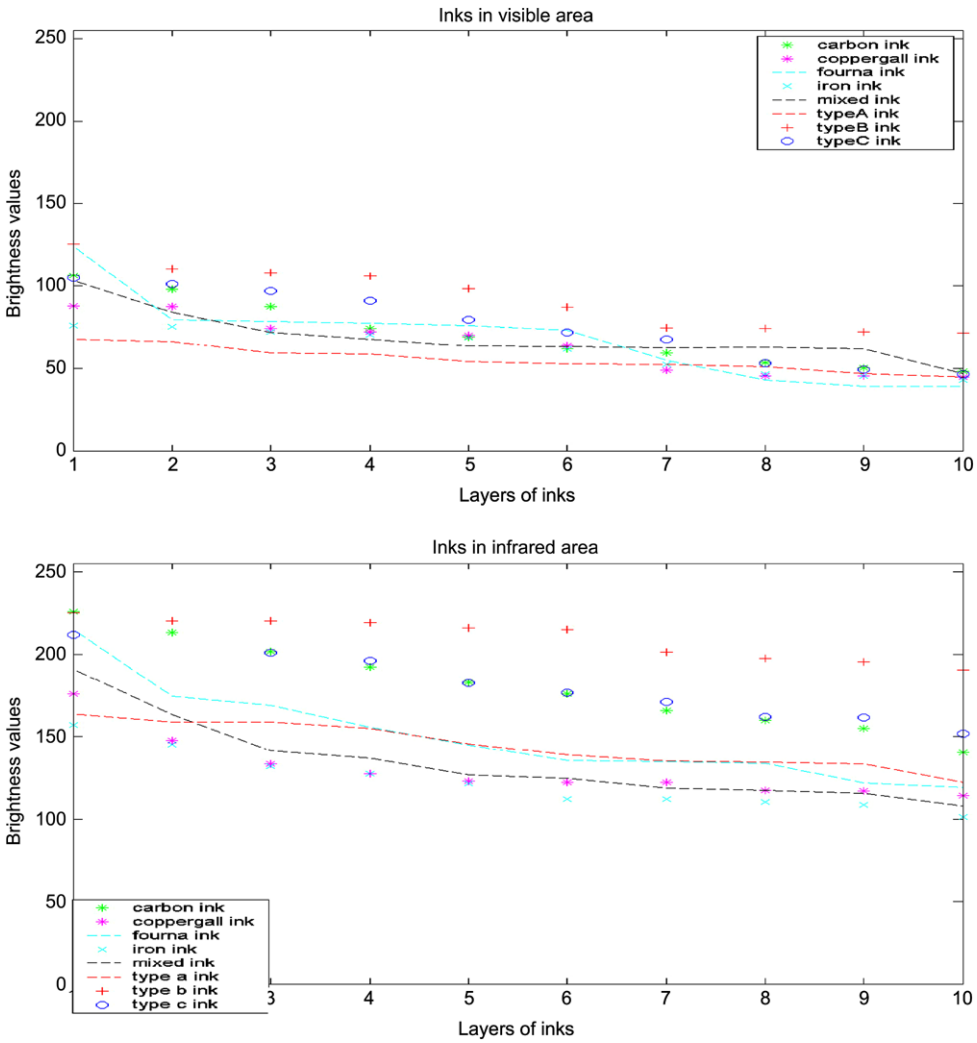

Fig. 2 Average intensity values for the eight inks in each of the ten different densities. *Top figure* shows the averages values in visible and the *bottom figure* in near-infrared

naked eye, and yet visible to an NIR imaging system. Figure 4 shows the co-occurrence matrices of the eight model ink types in visible spectrum where difference between the ink types is clearly exhibited. However, as it is shown in Fig. 5 the co-occurrence matrix of the same ink type in NIR has higher variance compared to the co-occurrence matrix in the visible spectrum.

We use co-occurrence matrices of gray-level intensities (GLCM) to model these second-order statistics (Haralick et al. 1973). The GLCM matrix with entries $P_{delta}(i, j)$ for a displacement vector $\delta(\delta_x, \delta_y)$ is defined as the number of occurrences of the gray-level i, and j at a distance δ. In our feature descriptor we impose the conditions that when observed direction is $\Phi_c = 0$ then $\delta_x = 1$, $\delta_y = 0$, and when $\Phi_c = \frac{\pi}{4}$ then $\delta_x = 1$, $\delta_y = 1$.

A number of useful texture features is possible from the GLCM, and in practise contrast and entropy are the most informative for ink textures, as the others are to some degree not independent from each other. Contrast is a measure of the clearness of ink regions. It is also a measure of the amount of local variation in the ink image. A low value of contrast results from images of uniform ink, whereas ink images with large variation produce a high value. Entropy quantifies the amount of different image intensity value pairs in the GLCM. For example, minimum entropy relates to the highly peaked distribution of a smooth and liquid ink texture, and maximum entropy to flat distribution due to the generous amount of differently shaded details in a viscous ink texture.

3.2 Weighted Sums of Off-diagonal Bands

The co-occurrence matrix has the property that off-diagonal entries represent pair of intensities of a specific difference. For example, the off-diagonal of rows i and columns $i + 2$ are all intensities pairs with a relative difference of two gray-levels, regardless of the absolute intensity values.

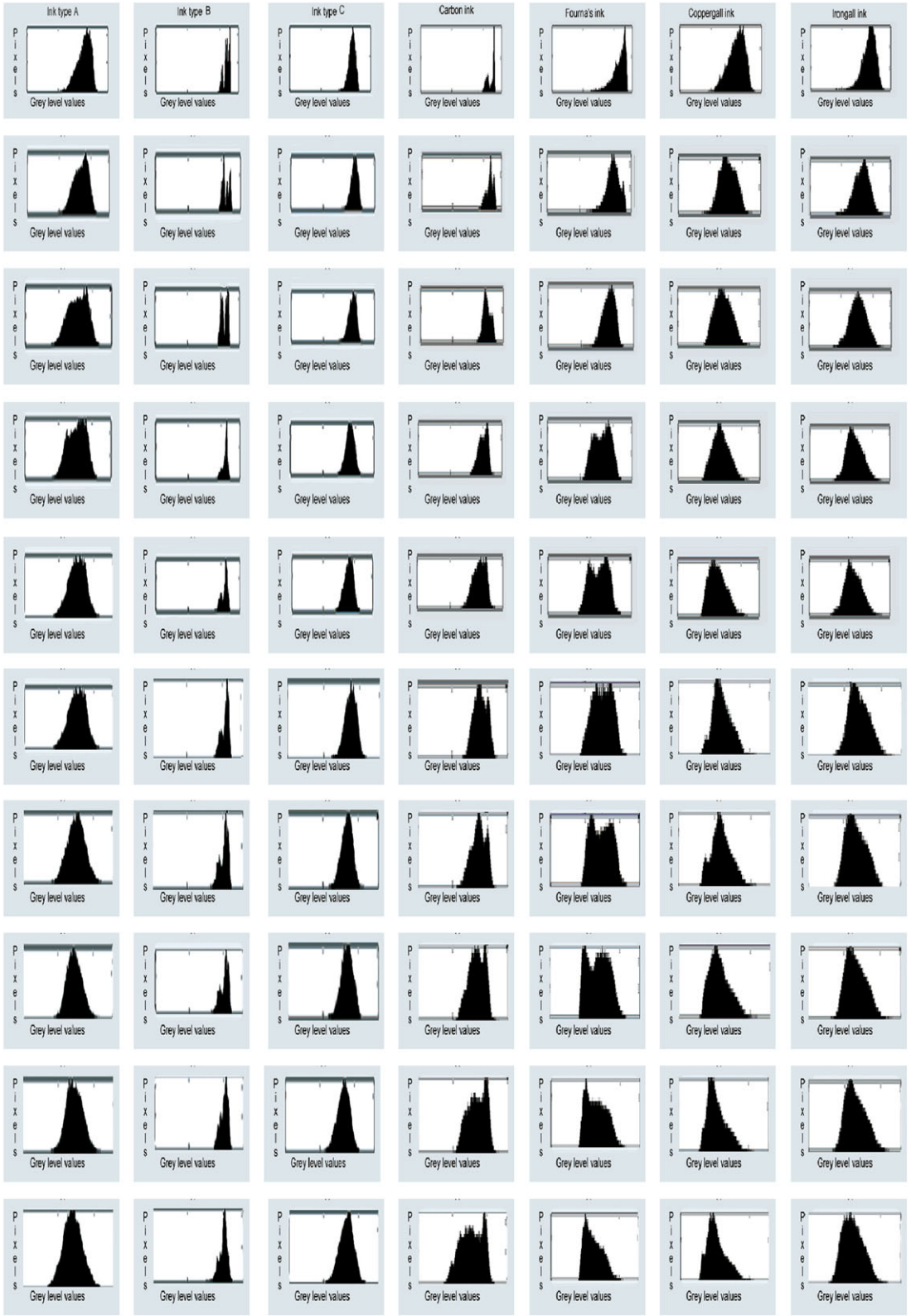

Fig. 3 Intensity histograms of the eight model ink types at ten different densities or thickness layers. *Columns* represent different ink composition, and *rows* represent different level of density for each ink. The *top row* represents the histograms after one layer of ink is applied, whereas the *bottom row* represents the histograms after ten layers of ink are applied. As can be noted in the figure, intensity distributions vary with ink type and to some degree with physical density, since as the number of ink layers increase the distributions tend to move towards the *darker areas* of the intensity range

Table 2 Statistical measures derived from NIR intensity distributions

Feature	Description
$\mu = \sum_{l=0}^{L-1}(b_l)\,p(b_l)$	Histogram mean
$\sigma^2 = \sum_{l=0}^{L-1}(b_l - \hat{b})^2\,p(b_l)$	Histogram variance
$\gamma = \sum_{l=0}^{L-1}(b_l - \hat{b})^3\,p(b_l)$	Skewness
$\beta = \sum_{l=0}^{L-1} 1 - \frac{1}{1+\sigma^2}$	Smoothness
$H_1 = -\sum_{k=1}^{L} p(b_l)\log_2 p(b_l)$	Histogram entropy

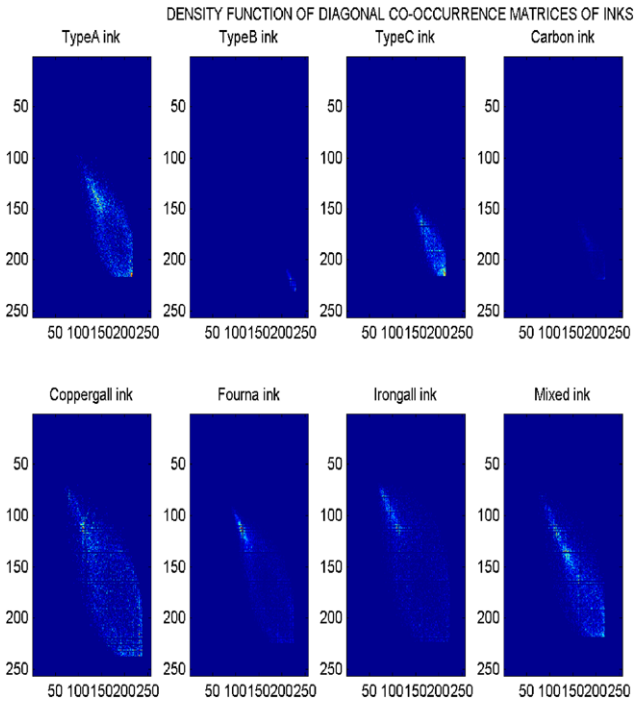

Fig. 4 An example of co-occurrence matrices for each of the eight model inks in the visible spectrum where the difference in the distribution is shown

The proposed set of four features from these statistics of different bands are,

$$\bigcup_{b=1}^{4}\left\{\sum_{w}^{2^{(b-1)}}\sum_{i=1}^{L-w} P_{i,i+w}\right\} \tag{1}$$

where, b represents the number of bands, and dummy variable w is the width of the band in off-diagonal units. Adding up entries of the same off-diagonal band is equivalent to create a texture statistic that is partially invariant to illumination intensity changes.

3.3 Co-occurrence Spectrum

For the second category of proposed features, we view the co-occurrence matrix as a collection of L-dimensional row vectors p_k, that is $P_{delta} = P_{delta}^T = [p_1, \ldots, p_L]^T$. Then,

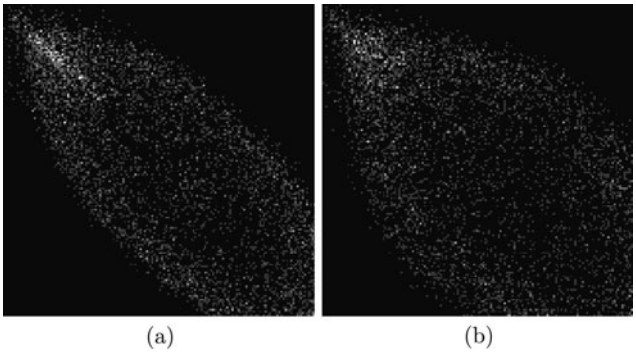

Fig. 5 Differences between the gray level co-occurrence matrices in the visible and near infra-red spectrum. (**a**) Grey-level co-occurrence matrix of an ink area of a carbon model image captured in the visible part of electromagnetic spectrum. (**b**) The corresponding co-occurrence matrix captured in the NIR spectrum

Table 3 Second-order statistical features derived from the co-occurrence matrices

Feature	Description
$\gamma_{\Phi_c} = \sum_{i,j}\{p(i,j)(i-j)^2\}$	Contrast (Φ_c rads)
$H_{\Phi_c} = -\sum_{i,j}\{p(i,j)\log_2 p(i,j)\}$	Entropy (Φ_c rads)
$\lambda_{\Phi_c}^{(i)} \in \Lambda_{\Phi_c} \Leftarrow Cov(GLCM_{\Phi_c})$	Eigenvalues
$S_{\Phi_c} = \bigcup_{B=0}^{4}\{\sum_{\delta=2^B}^{2^{(B+1)-1}}\sum_{i,j} p(i,j)\}$	Band Sums

the covariance matrix of the symmetric matrix P_{delta} in Fig. 5 provides information on the covariance of gray-level intensities with respect to all other neighboring intensities. The eigen decomposition of the covariance matrix provides a compact description of the intensity spread spectrum in terms of eigenvalues, one for each gray-level.

$$Cov(P_{delta})\,\Sigma = \Sigma\,\Lambda \tag{2}$$

where Σ is an orthonormal matrix of eigenvectors, and Λ is an $L \times L$ diagonal matrix of eigenvalues. The first six largest eigenvalues are then retained as features. Second order statistical features are listed in Table 3.

3.4 Ink Descriptor Space

The intensity histogram statistics and co-occurrence statistics are concatenated into 29-dimensional ink descriptors ink_n,

$$ink_n = \{\mu, \sigma^2, \gamma, \beta, H_1, \gamma_0, H_0, \lambda_0^{(i)}, S_0, \gamma_{\frac{\pi}{4}}, H_{\frac{\pi}{4}}, \ldots\} \tag{3}$$

where the first five descriptors represent the statistical measures derived from the NIR histogram distributions and the remaining represent the statistical features derived from the co-occurrence matrices.

Creating an augmented descriptor by combining intensity and co-occurrence type of descriptors poses some problems.

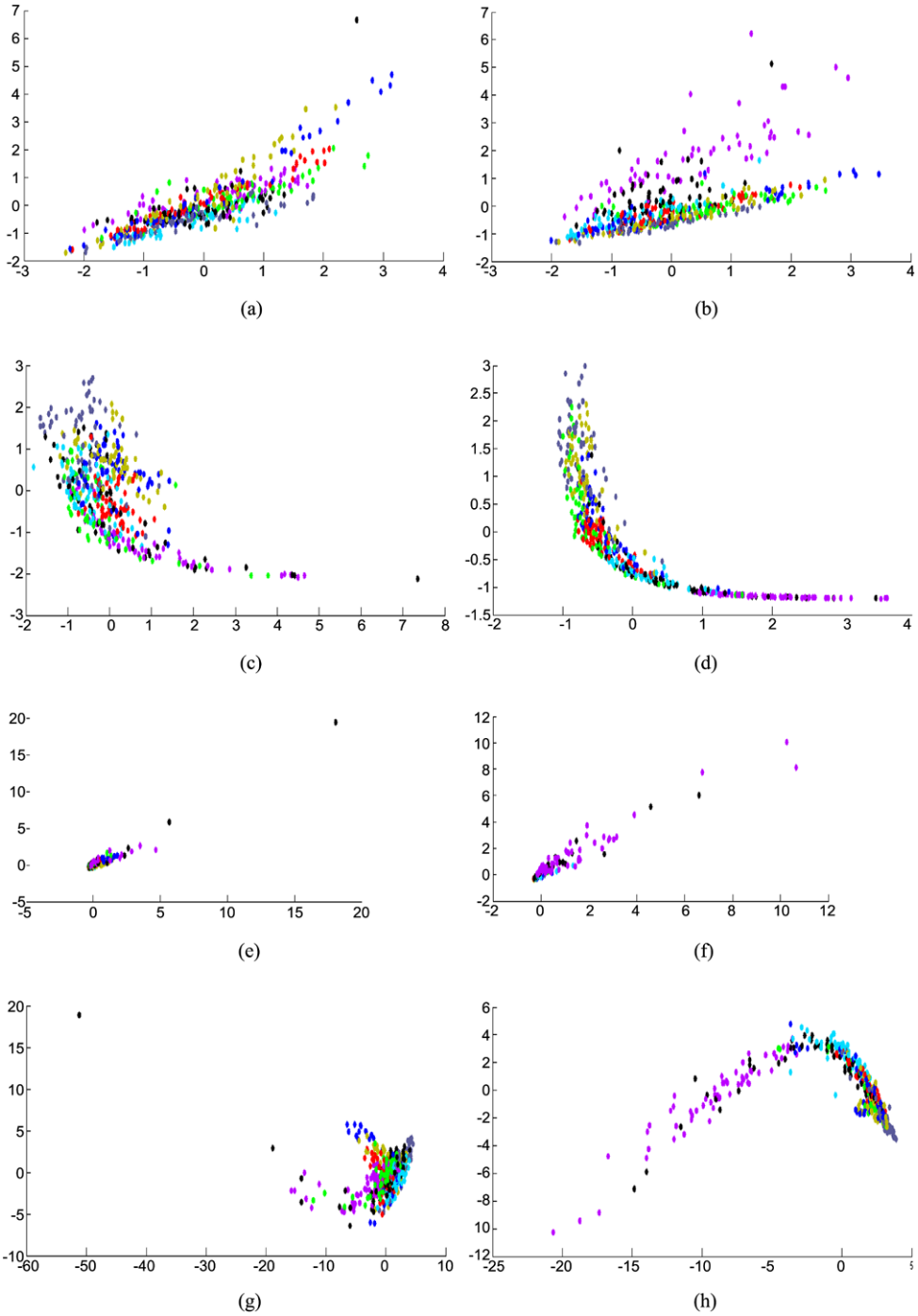

Fig. 6 Ink descriptor distribution in feature space. *Left column*: Histogram mean against histogram variance (**a**), entropy (**c**), covariance eigenvalues (**e**) and band sums (**g**) in the visible spectrum. *Right column*: corresponding plots in NIR spectrum

Combined descriptors have more difficult physical interpretation. In addition the features are derived from disparate intensity and co-occurrence family of descriptors whose values lie at scales not directly comparable to each other, requiring scale normalization. As the independence assumption of each dimension might be invalid, and dimensions are scaled differently when embedded in a subspace, we employ a multi-dimensional optimization of eigen-entropy function over scale.

In order to gain an understanding on the nature of the feature space captured by the ink descriptors we used text images written with eight inks of known composition. Figure 6 shows the histogram mean plotted against the histogram variance, entropy, covariance eigenvalues and band sums in both the visible and the NIR spectrum. Each of the eight ink models is represented in a distinct color. One characteristic of the ink descriptors ink_n is that they form clusters that are non-linearly distributed in feature space and form a

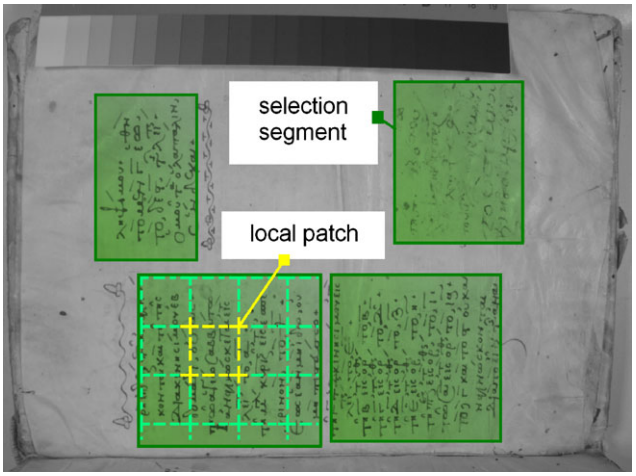

Fig. 7 Manuscript image, four segment selections of ink areas, and local patches from sliding window (*dash squares*)

manifold embedded in higher-dimensional feature space. Of interest is the spread noted in the NIR, and in particular the band sums of the co-occurrence matrix.

The first and second order statistical features previously described are extracted from samples of different image local patches containing ink (see Fig. 7). A sliding window breaks up the patches in smaller and slightly overlapping sub-patches, which makes the features invariant to mild image rotations and translations caused by the miss-alignments during document capture. The sub-patches have the property of being sufficiently small as to capture local structure but large enough to include amounts of ink pixels such that the requirements of the central limit theorem are met. Imposing these restrictions results in more robust ink statistics, as well as conveniently enforcing Gaussian distribution on the data.

4 Ink Clustering

Our modelling assumption is that we can model the ink space using a mixture of K Gaussian clusters of ink descriptors, where K is determined using Minimum Description Length. Each cluster is best explained by their distance to a centre of gravity μ_k, mass density w_k and ink appearance attributes with a sphere of influence directly related to the eigenvalues λ_l of the descriptor covariance matrix Σ_k. Another assumption is that all manuscript pages are explained by a single and shared computational model.

4.1 Ink Appearance and Descriptor Clustering

The ink found on a section of a manuscript is characterized by ink descriptors generated by clusters whose parameters $\Theta = \bigcup_{k=1}^{K}\{\mu_k, \Sigma_k, w_k\}$ maximize the likelihood $L(\Theta|Ink)$ of observing the set of all ink descriptors $\bigcup_{k=1}^{K} Ink_k$. We formulate the problem in the form of an objective function $E(Ink, \Theta)$ for the ink IR appearance cluster parameters that best explain the observed ink descriptors.

$$E(Ink, \Theta) = -\log\{L(\Theta|Ink)\} \tag{4}$$

We solve for the optimal cluster parameters with a simple Expectation-Maximization iterative estimation procedure (Bishop 1995), and store their description length value K_j. The procedure iteratively search for the overall K^* minimum (*Minimum Description Length*) over different values of plausible number of clusters K_j. For each selection segment S_q marginally overlapping local image patches W_r are sampled from a sliding window moving from top left to bottom-right of segment bounding box (see Fig. 7). The size of the patch W_r has to be large enough to estimate some textural statistics. We estimated the minimum size W_s to be at least L^2 pixels, where typically $L = 256$, that is the maximum number of gray-levels.

4.2 Probabilistic Voting of Ink Selection Segments

The posterior probabilities of cluster membership of descriptors $ink_r = \Psi(W_r)$ local patches W_r of a segment selection $S_q = \bigcup_{r=1}^{R}\{W_r\}$ are accumulated into K^* fractional bins, one for each candidate cluster.

The label l_i of the dominant cluster $k*$ in a segment S_q is found by probabilistic voting from the posterior probabilities:

$$l_q = \arg\max_k \left\{ \sum_{W_r \in S_q} p(\theta_k|ink_r) \right\} \tag{5}$$

This probabilistic voting procedure is equivalent to assign a label to an entire ink selection segment S_q, so that it can be treated like a discrete quantity, similarly to the idea of bag of words (see Fig. 8). Note that label $l_q \neq \arg\max_k p(\theta_k|\Psi(S_q))$. The resulting labelling of the segments is further exploit to build statistics over the proportions of ink appearance clusters on the entire set of manuscript pages, as explained in the next section.

4.3 Comparing Ink in Manuscripts Images

There are times when we wish to select a number of ink areas from carefully chosen manuscript pages, so to analyze and unveil ink appearance similarities. A selected ink area (segment) of a page has the property that the most influential cluster determines its ink appearance attributes. These attributes are summarized by the predominant cluster label l_i, and the distribution of the segments' labels characterizes the manuscript ink appearance. The label distribution obeys a PDF $p(l_k)$ approximated by the normalized label histogram whose bins are computed as,

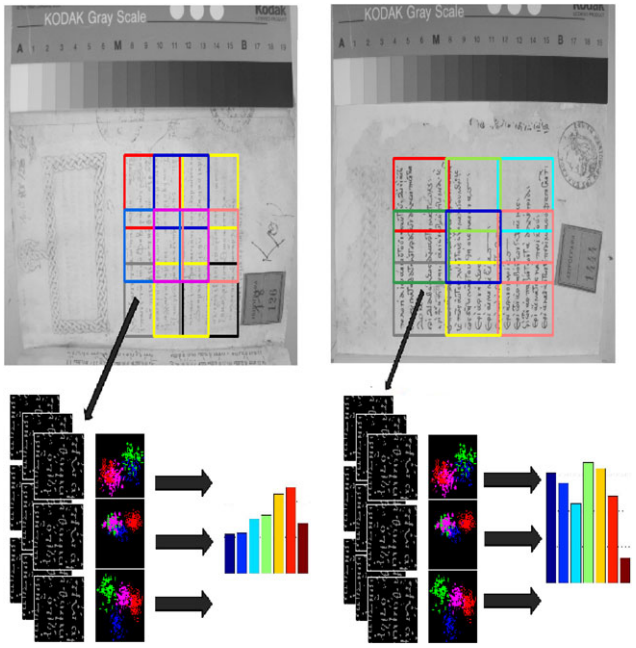

Fig. 8 Voting algorithm executed on local patches from a candidate ink selection segment. Test (*left*) and model (*right*) manuscripts are partitioned in several patches of ink pixels. For each such patch segmentation identifies the ink pixels and descriptor features are extracted to generate cluster probabilities. Voting consists in picking and accumulating the dominant cluster index in each patch into the final document histogram of ink labels

$$p(l_k) = \frac{\sum_{i=1}^{N_s}\{l_i(1-\min(l_i-k,1))\}}{N_s}, \quad \text{where } 1 \le l_i \le K \tag{6}$$

where N_s denotes the total number of segments in all manuscript pages.

The ink similarity between two images of manuscript pages is defined as follows,

$$SL(Ink_1, Ink_2) = 1 - \left\{\sum_{k=1}^{K} |p(l_k|Ink_1) - p(l_k|Ink_2)|\right\} \tag{7}$$

To largely similar ink descriptor sets Ink_1 and Ink_2 correspond a very small histogram difference of the label PDFs. Note that the order of label bins is irrelevant. Figure 9 gives a summary of the algorithm followed from the segmentation of text to the similarity measures of feature descriptors in the manuscripts.

5 Experiments

Ink found on the manuscript must be correctly separated from the support such as the background paper, parchment, or papyrus before feature extraction. Image acquisition of ink documents consistently took place in controlled laboratory conditions at similar color temperature, scale, position and orientation of the illuminants.

The images in the near infrared radiation were recorded using two tungsten photolamps, the 093 B+W optical filter and a CCD camera sensitive between from 380 nm to 900 nm. Each image included the manuscript page, a standard black and white scale and a colour scale. All images were captured with the same grey and colour scales with a resolution of 4256 × 2848 pixels. In addition to the above requirements all images were acquired with the following protocol regarding the positioning of the lights and the camera:

- The camera lens should be parallel to the page surface that should be completely flat.
- The lights should be in 45 degrees, to avoid the creation of shades.
- The distance between the lens and the page surface is the same for all images.
- During all acquisitions the same lighting should be used and the lights should be positioned in the same distance from the object.

Therefore our intensity normalization method is simpler than the one taken with more complicated document surface conditions as in Shi and Govindaraju (2004). At first, light intensity was normalized using a 3-D plane-fitting algorithm to correct gradients introduced by the illuminants. Later a simple piece-wise linear interpolation using a Kodak gray card was found to be more effective in correcting the gradients. The greylevels on the card help to create a lookup table for mapping ranges of gray tones. The last step, was to employ a local adaptive thresholding similar to the document binarization algorithm proposed by Sauvola and Pietikainen (2000) was employed in order to segment the low-contrast NIR images. A low-contrast image of inks is difficult to segment from the support, and the chosen approach resolves this issue by computing a different threshold $T(x, y)$ for each pixel by taking into account the local mean, $m(x, y)$, and local standard deviation $s(x, y)$ of the intensity in the neighborhood of that pixel at (x, y). This is computed with the equation $T(x, y) = m(x, y)[1 + k(s(x, y)/R - 1)]$, where R is the dynamic range of the standard deviation, and is set to $R = 128$. Parameter k is user generated and application dependent and set at $k = 0.5$.

Some of the results of our segmentation process is shown in Fig. 10. The images used during our experiments can be separated to those of known chemical composition and include both model and test images and those of unknown chemical composition that were taken directly from Medieval manuscripts.

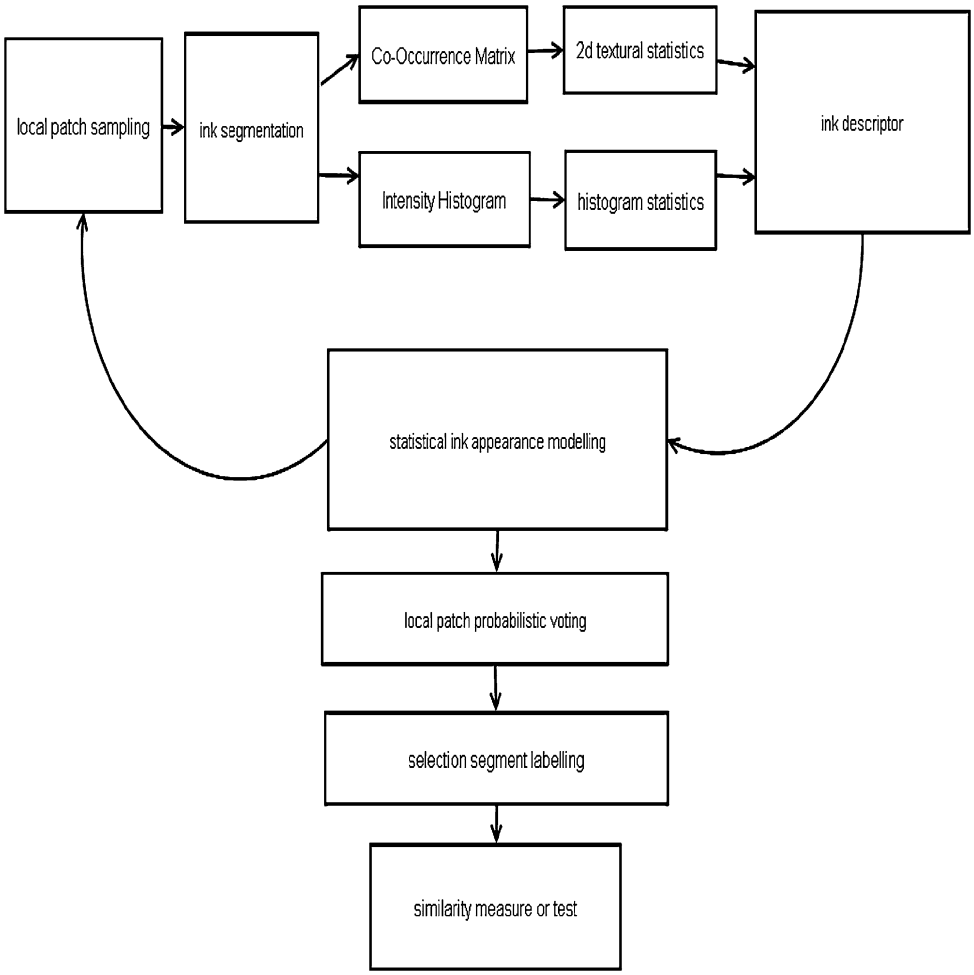

Fig. 9 The unsupervised clustering algorithm. The flowchart shows all the steps from text segmentation to the similarity measure of the feature descriptors

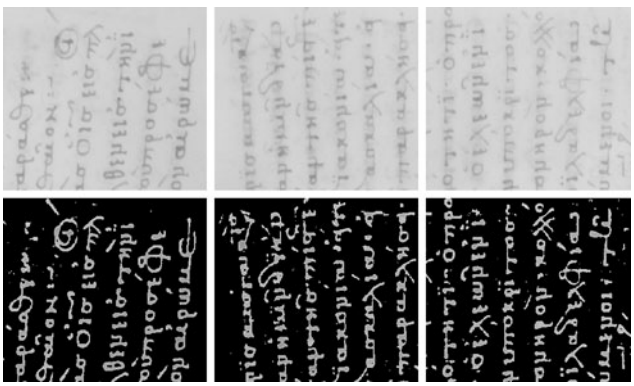

Fig. 10 The ink from the images was segmented and was used to build the intensity histograms and co-occurrence matrices from where the inks descriptors are derived

5.1 Images of Known Chemical Composition

We test the ink descriptor performance with the ink dataset used in Kokla et al. (2007). They were created using the eight model inks described in Section 2 and reflect the scripting conditions found in manuscripts and encapsulate:

- The varying thickness of the inks during scripting.
- The varying scripting formed due to the different means of writing used, such as quill, calamus and penna.
- The writing characteristics of different authors.

A total of 480 writings on paper were captured with a near-infrared camera. Hereafter we refer to these as the model images. All images were captured with same optics, and under consistent capturing conditions.

Model images of each ink type included all the capital and lower case letters of the Greek alphabet. In order to take into account the different writing styles, separate images were produced using quill, calamus and penna. In order to take into account the thickness of the inks, each of the inks were used in ten different thickness to create 80 model scripts (8 inks by 10 thickness of each). The ten different thickness of ink were created by first writing ten different samples of scripts. All but one of these samples were overwritten. Then all of but one of the overwritten samples were again overwritten. The procedure was repeated until the final sample alone was constructed of ten layers. The process was repeated with each of the eight model inks and a total

of 480 images (8 inks × 10 layers × 3 pens × 2 case letters) of the Greek alphabet were created.

Figure 10 shows examples of the images produced using 1 to 10 layers of model inks. Figure 11(a) shows the 10 layers of one of the model inks in low case letters. One can observe how each of the ten layers increases in ink density from top to bottom. Figure 11(b) shows the different variations that can occur depending on the different means of writing used.

5.2 Feature Descriptor Performance

The performance of the augmented feature descriptor is tested using the model ink images. A multi-layer perceptron (MLP) classifier (Bishop 1995) is trained on features from 240 model images and tested on the remaining 240 images. The proposed augmented feature descriptor discussed in Section 3 is tested by comparing its performance against four other descriptors: intensity values, co-occurrence sum of diagonals, Haralick co-occurrence descriptors (entropy, contrast, 2d energy) and first order histogram statistics.

Fig. 11 Test images. (**a**) The same ink was used to write ten layers of the Greek alphabet at ten different densities. (**b**) Test images capture the writing behavior of an author as it is influenced by pen type and letter size. Penna type (p) often results in accidental ink spills and larger spread areas, as opposed to the terser kalamus (k) and feather (f)

Figure 12 shows the recall performance of all five feature descriptors. The proposed ink feature descriptor of first and second order statistics expanded to eigen-co-occurrence and off-diagonal band features outperforms all other descriptors in all but one type of ink. Notice how all descriptors perform worse against ink type 4 (i.e. carbon ink) given this composition absorbs larger levels of light intensity even in the NIR part of the electromagnetic spectrum.

Table 4 shows the recall, and precision rates for various ink images recognized by the classifier trained on the ink augmented ink descriptor proposed. Trained images are classified correctly most of the times, and test images 74% of the times on average across ink type recipes. The results

Table 4 Performance of classifier trained with ink descriptors for NIR images. The columns correspond to ink composition types, and the vertical axis is the corresponding true positive rate using the test classifier. 1, 2, 3 = incomplete irongall, 4 = carbon, 5 = fourna, 6 = irongall, 7 = metalgall, and 8 = mixed inks. Notice that recall percentage for ink 4 (Carbon) is only 40% due to its high level of NIR absorbance

	1	2	3	4	5	6	7	8
Recall(TPR)	.77	.87	.80	.40	.77	.70	.80	.83
TNR	.97	.97	.95	.95	.96	.98	.99	.94
Precision	.79	.81	.69	.55	.72	.81	.92	.66

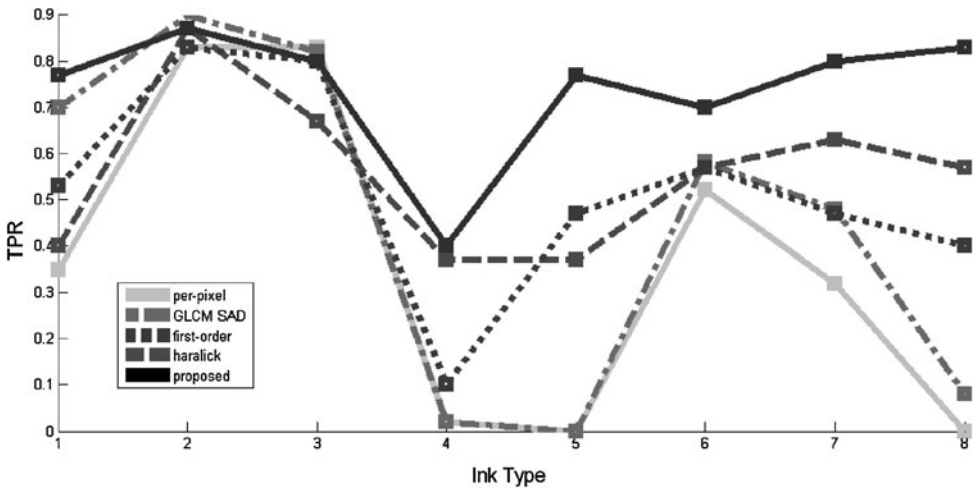

Fig. 12 Comparison of ink texture descriptors. The *horizontal axis* represent ink composition types: 1, 2, 3 = incomplete irongall, 4 = carbon, 5 = fourna, 6 = irongall, 7 = metalgall, and 8 = mixed inks. The *vertical axis* represents the recall performance of the descriptors for each ink. The recall performance for the descriptor based only on pixel intensities is very poor and the descriptor finds it difficult to discriminate between inks. Co-occurrence SAD, First-order, and Haralick descriptors are somewhat more discriminative than per-pixel intensities. The proposed descriptor combines eigen features, off-diagonal bands, first and second order statistics to create a richer representation of the ink type distributions and has the best recall performance between all descriptors for all inks

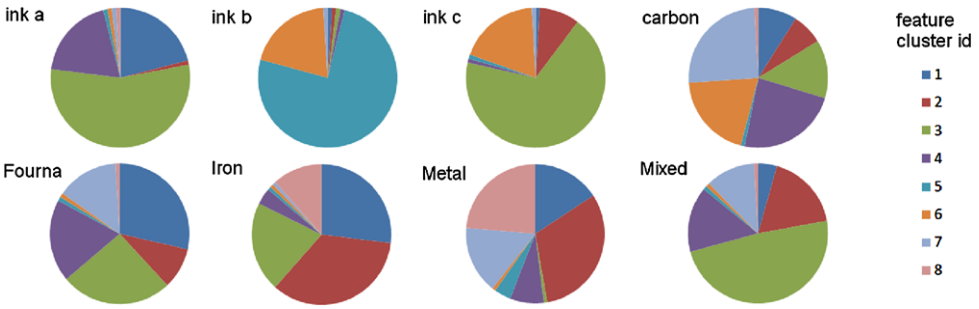

Fig. 13 Each of the eight inks are represented by a pie chart. Slices in the pie charts represent the proportion of each of the high level features contributing to each of the ink types. Different inks have different image-based properties and therefore different proportions of high level feature labels

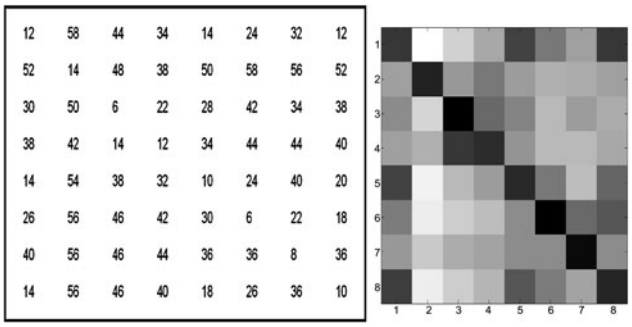

Fig. 14 Confusion Matrix. Difference in label distribution between pairs of ink types

show that descriptors for carbon ink are the most difficult to discriminate, even in NIR spectrum.

5.3 Unsupervised Clustering of Ink Models

The algorithm for unsupervised clustering described in Section 4 is first tested on the images of known chemical composition. The clustering of images of known inks is used to raise confidence in the ink appearance clustering of the model images, which in turn prove that the visual comparison of manuscript ink is possible, and meaningful. Unsupervised clustering assigns labels to each model image. We test the hypothesis that test images from the same model ink are likely to be labelled similarly.

As described in Section 4 the minimum number of clusters of each dataset is determined by using minimum description length. In the case of the model inks, the optimum number of clusters found to describe the dataset is eight, and each of these clusters is seen to correspond to a high level feature in the descriptor space. The key result of these tests is that different inks have different image-based properties and therefore different proportions of high level features labels. These proportions is what makes each ink type unique, and the variability of each ink with respect to the percentage of high level features is shown in Fig. 13.

A similarity distance function based on the difference in the label distribution of two compared pages provides insight of how two inks differ with respect to the clusters. We can see from Fig. 14 that the distance between images of same ink type is always smaller than the distance between different model ink images.

5.4 Comparison of Manuscript Images

In order to test the algorithms on manuscript inks, NIR images were selected from the on-line EU NOESIS database. The manuscripts and pages were selected by the NOESIS partners. The tasks performed were set by a preservationist that specializes in manuscript inks and included the comparison of sections of pages from the same or different manuscripts. The aim of all the tasks performed were to confirm the hypothesis of the preservationist regarding the similarity or not of the selected sections. As a qualitative interpretation sections that were given similarity measure between 0%–33% are considered as been written using different types of ink, whereas any similarity measure over 60% indicates that the same type of ink has been used. Test performed on manuscripts are summarized in Tables 5 and 6. Results in Table 5 focus on the comparison of sections in the same manuscripts, whereas results in Table 6 aim to indicate that pages from the same manuscript (which preservationists indicated that are written with the same scriber or ink) give higher similarity measures, compared to pages from different manuscripts. Images of the manuscripts used are shown in visible and NIR in Figs. 15 to 22. The selection in the infrared images indicate the sections of the images that were selected for comparison. The caption in each of the figures provides an explanation on the selection of the manuscripts and the hypothesis. As seen from the images and tables the results obtained by the similarity measures for images in Figs. 15 to 20 confirm the hypothesis of the preservationist, whereas the results for the images in Figs. 21 and 22 show examples where additional information and analysis is required to deduce the results.

Table 6 summarizes similarity measures for pages from manuscripts GLNR126 and GLNR666. In accordance to the hypothesis segmented text from the two manuscripts show small similarity and share in most cases 0% to 43% of ink characteristics with the exception of pages 52 from GNRL666 and 100 from GNRL126 where 73% of the segmented text have similar ink characteristics.

Table 5 Similarity measure results for pages (folios) from the same manuscripts. The pages are selected by preservationist in the field and the hypothesis was based on studies on the scripting and historical information in the manuscript. Similarity measures are given by the NOESIS system based on the unsupervised clustering algorithm

Manuscripts	Page comparisons	Preservationist's hypothesis	Similarity measure
GRNL666	52 vs 108	Page 52 is from	6%
	52 vs 180	different era	11%
	180 vs 108		78%
GRNL126	100 vs 50	Same ink	67%
GRNL2117	50 vs 2v	Script and color different	20%
GRNL2117	223	Undecided	0%
GRNL2536	1 vs 31	Similar inks difficult visual recognition	60%
GRNL2582	1 vs 73	Different inks	25%
GRNL72	190 vs 190v	Same inks	50%
GRNL194	3 vs 136	Different inks	25%
GRNL306	114 vs 149	Different inks	25%

Table 6 Similarity measure between pages from different manuscripts GRNL666, and GRNL126

	f.001	f.050	f.100	f.276
Folio 052	28%	73%	31%	43%
Folio 108	8%	0%	11%	17%
Folio 180	0%	11%	0%	0%

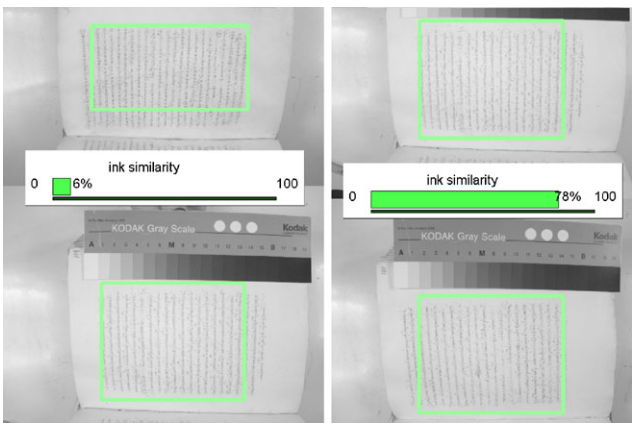

Fig. 15 Pages from manuscript GRNL666 are expected to be written with similar inks. Page 52 is dated by historians as being from a different era to the rest of the manuscript. This is confirmed by the similarity measure given by the NOESIS project (*left*). Pages 180 and 108 are dated from the same period (*right*)

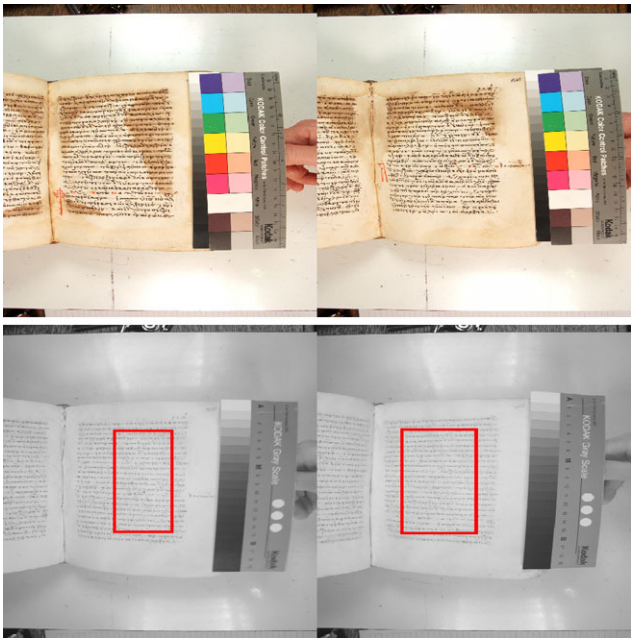

Fig. 16 The visible (*top*) and infrared images (*bottom*) of pages 50 (*left*) and 100 (*right*) of manuscript GRNL126. The scripting on the pages is the same but in visible spectrum the shade of the ink is different. Similarity measures indicate that the ink is the same, as hypothesized based on the historical information available

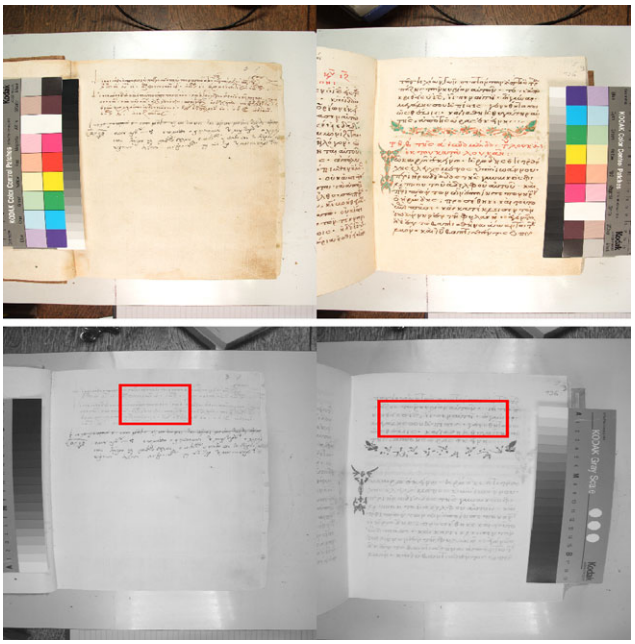

Fig. 17 The visible (*top*) and infrared images (*bottom*) of pages 3 (*left*) and 136 (*right*) of manuscript GRNL194. The selected sections show different scripting style, and the similarity measure of 25% confirms the difference in ink type

6 Conclusions

We have introduced a statistically-based feature descriptor for different ink compositions. The descriptor is enriched

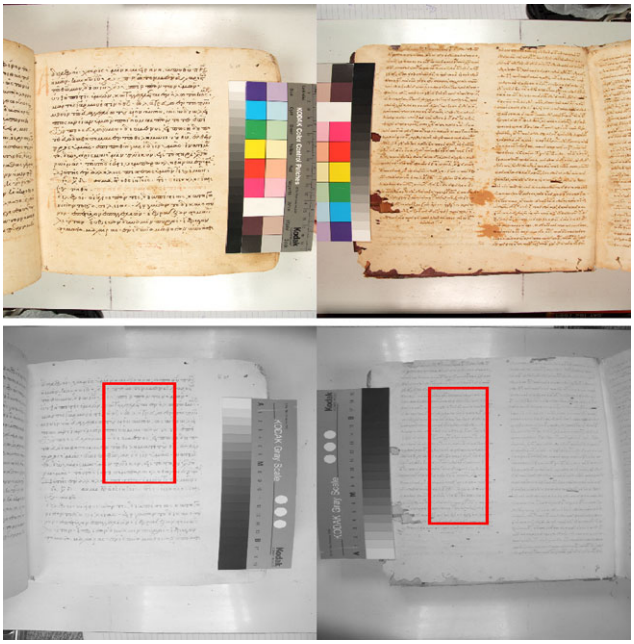

Fig. 18 The visible (*top*) and infrared images (*bottom*) of pages 50 (*left*) and 2 verso (*right*) of manuscript GRNL2117. The scripting and the shade of the ink in visible spectrum is different. The similarity measure of 20% confirms the difference in ink type

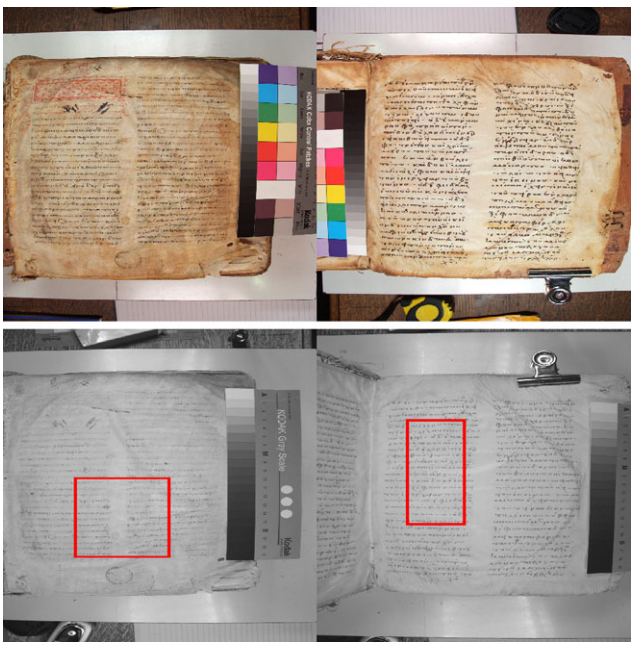

Fig. 19 The visible (*top*) and infrared images (*bottom*) of pages 1 (*left*) and 313 (*right*) of manuscript GRNL2536. The inks are the same but some of the writings were overwritten and some have been added at the end. The similarity measure of 60% confirms that the inks used is the same

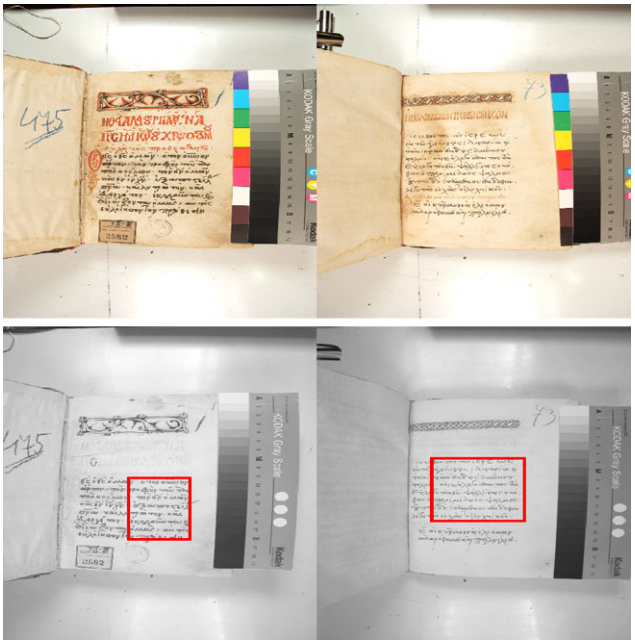

Fig. 20 The visible (*top*) and infrared images (*bottom*) of pages 73 (*left*) and 1 (*right*) of manuscript GRNL2582. The scripting is different and possibly the ink is different. The similarity measure returned is 26%

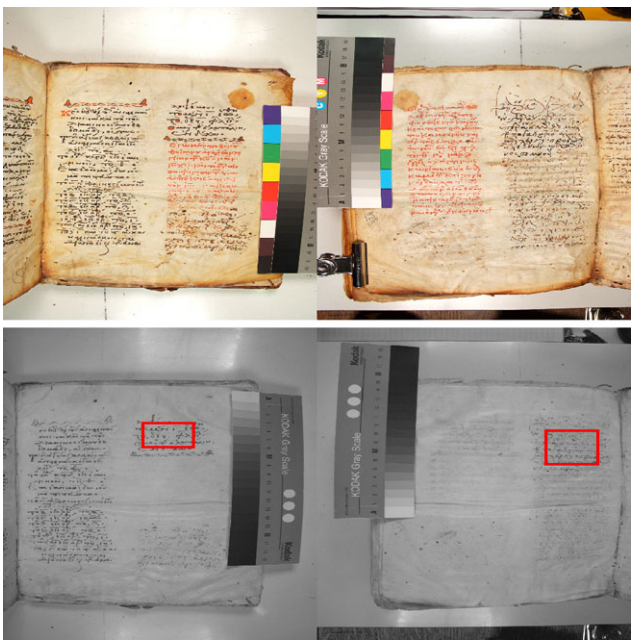

Fig. 21 The visible (*top*) and infrared images (*bottom*) of pages 190 (*left*) and 190v (*right*) of manuscript GRNL72. The hypothesis is that the inks are the same, however the faded inks made it difficult to confirm similarity. The selected sections in the infrared images give a similarity measure of 50%. This indicates that the ink used in the two sections is similar but this is not a strong result

by first-order and second-order statistics, and in particular we have introduced two new texture features categories. The first one based on weighed off-diagonal bands and the second on eigen decomposition of the covariant matrix of local joint intensity co-occurrences. The resulting feature descriptor has the disadvantage of being high-dimensional, and in

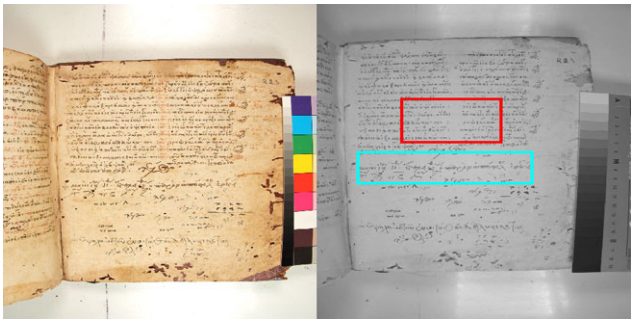

Fig. 22 The visible (*left*) and infrared images (*right*) of page 223 of manuscript GRNL2117. This is an example where the preservationist cannot tell with certainty whether the selected inks on the same page are the same or not. The similarity measure is 0% giving a strong indication that the inks are not the same and therefore probing further investigation

future work it would be interesting to discover and remove redundant descriptor dimensions with a suitable feature selection algorithm. On the other hand, the advantage of the proposed ink feature descriptor is that it is suited to discriminate among ink compositions in the NIR part of electromagnetic spectrum as the comparison tests show. It was tested by training suitable non-linear classifier and studying the discriminative power of the texture features. We also demonstrated that the second-order statistical nature of the features allows the descriptor to discriminate among ink texture of different chemical composition, a task the human visual system finds extremely challenging to accomplish given the small inter-class variance, which results in different ink composition surfaces to be perceived as identical. The descriptor was also tested with manuscripts from the NOESIS database, where the hypothesis of preservationist was confirmed by the similarity measure results given by our algorithm. Extensive comparisons are currently carried out in the NOESIS database in order to capture statistical results on the similarity measures across manuscripts of the same and different eras. In addition we are working in reducing the manual selection required for the comparison of sections of manuscripts.

Acknowledgements We would like to thank the EU for supporting this work under the project EU-Noesis, Non-dEStructive Image-based manuscript analysis System (research contract no. 509145, 6th Framework)

References

Alexopoulou, A., & Kokla, V. (1999). Physicochemical study of inks in manuscripts using uv and ir radiation. In *6th international conference on non-destructive testing and microanalysis for diagnostics and conservation of cultural and environmental heritage* (pp. 2049–2056).

Barrow, W. (1972). *Manuscripts and documents*. Charlottesville: University Press of Virginia.

Bishop, C. (1995). *Neural networks for pattern recognition*. London: Oxford University Press.

Brown, K., & Clark, R. (2004). Analysis of key anglo-saxon manuscripts (8–11th centuries) in the British library: pigment identification by Raman microscopy. *Journal of Raman Spectroscopy*, *35*, 181–189.

Clarke, M. (2001). The analysis of medieval European manuscript. *Reviews in Conservation*, *2*, 3–17.

Coggins, J., & Jain, A. (1985). A spatial filtering approach to texture analysis. *Pattern Recognition Letters*, *3*, 195–203.

Dasari, H., & Bhagvati, C. (2007). Identification of non-black inks using hsv colour space. In *Ninth international conference on document analysis and recognition (ICDAR 2007)* (Vol. 1, pp. 486–490).

Farrokhnia, F. (1990). *Multi-channel filtering techniques for texture segmentation and surface quality inspection*. PhD thesis, Computer Science Department, Michigan State University.

Flieder, F., Barroso, R., & Orvezabal, C. (1975). Analysis des tannins hydrolysable susceptibles d' entrer dans la composition des encres ferro-geliques.

Franke, K., Bunnemeyer, O., & Sy, T. (2002). Ink texture analysis for writer identification. In *Eighth international workshop on frontiers in handwriting recognition* (pp. 268–273).

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*, 610–621.

Huang, Y., Brown, M., & Xu, D. (2008). A framework for reducing ink-bleed in old documents. In *Conference on computer vision and pattern recognition (CVPR'08)* (pp. 1–7).

Janssens, K., Vittiglio, G., Deraedt, I., Aerts, A., Vekemans, B., Vincze, L., Wei, F., Ryck, I. D., Schalm, O., Adams, F., Rindby, A., Knöchel, A., Simionovici, A., & Snigirev, A. (2000). Use of microscopic xrf for non-destructive analysis in art and archaeometry. *X-Ray Spectrometry*, *29*, 73–91.

Kokla, V., Konstantinou, V., Psarrou, A., & Alexopoulou, A. (2000). Towards the creation of generalised computational models for the characterisation of inks used in byzantine manuscripts. In *15th world conference on non destructive testing* (pp. 169–175).

Kokla, V., Psarrou, A., & Konstantinou, V. (2007). Ink discrimination based on co-occurrence analysis of visible and infrared images. In *Proceedings of the ninth international conference on document analysis and recognition, IEEE (ICDAR'07)*, Curitiba, Brazil.

Lee, A., Mahon, P. J., & Creagh, D. C. (2006). Raman analysis of iron gall inks on parchment. *Vibrational Spectroscopy*, *41*, 170–175.

Malik, J., Belongie, S., Leung, T., & Shi, J. (2001). Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, *43*(1), 7–27.

Monique, D. P. (1975). *Tat des travaux effectues sur l' analyse des consititues des encres noires manuscripts par deux technique: chromatographie sur couche mince et electrophorese* (Tech. rep.). Comite pour la conservation de l' ICOM.

Picard, R., Elfadel, I. M., & Pentland, A. P. (1991). Markov/Gibbs texture modeling: aura matrices and temperature effects. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 371–377).

Sauvola, J., & Pietikainen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, *33*(2).

Shi, Z., & Govindaraju, V. (2004). Historical document image enhancement using background light intensity normalization. In *Proceedings of the pattern recognition, 17th international conference on (ICPR'04)* (Vol. 1, pp. 473–476).

Varma, M., & Zisserman, A. (2005). A statistical approach to texture classification from single images. *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, *62*(1–2), 6–81.

Zerdoun Bat-Yehouda, M. (1983). *Les encres noires au moyen age (jusqu'a 1600)*. Paris: Editions du CNRS.