QUANTITAVE HYPERSPECTRAL IMAGING FOR CLASSIFICATION AND MONITORING OF MATERIALS USED IN HISTORICAL DOCUMENTS

Roberto Padoan[1], Marvin E. Klein[2], Michael S. Brown[3], Gerrit de Bruin[1], Bernard J. Aalderink[2], Ted A.G. Steemers[1]

[1]Nationaal Archief - National Archives of The Netherlands, Postbus 90520, 2509 LM The Hague, The Netherlands; Phone: +31- 0615440041; e-mail: roberto.padoan@nationaalarchief.nl.

[2]Art Innovation BV, Zutphenstraat 25, 7575 EJ Oldenzaal, The Netherlands; Phone: +31 541 570720; e-mail: info@art-innovation.nl; web: www.art-innovation.nl.

[3] School of Computing, National University of Singapore, 13 Computing Link, Singapore, 117414.

Introduction

The creation of an objective documentation for the condition assessment of historical documents is the first step in a conservation process when both monitoring and/or restoration procedures are applied. In order to obtain a maximum amount of information about a single document it is crucial to take into consideration the typical heterogeneity of its composing materials. The numerous interactions of this mixture of substances with each other and with environmental components (oxygen, humidity, fungus etc.) in the course of its history make each document a unique object. The correct classification and condition monitoring of the composing materials of historical documents is therefore of primary importance when studying production techniques and degradation processes.

Spectrally resolved optical measurements on an archival document can provide valuable information for the identification of its component materials and for monitoring of their condition. Conventional single-beam instruments enable a detailed study of the spectral characteristics of one or more substances in a document, in fact their use in the field of book and paper conservation is constantly increasing. However, in order to exploit the spectral information for an objective characterization and monitoring of an inhomogeneous original document, a sufficiently high spatial resolution and location references on the object are required. It is for this very reason that quantitative hyperspectral imaging (QHSI) has been adopted in the *Nationaal Archief* (National Archives of The Netherlands) as an analytical and monitoring tool for the most valuable archival units preserved in this institution. The aim of the research project at the *Nationaal Archief* is to investigate the potential of this non-destructive analytical method by applying it to a selected number of original documents and artificially prepared samples. In addition, valuable data is collected and stored for the formation of a multi-instrumental database regarding materials composing the archival units and their degradation process. In this contribution, the QHSI instrumentation, experimental techniques and analytical methods will be described and a few representative results obtained from recent case studies will be presented.

Quantitative hyperspectral imaging

Quantitative hyperspectral imaging (QHSI) is a non-destructive remote sensing technique that provides calibrated spectral data with a high spatial resolution, which enables a reliable distinction and mapping of different areas on the measured document without the necessity to preselect a specific region of interest [Elachi 2006]. This characteristic is of primary importance when information for studying aging processes is collected, especially for long-term monitoring purposes. In fact, areas considered today as being representative for a certain type of degradation might in future studies be considered useless, whereas others, which had been excluded from earlier data analysis, might turn out to be the ideal regions of interest for the representation of a specific degradation process.

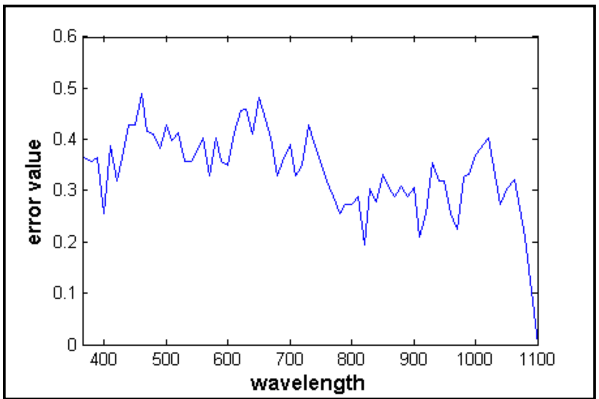
QHSI is a measurement of the optical spectral information with high spatial resolution on the object's surface. A series of greyscale digital images is recorded, where each image is taken at a specific narrow region of the light spectrum. After proper calibration and adjustments, the series of images are compiled in a three-dimensional data space called the hyperspectral data-cube. Basically, this data-cube contains for every individual QHSI image pixel the information of the entire spectral curve for the corresponding spot on the object. The availability of an entire spectral curve at each object point enables the application of complex analysis algorithms for an efficient discrimination of different materials on the recorded document surface and their compositions [Grahn 2007].

The experimental setting

The measurements at the *Nationaal Archief* are carried out with the SEPIA quantitative hyperspectral imager [Klein 2008], a prototype instrument specifically developed for this project by the company Art Innovation b.v. [AI 2011]. The SEPIA instrument is designed to simultaneously extract spectral reflectance values and spatial information from documents while the measurement itself has a minimum effect on their natural degradation processes. The SEPIA is based on two wavelength TUnable LIght Projectors (TULIPs) that illuminate the investigated document under an angle of 45º. The TULIPs are enclosed together with a monochrome digital camera mounted above the measurement area in a light-proof cabinet in which the document is placed. The projectors subsequently illuminate the document with a series of 70 well-defined optical wavelengths in the ultra-violet, visible and near-infrared wavelength range (365-1100 nm). At each wavelength a 4 megapixel grayscale image of a document area of 125 x 125 mm is recorded, corresponding to a resolution of 60 x 60 µm per pixel (ca. 400 dpi).

To translate the raw image pixel values into quantitative measurements of the local spectral reflectance of the document, the recorded image at each wavelength band has to be compared to recordings of a reference target. In this case a white reference target (Spectralon© target, supplied by Labsphere inc.) with known reflectance curve is used for this purpose. After this calibration processing, the value of each image pixel represents a precise measurement of the fraction of light reflected from the corresponding spot on the document at this particular wavelength and can be regarded as a local quantitative reflectance measurement [Klein 2008]. In order to achieve the maximum accuracy and reproducibility of the reflectance data, specific data recording and calibration procedures have been developed for the used SEPIA prototype instrument [Padoan 2009].

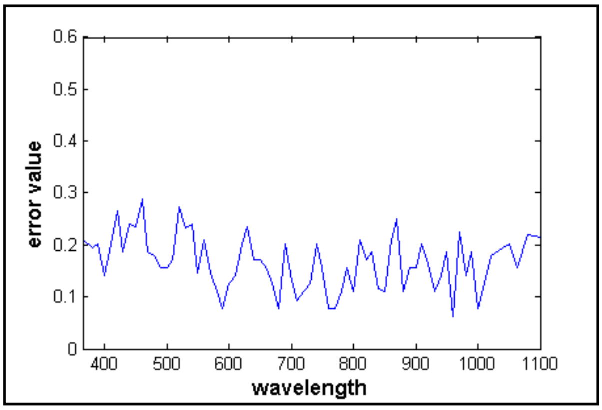
The alignment problem



Graph 1. Average pixel misalignment error per band within a QHSI data-cube.

When capturing the series of images that compose the QHSI data-cube, the instrument camera is refocused automatically at each wavelength to compensate for the inherent variation of the focal length of the objective lens. This refocusing results in slight misalignment of the individual images at different wavelengths. To counter this effect a calibration pattern with known markers is imaged at each wavelength. By observing how the known markers are displaced in each image it is possible to compute an image transformation algorithm that brings all images of the data-cube into an alignment. This type of internal misalignment of the data-cube images depends only on the "static" settings of the instrument and can be compensated by image transformation very accurately at less than 0.5 pixels average error (Graph 1).

Aligning multiple hyperspectral measurements is also necessary to compare multiple data-cubes of the same object imaged at different times. Since the document may be placed at slightly different orientations in the instrument at the time of recording, it is necessary to perform a software alignment between different QHSI data-cubes. To do this, the so-called SIFT (Scale Invariant Feature Transform) algorithm, which is well known in the field of image processing, has been implemented and used. The SIFT features between the two recordings are matched and a transform is computed to bring them into alignment. Based on experimental verification, it has been possible to perform this alignment between each band of two different QHSI data-cubes with an error below 0.5 pixels (Graph 2).



Graph 2. Average pixel alignment error between each band in two aligned QHSI data-cubes.

The quality of the alignment of single and multiple data-cubes is of utmost importance when multi-channel analyses such as principal component analysis (PCA) [Jocliffe 2002] are applied to hyperspectral data-cubes. This can be illustrated by comparing a PCA analysis carried out on a data-cube with and without prior image alignment. Figure 1A shows the real-color image of a measured section of a historic drawing [Vingboons 1665]. PCA was first applied by using the QHSI data-cube of this section without image alignment (fig. 1B) and then with a proper image alignment (fig. 1C), where only in the latter case all recorded wavelength images cover exactly the same area on the document. A comparison of the resulting PCA images (3rd component) shows that in figure 1C the fine underdrawings are distinguished very well from the ink drawings, whereas in figure 1B the use of unaligned images for the PCA processing makes the rendering of such fine features impossible.

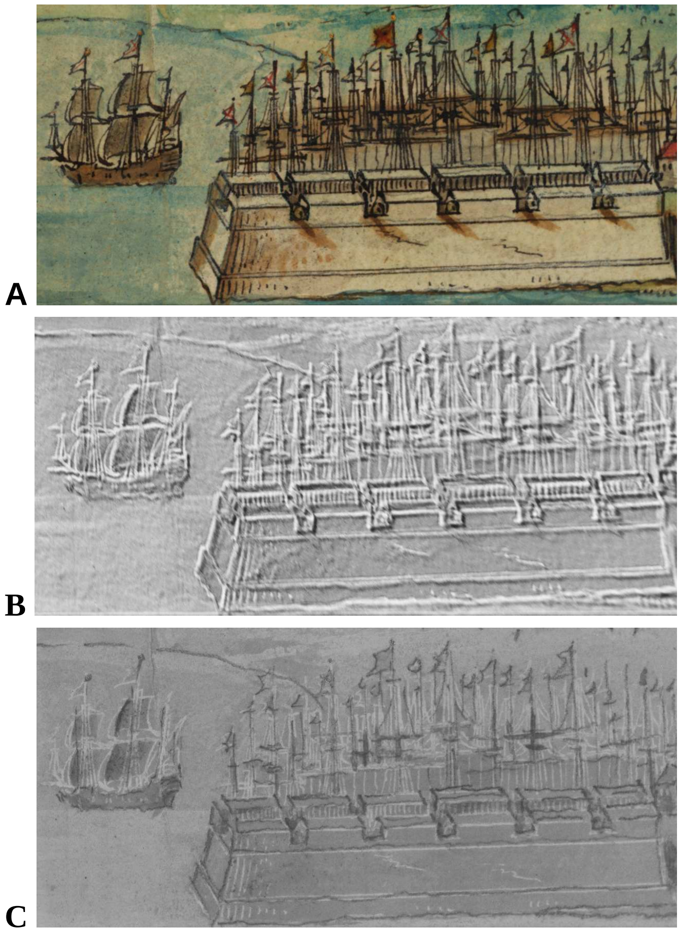


Figure 1A: Real-colour image of the panorama drawing of Havana [Vingboons 1665]. **B:** PCA performed on the hyperspectral data-cube without prior image alignment. **C:** PCA performed with prior image alignment.

Results obtained by single QHSI measurements

Once the digital information composing the data-cube is correctly calibrated and aligned, hyperspectral data from a single measurement can be analyzed with mathematical techniques to obtain a classification of the different surface areas of a historic document. By doing this it is in many cases possible to answer for example practical questions concerning manufacturing techniques, dating and history. Applications such as distinguishing inks and pigments and enhancing faint features such as hidden patterns, palimpsests, erased writings etc. are often performed in the *Nationaal Archief* to assist archivists and historians in their researches. A case study of major importance that can be considered representative for this type of application is the study of the Anjou Bible [Anjou], an illuminated Neapolitan codex of the 15th century that had been manufactured for the royal court of king Robert I of Anjou. The QHSI analysis of the Anjou Bible involved a number of advanced mathematical algorithms and various methods for visualizing the results. In particular, the best results have been obtained by using PCA, false-colour infrared rendering, comparison of spectral curves extracted from manually selected regions-of-interest, and the modified spectral-angle similarity (MSAS) algorithm [Homayouni 2004]. Using these analysis techniques, the classification of materials composing the illumination and the study of hidden patterns has revealed valuable information for its codicological study. Specifically, discrepancies have been found in the type of colours used to realise blue areas of an illuminated illustration that appeared similar under visible light. In addition, traces of a coat of arms belonging to another family previously owning the manuscript was discovered underneath the present coat of arms drawing. Furthermore, the legibility of an *ex libris* notation has been enhanced. These last two findings support existing theories on the history of the codex. For a more detailed description of the analysis procedures and the obtained results please refer to [Padoan 2010].

Results obtained by multiple measurements

A second type of applications of QHSI is possible when an entire series of measurements of the same object area is carried out repeatedly in the course of a certain period of time. After aligning the different data-cubes of the measurement series with each other, the spectral changes can be determined with high spatial resolution. This makes it possible, for example, to perform a quantitative analysis of material degradation processes, to verify the effects of conservation treatments, and to implement long-term condition monitoring of the document.

The goal of the Bihanne research project (named in honour of our beloved colleague Bihanne Wassink 1957-2008) is to explore the application of QHSI for monitoring historical documents, and to investigate the accuracy and reproducibility of the data obtained with the SEPIA instrument. Within this project a set of reference materials has been created, artificially aged in different steps under a variety of controlled climate conditions and repeatedly recorded in between the subsequent aging cycles. For the studies described in this article, discarded original document samples (single sheets from a discarded 19th century document with relief printing on machine made paper) were cut to squares of 120x120 mm in size, which were further separated into four strips. Each strip was aged under different conditions: 1) natural aging in the dark and under controlled conditions in the archives' repositories (no artificial aging applied); 2) light aging; 3) thermal aging with fluctuating relative humidity; 4) a combination of light aged and subsequent thermal aging with variable RH. For each aging cycle the following environmental conditions were applied: Light aging at 1200 lux for 5 hours with lamps type Whitestar UV-P (12v 50W 4200K); thermal aging for 15 hours in climate chamber type Vötsch VTRK 150 at a stable temperatures of 80° C and three cycles of fluctuating RH (35%-80%).

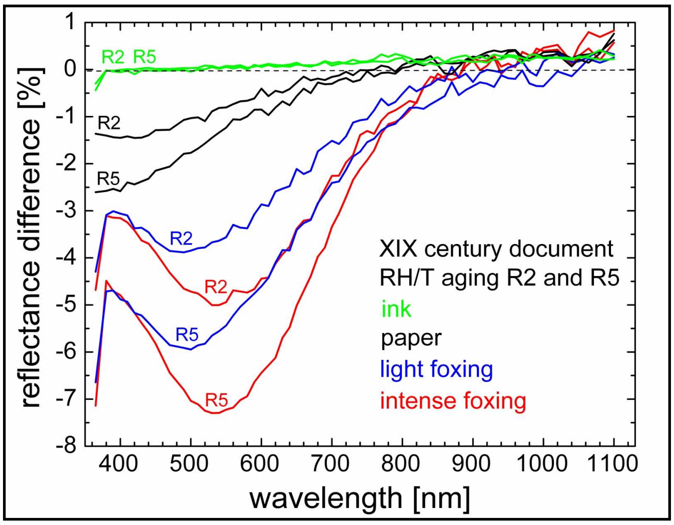
Two case studies regarding the developing of foxing and the degradation of printing ink have been chosen to show the potential uses of QHSI for monitoring applications. Both analyses have been executed on the same reference sample of the Bihanne project, which demonstrates the multi-task use of the QHSI data enabled by the fact that no pre-selection of a region of interest is required during the data collection.

Monitoring of foxing

Foxing is a localized paper degradation phenomenon that is mainly caused by oxidation processes and typically shows as an irregular distribution of brown stains on a sheet. This type of degradation can be caused by a free radical reaction of the paper (auto oxidation), the hydrolytic reaction caused by acids or other catalysis, or by biodegradation [Choi 2007]. In cases of very extensive foxing the physical and chemical stability of a paper sheet can be reduced. However, less extensive foxing is often considered to be a significant problem when it affects works of graphic arts with a relevant esthetic value. The progress of foxing typically results in an increasing number of affected spots, with an extension of the discolored areas and an increasing intensity of the discoloration of the individual spots. Due to the diversity in the causes of foxing formation in paper the need for an accurate documentation regarding its position, dimension and development is essential for both the condition assessment of a document and for a better understanding of this type of degradation. In fact up to now foxing is still at the center of an animated discussion of conservators and conservation scientists.

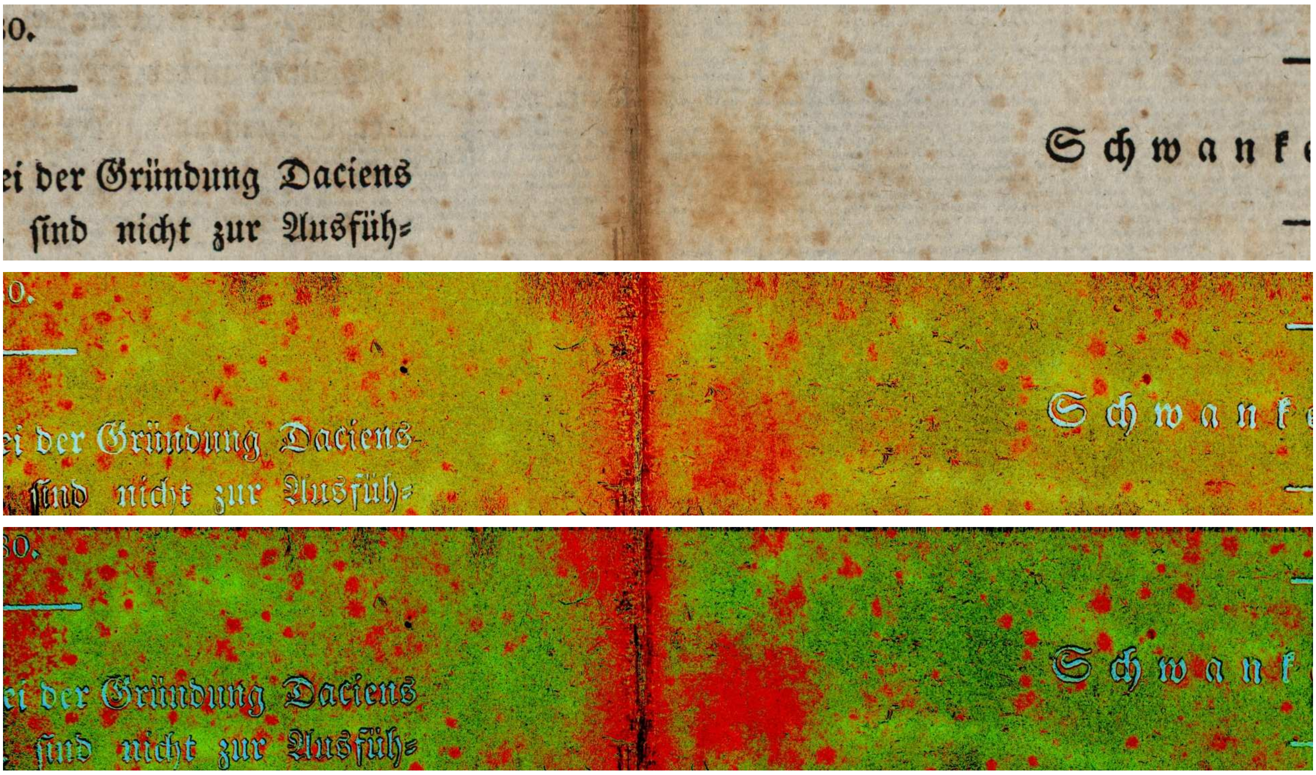
Figure 2A shows the real-colour image of a strip of the 19th century document, which already had foxing damage before the thermal aging was applied. This real-colour image was generated by effectively calculating a weighted sum of corresponding pixel values in the 41 spectral images contained in the QHSI data-cube for the wavelengths 380 nm to 780 nm (visible range). This is done according to the standardized calculation rules and coefficients given by the CIE, assuming the D65 standard illuminant as the light source and a 10º observer under photopic conditions [Hunt1998].

For demonstrating the use of the QHSI technique for degradation mapping and monitoring of foxing the spectral characteristics of this paper strip are compared before artificial aging (R0 measurement) and after 2 and 5 thermal aging cycles (R2 and R5 measurements, respectively). As a first step of the analysis regions of interest (ROI's) can be defined manually in areas of unaffected paper, ink, light foxing and intense foxing. From each of the three data-cubes (R0, R2 and R5), the mean spectral curves of these 4 ROI's are extracted. Graph 3 shows for each ROI the difference between the R2 and R0 curve and between the R5 and the R0 curve, corresponding to the spectral changes induced by 2 and by 5 aging steps, respectively. The reflectance curve of the ink area remains practically unchanged by the aging and the paper area unaffected by foxing shows smaller changes than the light and intense foxing areas. In addition, a clear distinction can also be found between the degradation of light and intense foxing areas with the spectral changes of the intense foxing area being faster than for the light foxing areas. A more detailed comparison of curves in Graph 3 shows that for both types of areas the spectral changes are practically the same in the UV-blue region of the spectrum (365-440 nm) and only small differences between light and intense foxing areas are found for wavelengths longer than 900 nm. In the spectral region from 440 to 900 nm the differences in the degradation process of light and intense foxing become clearly visible as the maximum spectral change is larger for intense foxing and it occurs at a different wavelength than that of the light foxing area (540 nm instead of 490 nm).



Graph 3. Differences of spectral curves extracted in different regions of interest (ROI) of the analyzed sample.

In a second analysis step, the differences observed in graph 3 of how the spectral curves of the different types of area typically change when subjected to thermal aging can be used to map the different types of areas on the document. In order to do so, for each pixel of the three aligned R0, R2 and R5 data-cubes the differences of the spectral curves of R2 and R0, and of R5 and R0 are calculated. Then each of these pixel difference spectral curves is compared with the difference curves of each ROI as shown in Graph 3. As a measure for the similarity of the pixel and the ROI curves the so-called spectral distance similarity (SDS) value [Homayouni 2004] is calculated over the spectral region 380-800 nm. For each ROI the SDS calculation results in a grayscale image, in which a high pixel value means that the spectral curve of this pixel has changed from R0 to R2 (or from R0 to R5) in a way similar to the change of the mean spectral curve of the ROI. Three of such grayscale images can be combined to a false-colour image (Figs. 2B-2C). In these images the colours indicate the same grade of similarity with the corresponding ROI difference spectral curves, where red pixels indicate spectral changes similar to the intense foxing ROI, green pixels changes similar to unaffected paper, and blue pixels for changes similar to the ink.



A

B

C

Figure 2. A: Real-colour image of the analyzed document (aged with RH/T) calculated from the 41 spectral images of the data-cube in the wavelength range 380 to 780 nm. **B:** False-colour image showing differences in the degradation process of paper after two aging cycles. Red pixels = intense foxing degradation; green = paper degradation; blue = ink degradation. **C:** False-colour image as in **2B** after five aging cycles.

The colour distribution in figures 2B and 2C visualizes the diversity of the types of degradation processes that, after respectively two and five aging cycles, have affected the spectral curves of the different object areas in different ways. In order to evaluate the effect of the different types of accelerated aging on the visual appearance of the paper the colour shift induced by the aging was calculated for all four sample strips of the XIX century document. In the CIELAB colour space for each spot on the sample the difference ΔE ($L^*a^*b^*$- CIE 1976) between its colour before accelerated aging was applied (R0-measurement) and its colour after five cycles of aging (R5-measurement) is calculated by

Eq. 1

$$\Delta E\left(L^*a^*b^*\right)=\sqrt{\left(L^*_{R5}-L^*_{R0}\right)^2+\left(a^*_{R5}-a^*_{R0}\right)^2+\left(b^*_{R5}-b^*_{R0}\right)^2}$$

where the $L^*a^*b^*$ values are calculated for each pixel from the 380 nm to 780 nm spectral images contained in the aligned R0 and R5 data-cubes according to the standardized CIE formula (illuminant D65, 2º-observer, photopic conditions) [Hunt1998]. Based on this calculation false-colour images (Fig. 3) have been generated, showing as dark-violet pixels those areas with ΔE = 0, as green pixels for ΔE = 5 and as red pixels ΔE = 10. It is generally said that a trained eye under optimal viewing conditions can distinguish colours in direct comparison if ΔE > 1.

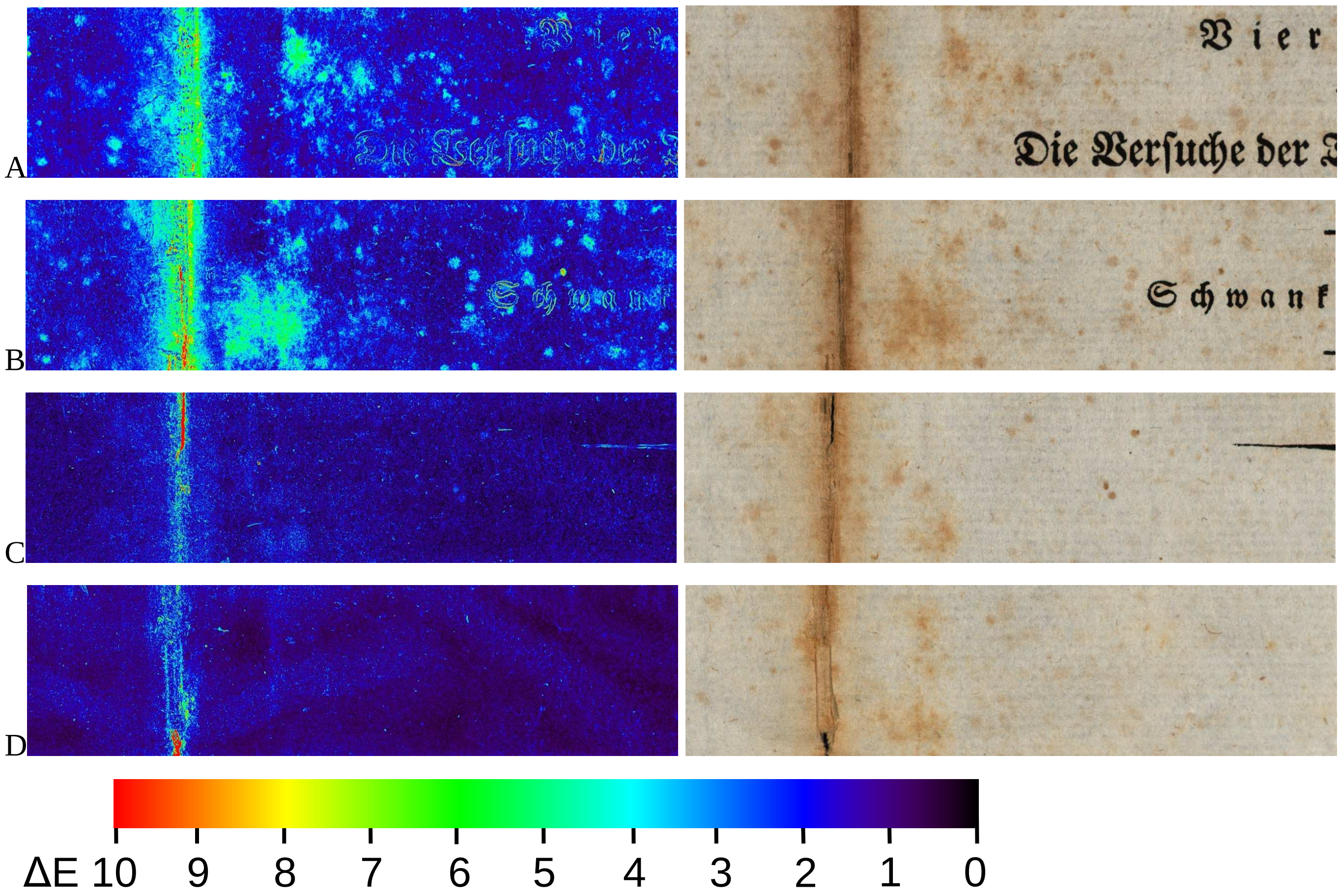
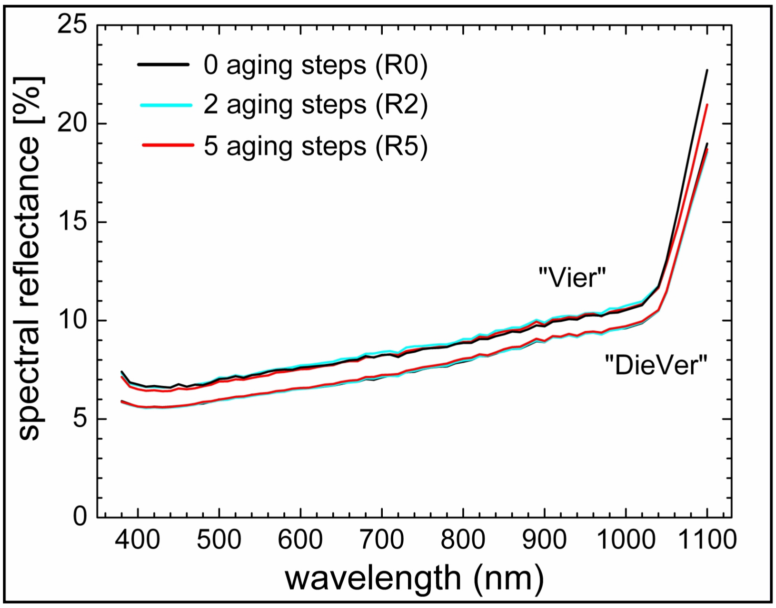


Figure 3. Mapping of the colour differences induced by five aging cycles. **A:** Accelerated combined thermal and light aging. **B:** Accelerated thermal aging. **C:** Accelerated light aging. **D:** Natural aging.

Figures 3A and 3B show that the largest colour change occurs in the intense foxing areas and along the fold line of the sheet for the combined and thermal-only accelerated aging. For the light-aging strip (fig. 3C) the colour change of the foxing areas is not larger than that of the paper areas, and it is in general less than for thermal aging. The colour change of the reference strip (Fig. 3D), which was only subjected to the natural aging, is very small as expected. By the study of these results it is possible to conclude that the specific type of foxing affecting this particular document is sensitive to changes in relative humidity and temperature but it is not especially sensitive to light aging.

Monitoring of printing ink

From the data-cubes of the XIX century document used for the foxing analysis, information can also be extracted for other composing materials such as the printing ink. Figure 4A shows the real-colour image of the sample strip that was subjected to the combination of light aging followed by thermal and RH accelerated aging.



Graph 4: Spectral curves extracted from different ROI's of the analyzed ink showing differences in the reflectance values even though only a single ink was used to print the analyzed page.

Graph 4 shows the spectral curves extracted for the three aging steps (R0, R2, R5) from two ROI's of the ink, namely the first line ("Vier") and the second line on the right side ("Die Ver"). For both ROIs changes caused by accelerated aging are minimal and much smaller than the initial differences of the curves for the two ink areas. However, it can be noticed that in the near-infrared range the curve for the ROI "Vier" changes due to the aging whereas the curve for the ROI "Die Ver" remains practically constant. In order to map the ink areas that show similar changes of the reflectance curves as either of the two ROI's, difference images of the aligned R5 and R0 measurements were calculated for all wavelengths from 380 to 1100 nm. For each pixel, the wavelength was determined for which the difference value was minimal, i.e. for which the largest reduction of the spectral reflectance was observed. The false-colour image Figure 4B shows blue pixels where this largest reduction occurs at short wavelengths (380 nm), green pixels for medium wavelengths (700 nm) and red pixels where it occurs at long wavelengths (1100 nm). Note that as opposed to conventional false-colour representations, the colour of a pixel in figure 4B does not depend on the *amount* by which the spectral reflectance decreases but only on the *wavelength* at which the greatest reduction occurs.

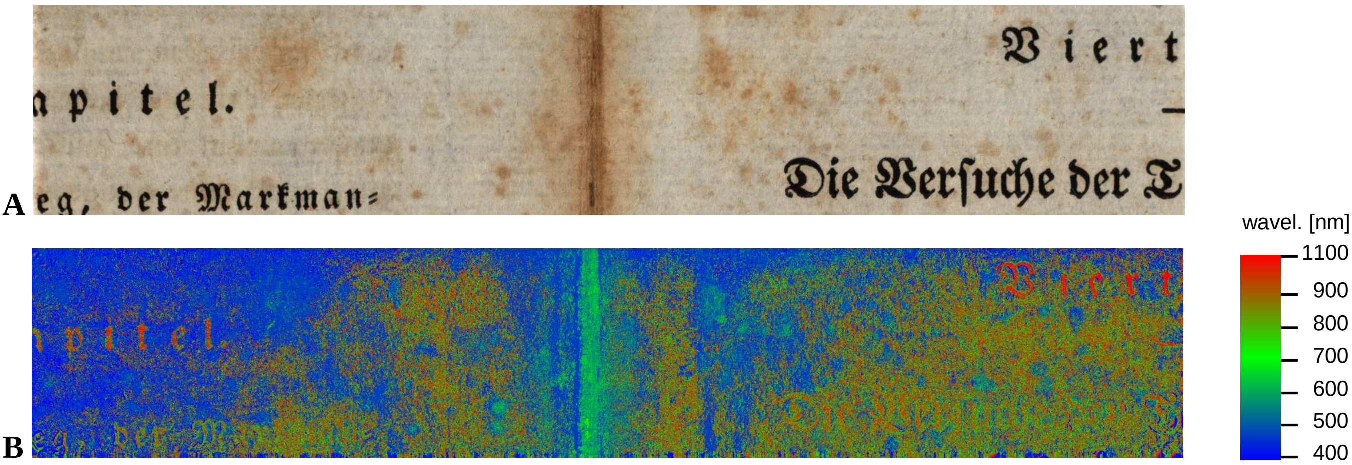


Figure 4. A: Real-colour image of the RH/T/Light aged document strip calculated from the 380 to 780 nm spectral images of the R0 data-cube. **B:** Colour-coded representation of the wavelengths at which the greatest reduction of the spectral values is induced by five cycles of combined thermal and light aging. Blue pixels for short, green for medium and red for long wavelengths.

In Fig. 4B, it can be seen that the first line on the left ("pitel") and the right one ("Viert") are fairly homogeneous and show the greatest reflectance reduction at long wavelengths (1100 nm). As opposed to this, the ink areas "der Markman" and "Die Versuche der" have a more inhomogeneous distribution of the wavelength with the greatest reflectance reductions, which occurs typically at shorter wavelengths than for the first ink areas. This difference cannot be explained by a variation in the composition of the ink as the text was certainly produced in a single impression by a uniformly inked plate. A difference in the concentration of the ink or in the local pressure during printing (different letter spacing) might be a more plausible reason, however, more research will be required to confirm this.

Discussion

The use of QHSI for the classification and monitoring of materials composing historical documents has been demonstrated using the example of an artificially aged XIX century document. Its high sensitivity for subtle and diverse changes in the spectral curves at any location on the measured object area make this technique a very important tool in standard conservation procedures for accurately recording the degradation processes of historical documents. QHSI provides the possibility to distinguish different types and speeds of degradation processes on the same analysed sample, allowing not only the classification of different materials in/on the same object but also to distinguish different aging processes of the same material. This type of classification is essential when micro-destructive or non-destructive spot tests need to be performed on documents for condition assessment and monitoring. The application of single-spot instruments can be easily guided by a colour-coded image of the analysed document showing the regions of interest that are most representative for a certain type of material or degradation process. This is why the *Nationaal Archief* is investing in building a multi-instrumental database of its most important documents using the QHSI mapping data as the reference for fixing information derived by different analytical tools. In parallel, research at the *Nationaal Archief* focuses on the further development and standardization of data acquisition and processing procedures to enable the implementation of the QHSI technique in a standard conservation working flow.

Bibliography

[AI 2011] Detailed information on a commercial version of the SEPIA quantitative hyperspectral imager can be found on the company's website www.art-innovation.nl.

[Anjou] Anjou Bible, Ms. 1, Maurits Sabbe Bibliotheek, K.U. Leuven (Belgium).

[Choi 2007] Soyeon Choi, Foxing on paper: a literature review Journal of the American Institute of Conservation, 46-2007, pp. 137-152.

[Elachi 2006] Charles Elachi - Jakob van Zyl, Introduction to the Physics and Techniques of Remote Sensing, Wiley-Interscience, New Jersey 2006.

[Grahn 2007] Hans F. Grahn – Paul Geladi, Techniques and applications of hyperspectral image analysis, Willey, Chichester 2007.

[Homayouni 2004] S. Homayouni – M. Roux, Hyperspectral image analysis for material mapping using spectral matching, ISPRS Congress Proceedings, 2004.

[Hunt 1998] R.W.G. Hunt, Measuring Colour, Fountain Press, Kingston-upon-Thames 1998.

[Jocliffe 2002] Jolliffe, I.T. 2002. Principal Component Analysis, 2nd ed., Springer, New York.

[Klein 2008] M.E. Klein et al., Quantitative Hyperspectral Reflectance Imaging, Sensors 2008, 8, 5576-5618 (PDF format, 3620 K) DOI: 10.3390/s8095576.

[Padoan 2009] R. Padoan, et al., Monitoring aging processes in archival documents by means of quantitative hyperspectral imaging: A part of the hyperspectral project at the Nationaal Archief (National Archives of the Netherlands), The Book and Paper Group Annual (AIC), Volume 28, 2009.

[Padoan 2010] R. Padoan et al., Quantitative Hyperspectral Study of the Anjou Bible, in The Anjou Bible: A Royal Manuscript Revealed, Peeters Publishers, Leuven 2010, pp. 171-185.

[Vingboons 1665] Johannes Vingboons, View of Havana, ca. 1665, Nationaal Archief inv. no. 4.VELH 619-57.