BERNARD J. AALDERINK, MARVIN E. KLEIN, ROBERTO PADOAN,
GERRIT DE BRUIN, AND TED A. G. STEEMERS

Clearing the Image: A Quantitative Analysis of Historical Documents Using Hyperspectral Measurements

ABSTRACT

Quantitative hyperspectral imaging is an optical measurement technique that has a great potential for historical document analysis. With the technique described here, optical images in seventy different wavelength bands are recorded. These images are subsequently spectrally calibrated by comparing them to recordings of a white calibration target with a known spectral reflectance curve. Post-processing algorithms for feature extraction and classification applied to the calibrated spectral image data can be used to investigate document aging phenomena and enhance the visibility of hidden features such as underdrawings.

In a first example a supervised feature extraction and classification algorithm is used to map the distribution of iron-gall ink corrosion within a nineteenth-century handwritten letter. The example shows that with the described procedure it is very possible to discern printing ink, uncorroded iron-gall ink, and corroded iron-gall ink.

The second example describes an unsupervised feature extraction and classification method, which was used to visualize underdrawings on a seventeenth-century historical map. The resulting false-color image shows a drastic increase in contrast between the underdrawings and the surrounding ink lines and pigments. Both examples demonstrate the great potential of the QHSI technique in combination with well-chosen, post-processing algorithms for investigation of historical documents.

INTRODUCTION

Major public archives and libraries worldwide have two competing tasks. In the first place, documents of cultural value should be made accessible to researchers and the general public of the present generation. Secondly, the institutes are responsible for keeping these documents in optimum condition for future generations. For the first task, new quantitative techniques are required that help to extract as much information as possible from already-degraded and fragile documents in a completely non-destructive way. For the second task, quantitative techniques are required to support an objective evaluation of the document condition and environmental influences therein. Hyperspectral Imaging (HSI) is a technique with which objects are imaged in tens or even hundreds of wavelength bands. Typically, HSI covers not only the visible, but also the near-ultraviolet and near-infrared, spectral range. This technique has already proven its worth in various fields such as agricultural research, space exploration, environmental investigations, and defense (Stein et al. 2001; Kruse et al. 2003; Lawrence et al. 2003). Often, these HSI systems are mounted on airborne or space-borne vehicles. On a microscopic scale, HSI technology is increasingly becoming a valued research tool especially in biomedical research (Schultz et al. 2001, Qin and Lu 2006).

However, it is also possible to use HSI for investigation of objects of cultural heritage. In the last few years this technique has been introduced for research in the area of document conservation and document analysis (Melessanaki et al. 2001; Mansfield et al. 2002; Casini et al. 2002; Attas et al. 2003; Klein et al. 2006; Kubik 2007; Padoan et al. 2008; Klein et al. 2008).

In this contribution we discuss the application of advanced numerical hyperspectral image processing techniques applied to the analysis of historical documents. Powerful computer algorithms for spectral feature extraction and classification allow one to detect even very subtle correlations in the hyperspectral data that cannot be detected with conventional visual comparison. Such algorithms can provide reproducible, quantitative results that enable a comparison of different objects or of the same object measured at different times, so that it becomes possible to establish databases and to measure the influences of treatments or aging. In the following, two particular applications will be discussed: 1) Evaluation of iron-gall ink corrosion and 2) detection of underdrawings.

Poster presented at the AIC 37th Annual Meeting, May 20–23, 2009, Los Angeles, California.

MEASUREMENT PRINCIPLE

The SEPIA quantitative hyperspectral imager used in these experiments is based on two wavelength TUnable LIght Projectors (TULIPs) that illuminate the document under an angle of 45º, and a monochrome digital camera that records the document from above (fig. 1). The light sources subsequently illuminate the document with a series of seventy well-defined optical wavelengths in the ultraviolet, visible, and near-infrared wavelength range (365–1100 nm). At each wavelength, a 4 megapixel grayscale image of a document area of 125 mm x 125 mm is recorded, corresponding to a resolution of 60 μm x 60 μm per pixel (ca. 400 dpi).

In order to translate the pixel values of each image into quantitative measurement values of the local spectral reflectance of the document, the recorded images have to be compared to recorded images of a reference target for each wavelength band. In this case recordings of a white reference target (Spectralon target, supplied by Labsphere, Inc.) with known reflectance are used for this calibration step. After this calibration, the value of each image pixel represents a precise measurement of the fraction of light reflected from the corresponding tiny document area at this particular wavelength and can be regarded as a local quantitative reflectance measurement. This hyperspectral imaging technique is therefore referred to as Quantitative Hyperspectral Imaging (QHSI) (Klein et al. 2008). The entire set of these (calibrated) spectral images is called the *hyperspectral data cube*. It contains for each pixel the entire spectral reflectance curve (fig. 2).

The spectral information in the hyperspectral data cube can then be used to distinguish different writing materials, such as inks and pigments, to measure the progress of aging processes, to enhance the legibility of degraded texts, and to determine deterioration effects on a document following an exposition or treatment.

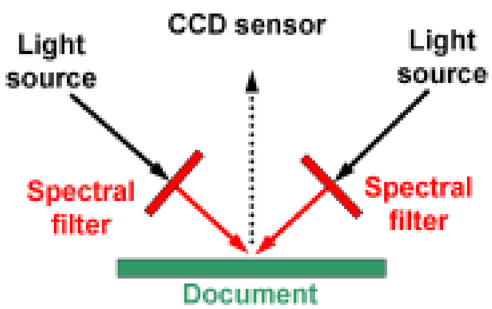


Fig. 1. Schematic overview of the QHSI instrument setup

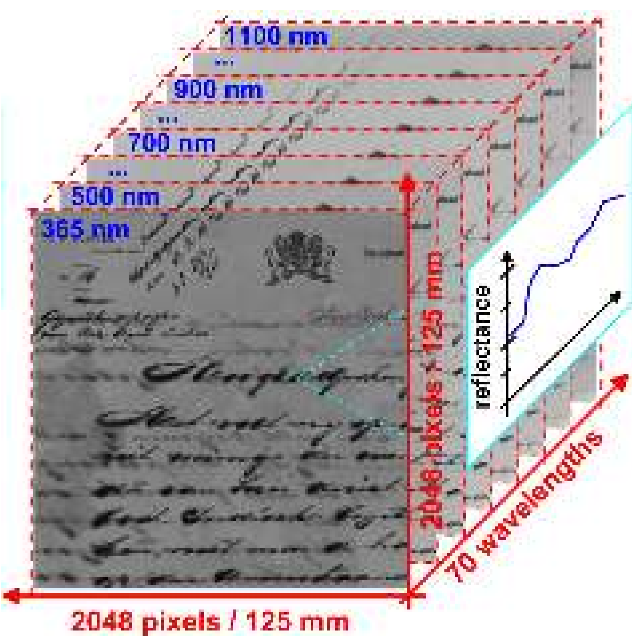


Fig. 2. Schematic representation of the hyperspectral data cube

MAPPING OF IRON-GALL INK CORROSION BY SUPERVISED CLASSIFICATION

In the field of conservation of documents, the behavior of ink over time is often studied (Public Record Office 1999; Havermans et al. 2003). Many of these studies focus on the deterioration of metal gall ink, due to its use in the production of handwriting inks for over one-hundred years and its well-known instability.

One important task for assessing the condition of a document is to distinguish ink areas that exhibit different degrees of corrosion. Figure 3 shows a section of an old letter (dated 1885) that contains both printed and handwritten text on a paper substrate. The handwritten text consists of iron-gall ink, which is in some places heavily corroded. A supervised classification procedure using QHSI measurements was used to assess the degree and distribution of the ink corrosion within this document.

The first step in the classification procedure is the definition of the different classes that are to be discerned (different spatial regions that can be identified on the document). In this case the classes *substrate* (paper substrate without ink), *corroded* (corroded iron-gall ink), *uncorroded* (uncorroded iron-gall ink), and *print* (print ink) have been defined. For each class a so-called Region-Of-Interest (ROI) is defined within the image, which represents a set of image pixels that can be clearly ascribed to one specific class. In our example, we defined the four ROIs as shown in figure 4. Using a so-called Spectral Distance Similarity (SDS) feature extraction algorithm (Homayouni and Roux 2004; Klein et al. 2008) all relevant spectral information for each pixel, which in this case is distributed over the seventy different wavelength bands, is compressed to four values per pixel (four, since four classes have been defined).

LEFT TO RIGHT

Fig. 3. A real-color image of a part of an old letter (dated 1885) containing printed and handwritten text on a paper substrate

Fig. 4. Overview of the defined regions-of-interest for the classes substrate, corroded, uncorroded, and print

Since for every pixel an absolute reflectance curve is available within the hyperspectral data cube, an average spectral curve can be calculated for each ROI (fig. 5). The SDS algorithm is subsequently used to measure for each pixel the similarity of its spectral content with each of the four ROI spectral curves. The result is a set of four so-called SDS *feature images* shown in figure 6. The pixel values in each image are a measure for the similarity of the pixel spectral curve with the average spectral curve of the corresponding ROI. A high pixel value (bright pixel) corresponds to a high similarity, a low pixel value (dark pixel) to a low similarity value.

Based on its four SDS values visualized in the feature images, each pixel has to be assigned to one of the four predefined classes. In order to do this, one has to define the so-called *decision boundaries*, which are mathematical conditions that are applied to the set of four SDS values of each pixel to determine to which class the pixel belongs. Instead of using the SDS values of the pixels in their spatial context, as described by the feature images, the mathematical conditions can be defined more easily if one regards each pixel as a certain point in four-dimensional (4D) mathematical space. The coordinates of the point on each of the four axes in this so-called *feature space* are given by the SDS values of each image pixel. While it is of course not possible to plot (and not even to truly imagine) a distribution of points in a 4D-space, one can select any two of the four coordinates to generate a so-called *scatter plot*. A scatter plot is a diagram, in which each pixel is plotted at a point whose x and y coordinate are given by the SDS values of the pixel in two selected feature images. In a mathematical sense, a 2D scatter plot is thus a particular projection of the pixel distribution in the 4D feature space.

Figure 7 shows the scatter plot of all pixels of the hyperspectral data cube for the two feature space axes that correspond to the corroded and uncorroded ink ROIs, respectively. Pixels that belong to one of the three ROIs that had been defined in the corroded, uncorroded and print ink area are

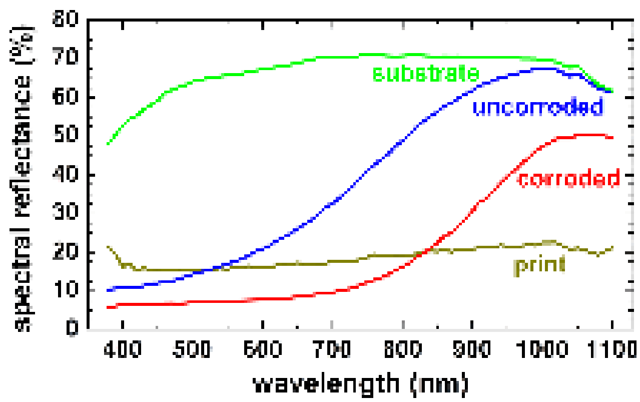


Fig. 5. Average spectral curves for each of the defined regions-of-interest

Fig. 6. Feature images calculated with the SDS algoritm: sub-strate (top left); print (top right); uncorroded ink (bottom left); and corroded ink (bottom right)

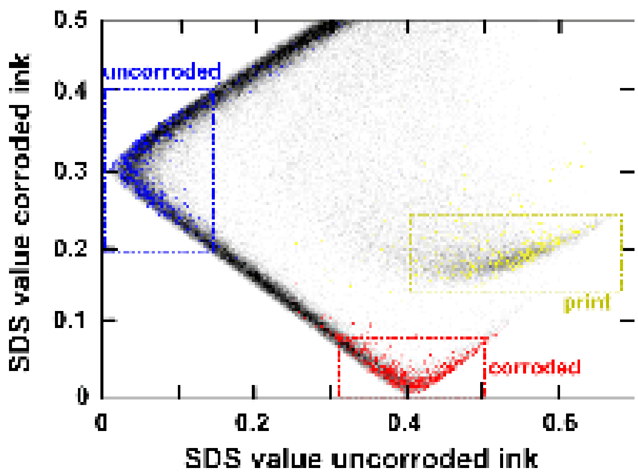


Fig. 7. A graph showing the distribution of two of the four SDS values for each pixel, together with the decision boundaries for the classification. Every pixel that falls outside the uncorroded, the corroded, or the print box is assigned to the class substrate.

Fig. 8. A color-coded graph showing the result of the classification process

Fig. 9. A real-color image of the map of New York (dated 1665), drawn by J. Vingboons. Part of the inventory of the Nationaal Archief

represented in the corresponding colors. Guided by the distributions of the colored ROI pixels in the scatter plot, three non-overlapping boxes are marked manually. Each box defines the decision boundaries for one of the ink classes. If due to its SDS values a pixel lies within such box, it is assigned to the corresponding ink class. If the pixel lies outside all three boxes, it is assigned to the substrate (background) class. The result of this classification process is depicted in figure 8. This image shows that ink corrosion is present in areas where the iron-gall ink is applied thickest.

DETECTION OF UNDERDRAWINGS BY UNSUPERVISED CLASSIFICATION

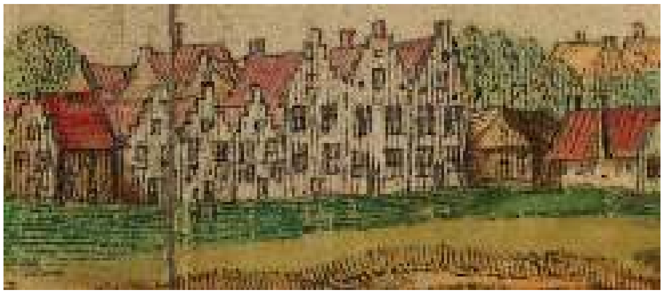
Besides the detection and mapping of deterioration effects, hyperspectral imaging can also be used for the detection of underdrawings (Kubik 2007). This example describes the results of using the described QHSI technique and post-processing algorithms to detect underdrawings in a historical map with a view of New York, drawn around 1665 by J. Vingboons in the Netherlands (fig. 9). This map is in possession of the Nationaal Archief (the National Archives of the Netherlands) in The Hague.

The historical map contains underdrawings that are visible with the naked eye in some places, but which cannot be discerned clearly in other areas. QHSI recordings of a small part of the map were made (fig. 10) in an effort to enhance the contrast between the underdrawings and the ink and pigments that are present. The resulting hyperspectral data cube was subjected to an unsupervised classification procedure.

In this classification procedure, the first step is to automatically extract the most important spectral information from the large number of images using a Principal Component Analysis (PCA) algorithm (Jolliffe 2002). This algorithm condenses the information from all spectral bands into only a few images. In this case the first five components were found to contain all relevant spectral data, thus resulting in a 5-dimensional feature space (the information is condensed to five values per pixel).

Subsequently, four classes were defined: substrate, ink, underdrawing, and road pigment (a brownish pigment used for coloring the road in front of the houses, see figure 10). The probabilities for each pixel to be a member of one of these classes were calculated using a Quadratic Maximum Likelihood (QML) classification algorithm (Hsieh and Landgrebe 1998). The result is visualized in a color-coded image that clearly shows the difference between underdrawings and ink drawings (fig. 11).

The false-color image shows that the contrast between the underdrawings (yellow) and the substrate with ink and pigments has increased drastically and that the underdrawings can now easily be discerned within the entire imaged part of the document.



LEFT TO RIGHT

Fig. 10. A real-color image of a recorded section of the map of New York where some of the underdrawings can be seen with the naked eye

Fig. 11. A false-color image of a recorded section of the map of New York, showing a strong contrast between the underdrawings (yellow), the substrate (blue), the ink (red), and the brownish pigment used in the road area (brown)

SUMMARY AND CONCLUSION

In summary, this article discusses the use of the quantitative hyperspectral imaging technique for the analysis of historical documents. Through advanced image processing algorithms a classification of the pixels in the recorded area is achieved. This can be used to map the distribution of different types of ink and of corroded and uncorroded areas for investigating degradation effects. As a second application, the use of the technique to enhance the visibility of hidden features such as underdrawings is demonstrated.

In conclusion, quantitative hyperspectral imaging is a very capable non-destructive analysis technique that can help to balance the needs of retrieving more contents information and preventing the decay of valuable historical documents. Current research at the Nationaal Archief and Art Innovation concentrates on further developing the hyperspectral imaging technique for investigation of aging and deterioration processes of historic documents.

REFERENCES

Attas, M., E. Cloutis, C. Collins, D. Goltz, C. Majzels, J.R. Mansfield, and H.H. Mantsch. 2003. Near-infrared spectroscopic imaging in art conservation: Investigation of drawing constituents. *Journal of Cultural Heritage* 4 (2): 127–136.

Casini, A., F. Lotti, M. Picollo, L. Stefani, and A. Aldrovandi. 2002. Fourier transform interferometric imaging spectroscopy: A new tool for the study of reflectance and fluorescence of polychrome surfaces. In *Conservation science 2002: Papers from the conference held in Edinburgh, Scotland 22–24 May 2002*, 249–253. London: Archetype Publications.

Havermans, J., H.A. Aziz, and H. Scholten. 2003. Non destructive detection of iron gall inks by means of multispectral imaging. Part 2: Application on original objects affected with iron gall ink corrosion. *Restaurator* 24 (2): 88–94.

Homayouni, S., and M. Roux. 2004. Hyperspectral image analysis for material mapping using spectral matching. In *Proceedings of the XXth ISPRS congress: Geo-imagery bridging continents, 12–23 July 2004, Istanbul, Turkey*, vol. 25, pt. B7, commission 7, 1682–1750. *International Society for Photogrammetry and Remote Sensing.*

Hsieh, P.F., and D. Landgrebe. 1998. *Classification of high dimensional data*, PhD diss., Purdue University.

Jolliffe, I.T. 2002. *Principal component analysis*, 2nd ed. New York: Springer.

Klein, M.E., J.H. Scholten, G. Sciutto, Th.A.G. Steemers, and G. de Bruin. 2006. The quantitative hyperspectral imager: A novel non-destructive optical instrument for monitoring historic documents. *International Preservation News* 40: 4–9.

Klein, M.E., B.J. Aalderink, R. Padoan, G. de Bruin, and Th.A.G. Steemers. 2008. Quantitative hyperspectral reflectance imaging. *Sensors* 8: 5576–5618.

Kruse, F.A., J.W. Boardman, and J.F. Huntington. 2003. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing* 41 (6): 1388–1400.

Kubik, M. 2007. In *Physical techniques in the study of art, archaeology and cultural heritage*, vol. 2, Dudley Creagh and David Bradley, eds., 199–259. Amsterdam: Elsevier.

Lawrence, K.C., B. Park, W. R. Windham, and C. Mao. 2003. Calibration of a pushbroom hyperspectral imaging system for agricultural inspection. *Transactions of the ASAE* 46 (2): 513–521.

Mansfield, J.R., M. Attas, C. Majzels, E. Cloutis, C. Collins, and H.H. Mantsch. 2002. Near infrared spectroscopic reflectance imaging: A new tool in art conservation. *Vibrational Spectroscopy* 28 (1): 59–66.

Melessanaki, K., V. Papadakis, C. Balas, and D. Anglos. 2001. Laser induced breakdown spectroscopy and hyper-spectral imaging analysis of pigments on an illuminated manuscript. *Spectrochimica Acta Part B: Atomic Spectroscopy* 56 (12): 2337–2346.

Padoan, R., Th.A.G. Steemers, M.E. Klein, B.J. Aalderink, and G. de Bruin. 2008. Quantitative hyperspectral imaging of historical documents: Technique and application. In *Proceedings of the Art 2008 Conference*. Jerusalem: International Seminars Ltd.

Public Record Office. 1999. *An introduction to printing inks*. Richmond, UK: Public Record Office.

Qin, J., and R. Lu. 2006. Hyperspectral diffuse reflectance imaging for rapid, noncontact measurement of the optical properties of turbid materials. *Applied Optics* 45 (32): 8366–8373.

Schultz, R.A., T. Nielsen, J.R. Zavaleta, R. Ruch, R. Wyatt, and H.R. Garner. 2001. Hyperspectral imaging: A novel approach for microscopic analysis. *Cytometry* 43 (4): 239–247.

Stein, D., J. Schoonmaker, and E. Coolbaugh. 2001. Hyperspectral imaging for intelligence, surveillance, and reconnaissance. In *Space and Naval Systems Warfare Center (SSC) San Diego Biennial Review 2001*, 108–116. San Diego: Space and Naval Warfare Systems Center.

BERNARD J. AALDERINK
Art Innovation, B.V.
Oldenzaal, the Netherlands
benno.aalderink@art-innovation.nl

MARVIN E. KLEIN
Art Innovation, B.V.
Oldenzaal, the Netherlands
marvin.klein@art-innovation.nl

ROBERTO PADOAN
Nationaal Archief
The Hague, the Netherlands
roberto.padoan@nationaalarchief.nl

GERRIT DE BRUIN
Nationaal Archief
The Hague, the Netherlands
gerrit.de.bruin@nationaalarchief.nl

TED A. G. STEEMERS
Nationaal Archief
The Hague, the Netherlands
ted.steemers@nationaalarchief.nl