

UNIVERSITY OF WESTMINSTER



WestminsterResearch

<http://www.wmin.ac.uk/westminsterresearch>

Ink discrimination based on co-occurrence analysis of visible and infrared images.

Vasiliki Kokla
Alexandra Psarrou
Vassilis Konstantinou

Harrow School of Computer Science

Copyright © [2007] IEEE. Reprinted from the Proceedings of the Ninth International Conference on Document Analysis and Recognition, 23-26 Sept. 2007 (ICDAR 2007). IEEE, Los Alamitos, USA, pp.1148-1152. ISBN 9780769528229.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Westminster's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of the University of Westminster Eprints (<http://www.wmin.ac.uk/westminsterresearch>).

In case of abuse or copyright appearing without permission e-mail wattsn@wmin.ac.uk.

Ink Discrimination based on Co-occurrence Analysis of Visible and Infrared Images

Vasiliki Kokla, Alexandra Psarrou, Vassilis Konstantinou
University of Westminster, Harrow School of Computer Science
Watford Road, Harrow HA1 3TP , UK
psarroa@wmin.ac.uk

Abstract

Inks found in Byzantine manuscripts are semi-transparent pigments and their examination and analysis provide an invaluable source of information on the authenticity and dating of manuscripts and the number of authors involved. However, inks are difficult to characterize because their intensity depends on the amount of liquid spread during scripting and the reflective properties of the support. Most existing methods for the analysis of ink materials are based on destructive testing techniques that require the physicochemical sampling of data. Such methods cannot be widely used because of the historical and cultural value of the manuscripts. In this work we show that manuscript inks can be represented through a mixture of Gaussian functions and can be characterised using co-occurrence matrices.

1 Introduction

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 physicochemical sampling of data. Such methods cannot be used widely because of the historical value of the manuscripts. Non-destructive techniques such as spectroscopy and reflectography are more suited to the study and conservation of old manuscripts, where the optical properties of the pigments are studied under illumination beyond the visible spectrum [3,7]. Most of the image-based research in inks and pigments found in artifacts is focused in the generation rather than analysis and are mainly applied in the restoration of colors in paintings. Alternatively research in machine vision is carried out in the analysis and modelling of color and mainly focuses on the visual retrieval of information in digital image libraries.

Our aim in this work is to use image-based techniques to complement the reflectographical methods of analysis by computing models and interpreting the visual properties of the inks in the visible and infrared areas of the spectrum, therefore providing an *in situ* and portable quantitative method for the discrimination of the types of inks used in old manuscripts. Inks, however, are semi-transparent pigments and difficult to characterize because their intensity depends on the amount of liquid spread during scripting and the reflective properties of the support. In this work we show that manuscript inks can be represented in the infrared area of the spectrum through a mixture of Gaussian functions and can be characterised based on correlations of image variations in the visible and infrared areas of the spectrum. Such variations are studied using co-occurrence matrices that capture the behaviour of the inks during the scripting process.

In the remaining of this short paper, in Section 2 we give a short description of the composition of inks that were used during this study and present the model and test images used during our experiments. In Section 3 we describe the co-occurrence matrices used in our experiments. In Section 4 we present our results before we conclude in Section 5.

2 Background

The inks used in this study date from the 11th to the 18th 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 standard of inks manufactured according to the recipes given in [8] to have a basis for comparison with unknown inks. 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

- Metalgall ink. This category contains the Coppergall inks and Irongall inks
- Incomplete inks. This group includes inks, that have a similar composition to that of metalgall inks, although their composition does not include all of the ingredients of metalgall inks and we treat them as subclasses of metalgall inks (type A, B and C).
- Mixed ink. This category contains inks that have ingredients of the last two categories.

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 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. Reflectographical studies on the optical behaviors of the inks under visible and infrared radiation have shown that inks that have very similar photometric properties under visible light can be separated when viewed under infrared radiation [1]. The differentiation is mainly due to the different chemical composition of the inks and can be represented using mixture of Gaussian functions as shown in Figure 1. However, even though there is a difference in the intensity distribution of the inks under infrared variation, this alone is not sufficient to discriminate between the different inks and therefore additional characteristics have to be taken into account. As inks are composed from different materials and bindings, assuming they are used on the same substrate, inks tend to spread differently during the scripting process. In this paper we study the behaviour of the inks during the scripting process by examining the image variations of the inks during scripting using co-occurrence matrices.

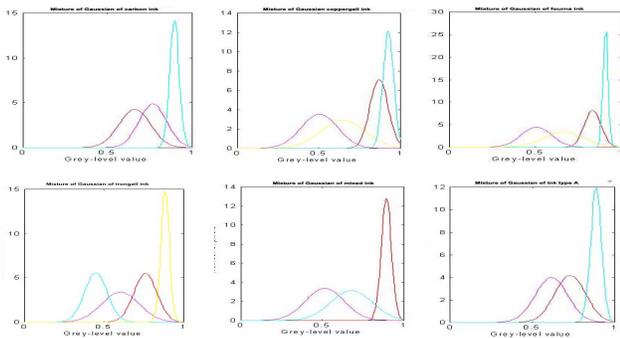


Figure 1. Intensity distribution curves using mixtures of Gaussians in the infrared area of spectrum. From Left to Right: Top row: Carbon, Coppergall, Fourn inks. Middle

2.1 Model and Test Images

During our experiments we created images to 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.

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 Byzantine manuscripts where alternative X-Ray Fluorescence spectrography (XRF)[7] method is employed to establish the ink composition used in the manuscript. This is performed in order to verify the results derived from the image-based technique. Figure 2 shows examples of model images produced using 1 to 10 layers of varying thickness inks during scripting. A total of 480 images (8 inks x 10 layers x 3 pens x 2 case letters) of the Greek alphabet were created. These were grey level images and included all categories of inks, writings produced by various script materials and different script styles.

Script of ten layers with low case letters

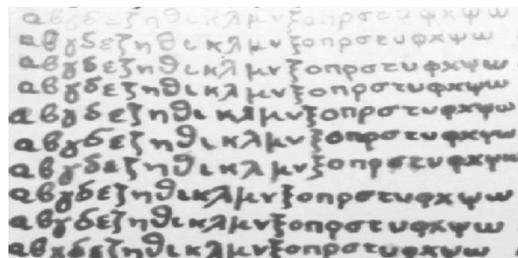


Figure 2. Examples of model images using 1 to 10 layers of inks.

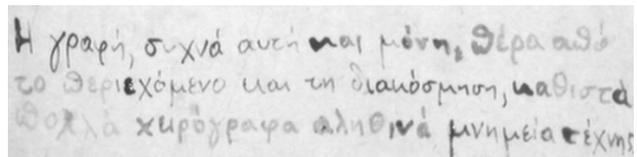


Figure 3. A sample script test image of known ink composition.

2.2 Digitization and Ink Segmentation

The digitization of the images in the visible and near infrared radiation has been performed using two tungsten photolamps, the 489 and 093 optical filters and a CCD infrared reflectograph sensitive up to 1200nm. A standard black and white scale of 14 grey level tones has simultaneously been recorded in each reflectogram in order to control the digitization conditions and acquisition parameters. A standard set-up regarding distance and position of the lamps has been used in order to minimize any variations due to distortions. The Expectation-Maximization algorithm is used to segment the ink areas from the manuscript support, by computing a mixture of Gaussian functions for the background and foreground distributions.

3 Co-occurrence of Image Variation

In the isolated ink areas of the images we apply co-occurrence matrices in order to compute the images variations that characterize each ink. Although inks may have similar chemical composition, they exhibit different scripting behavior due to their difference in fluency and their spread on the support. Our studies have shown that such variations can be captured in co-occurrence matrices that are used as the underlying structure for storing the spreading characteristics of the inks during the scripting process. The co-occurrence matrix describes the relationships of neighboring pixels in the inks, based on the distance d between two pixels in any direction. The direction between the two pixels is a concern, and this is implemented by multiple matrices, for four directions: horizontal, vertical and two diagonals. Employing the method of the grey level co-occurrence matrix, models are created for each layer of each ink. Each model incorporates all images that were produced by the three different pens (quill, calamus and penna) and the upper and lower case letters. For all model scripts co-occurrence matrices are formed in angles of 0° , 45° , 90° , 135° degrees and distance ($d=1$) of neighboring pixels, giving a total of 320 models.

Typically the co-occurrence matrices used record intensity values between 90-200, therefore the size of a co-occurrence matrix is approximately 110 x 110. The measurements in the co-occurrence matrices are normalized and therefore the results are independent of the size of the images. Computing the correlations of neighboring pixels with the co-occurrence matrix produces sets of characteristic variations for each model ink and the inks used in the test images. In order to verify the validity of the approach the co-occurrence matrices of the ink model are compared with:

- Each of the images that contribute to the creation of the models. The computation of the model of each ink

in a single co-occurrence direction includes 6 images (3 pens x 2 letter cases).

- The test scripting images that are created by different authors.
- Images of unknown ink composition taken from the manuscripts.

Co-occurrence matrices obtained from the model and test images are compared by calculating the Euclidean Distance between the co-occurrence matrices x and y given by:

$$h(x, y) = \frac{1}{\|x - y\|} = \frac{1}{\sqrt{((x_1 - y_1)^2 + \dots + (x_N - y_N)^2)}} \quad (1)$$

An example of the distribution of inks captured by the co-occurrence matrices in visible is shown in Figure 4. The eight figures show the co-occurrence matrices calculated in the vertical direction for each one of the eight inks used in our studies. Co-occurrence matrices in the four directions in visible and infrared are modelled as density functions which are then used for discriminating between the various types of inks.

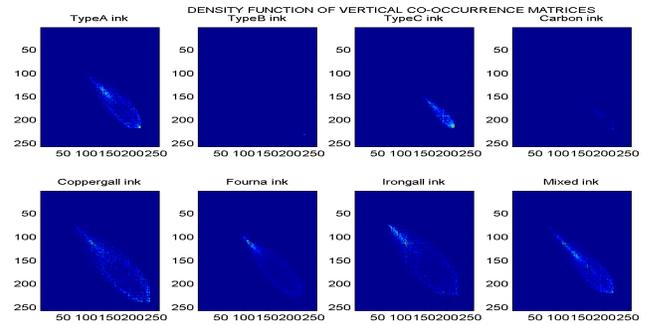


Figure 4. Vertical co-occurrence matrices for each of the eight model inks.

4 Results

The computation of the ink variations gave us important results in our attempts to characterize manuscript inks and as the results show in most cases, the identification of inks is feasible. All the results are presented in Figures 5 and 6.

Our results fall into three categories:

Successful A result is to be considered as successful when the correct model ink is identified.

Screening A result is to be considered as screening when the correct model ink is included among the first three results.

Unsuccessful A result is to be considered as unsuccessful when the correct model ink is not included among the first three results.

4.1 Model images

Figure 5 shows the percentage of the *successful* and *screening* results when the inks models are tested against images that were incorporated in the computation of the model inks. The computation of the model of each ink in a single co-occurrence direction included 6 images (3 pens x 2 letter cases).

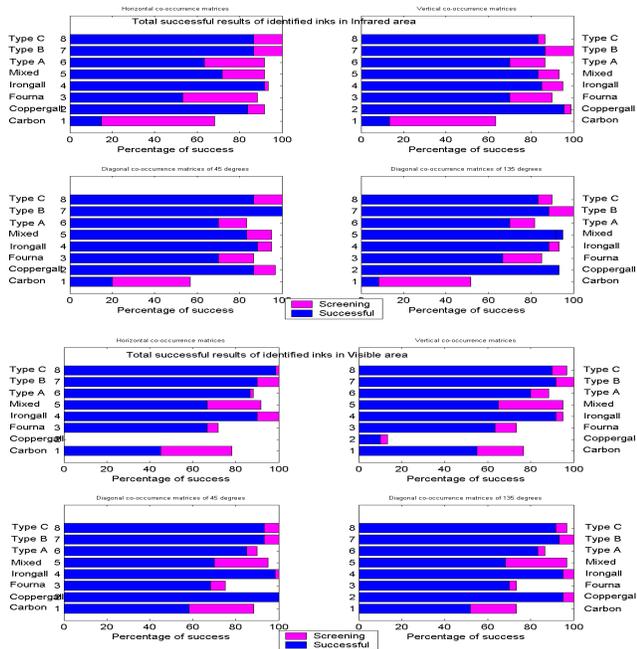


Figure 5. Top two row: Percentage of successful and screening results in the infrared area of the images used in the computation of the model co-occurrence matrices in four directions (horizontal, vertical, two diagonals). Bottom two row: Percentage of successful and screening results in the visible area of the images used in the computation of the model co-occurrence matrices in four directions.

The results are based on the computation of the four co-occurrence matrices under visible and infrared illumination. The following observations are made:

- Most of inks are capable of classification with over 80% success rate when the screening results are taken into account. The only exception is the carbon ink, which presents the lowest successful results. However,

carbon ink can be screened in the infrared area (over 60% success rate) and the visible area (over 80% success rate). Therefore it is possible that the classification of carbon ink can be achieved through the examination of the ink in the visible area of the spectrum in contrast to other types of ink.

- Among the four different measurement of co-occurrence matrices, the best results are given by the vertical and diagonal co-occurrence matrices in the infrared area and also in the diagonal co-occurrence matrices in the visible area.

4.2 Test images

Figure 6 shows the results of the scripting test images in the infrared (top two rows) and visible (bottom two rows) areas prepared by five different authors.

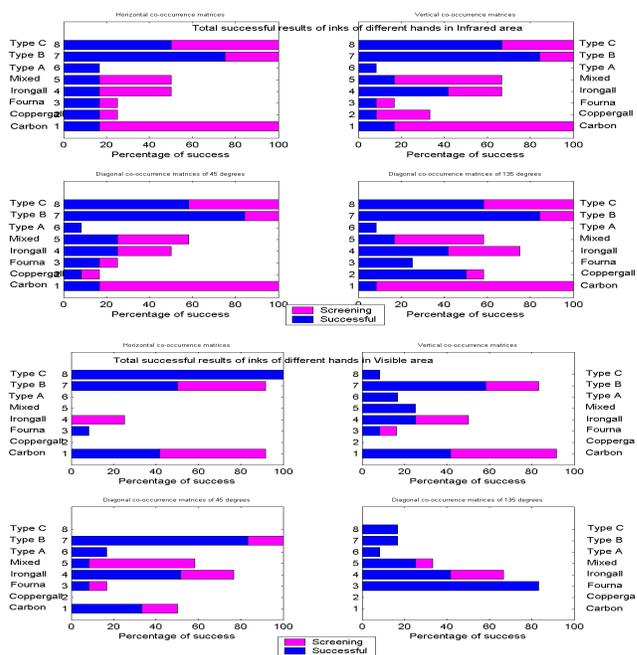


Figure 6. Top two rows: Percentage of successful and screening results of the test images in the infrared area in four directions. Bottom two rows: Percentage of successful and screening results of the test images in the visible area.

Examination of the results shows that it is possible to identify or screen seven out of eight inks with a success rate of 60% or more. In particular:

- typeC, typeB and carbon inks can be identified or screened in the infrared area with a certainty of 100% using any co-occurrence matrices.

Manuscripts	XRF	Infrared image-based
Memosa	Fe	TypeA and Irongall
Memosaa	Fe	TypeA and Irongall
Memosb	Fe and Cu	TypeC, Mixed, Coppergall
Memosc	Fe and Cu	TypeB

Table 1. Comparison between XRF and image-based results on the manuscripts

- Mixed, irongal and coppergal inks can be identified or screened in the infrared area with a certainty of 60% to 75% , using co-occurrence matrices of 90⁰ and 135⁰degrees.
- Fournal ink can be identified in the visible area.
- TypeA cannot be recognised in either the infrared or the visible area and is commonly mistaken for type B or type C inks and further investigation is needed in order to establish the causes.

4.3 Manuscript images

The ink models were also tested against Medieval manuscripts of unknown ink composition. The ink composition of the manuscript images were tested using XRF in order to compare it with the results given by the image-based analysis of the manuscripts.

A comparison of the results derived by the XRF method and infrared image-based is shown in Table 1. The results show that the constituency of the inks used in three out of the four manuscripts can be determined by the image-based results. In particular:

- The inks of TypeA and Irongall which have been identified as the correct models as the inks of the manuscripts *memosa* and *memosaa* include in their composition iron, as shown in the XRF measurements for these two manuscript inks.
- The inks of TypeC and Mixed which have been identified as the correct models as the ink of manuscript *memosb* include in their composition Copper, as shown in the XRF measurements for this manuscript ink. Irongall is indicated in the 45 degrees co-occurrence matrix.

We have not been able to verify the ink composition of the ink used in manuscript *memosc*, however further test are required using XRF to verify that iron and copper particles found by the XRF tests are indeed from the ink and not the support of the manuscripts.

5 Discussion and Future Work

In this paper we propose an image-based method for the discrimination of manuscript inks. Our method is based on the analysis of the image variations exhibited by the writing inks in the visible and infrared areas of the spectrum. Analysis in the visible areas mainly reflect the ink intensity whereas analysis in the infrared area reflects the ink composition. The analysis is based on the computation and comparison of co-occurrence matrices that capture the underlying structure of the ink fluency during the scripting process. Models of the inks are computed based on four co-occurrence matrices and we have taken into account scripting with different pens and authors and the thickness of the inks.

Based on the results presented we can conclude that co-occurrence matrices computed in the infrared area of spectrum can provide reliable information towards the discrimination of inks. This is supported by the fact that co-occurrence matrices could be used to identify or screen all of the inks used in the model images as well as five of the eight inks used in the scripting test. In addition the composition of three of the four unknown inks found in Byzantine manuscripts were determined. However, further work is currently undertaken to combine these results with other statistical measurements in order to increase their discriminatory ability as well as enhance the strength of the belief in the results.

6 Bibliography

1. Alexopoulou A., Kokla V., "Physicochemical study of inks of manuscripts using ultraviolet and near infrared radiation", 6th Int. Conf. on "Non-Destructive Testing and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage", 1999.
2. Barrow W.J., "Manuscripts and Documents", University Press of Virginia, 1972.
3. Cryssoulakis Y., Chassery J.-M., (1989),"The application of Physicochemical methods of analysis and image processing techniques to painted works of art", ERASMUS project
4. De Pas Monique," Etat des travaux effectues sur l' analyse des consituites des encres noires manuscripts par deux technique: chromatographie sur couche mince et electrophorese", Comite pour la conservation de l' ICOM, 1975.
5. Flieder F. *et al*, "Analysis des tannins hydrolysable susceptibles d' entrer dans la composition des encres ferro-geliques", 1975.
6. Gevers, T. and Smeulders, A., "A comparative study of several colour models for colour image invariant retrieval", Proc of First Int. Workshop on Image Databases and Multimedia Search, pp 17-27, 1996.
7. Janssens,K. *et al* "Use of Microscopic XRF for Non-destructive Analysis in Art and Archaeometry", X-Ray Spectrometry 29, 73-91, 2000.
8. Zerdoun Bat-Yeouda Monique, (1983),"Les encres noires au moyen ages(jusqu' a 1600)", edition du CNRS.