Multispectral Image Archiving of Watermarks in Historical Papers

Peter Meinlschmidt; Wilhelm-Klauditz-Institut, Fraunhofer-Institute for Wood Research; Braunschweig, Germany

Volker Märgner; Institute for Communications Technology, Technische Universitaet Braunschweig; Braunschweig, Germany

Abstract

Multispectral imaging is a technique used today in many quality control applications in industry or the cultural heritage field. Determining the origin of documents is an important task of historians. For this task not only is the visible text relevant but also further information about the physical paper document, such as watermarks. Today historic documents are mostly scanned using reflected visible light but such scans does not provide information about watermarks. For this, historians have to examine the original paper document. But the watermark may not be visible even with the original document due to heavy interference with overlaying ink. In this paper the digitization of historic papers using different spectral wavelength bands is discussed in general and the use of infrared bands for the detection of watermarks in particular. The details of a watermark detection system are presented and a concept for combining visual and infrared image digitization is shown.

Introduction

Most historic papers are so fragile that any additional handling has the potential to destroy valuable cultural heritage. All possible information should therefore be taken from any such document in just one step. Not only the text itself but also watermarks are important in determining the historical and geographical background of a document.

Visualizing watermarks in historic paper appears at first sight to be a very simple task. The paper only needs to be placed between a light source and the camera and the watermark can be seen instantly in this transmitted light. To obtain an undistorted image of the watermark the light source should be as homogeneous as possible for digital photography or a special scanner for transmitted light should be used.

This kind of recording technique works adequately enough as long as the paper is sufficiently transparent to allow recognition of the contours of the watermarks. The technique does however reveal its limitations once the surface of the paper is covered with different types of inks or drawings.

The development of digital NIR cameras and new image processing techniques has seen increasing interest in the digitization and analysis of historical papers. In these new techniques scientists have seen a real opportunity for separating the watermark from the overlaying writing. These cameras take high-resolution images which can be stored on the computer for later image enhancement and pattern recognition with the following aims:

- Reduction of the overlaying writing and drawing,
- Minimization of background noise,
- Extraction of the watermark emblem itself as well as chain and laid lines, and
- Archiving the image in the best possible quality.

In this paper we present a new infrared technique for visualizing watermarks, invented in close collaboration between the Fraunhofer Institute for Wood Research (WKI) and the Institute for Communications Technology (IfN) of the Technical University of Braunschweig.

The method is based on one hand on the fact that most of the different types of ink and colors are invisible and transparent in infrared wavelengths and on the other hand on the certainty that the transmitted infrared radiation is differently scattered or absorbed by the paper and the watermark.

The present paper is organized as follows. We start with a short presentation of the digitization process in the visual wavelength band with an example of a historic paper. The next section presents the information which becomes available with selection of bands from the visual light range through the near-infrared and up to mid-infrared wavelengths. After this we describe the concept behind the proposed infrared technique for visualizing watermarks in paper. Finally we present the concept of a combined visual and infrared image digitization and close with some concluding remarks.

Visual Image Acquisition for Text and Surface Structure Archiving

For many years extended scanning has been a task performed by almost every library all over the world. Virtual libraries are an important part of the internet and more and more images of historic documents can be viewed in an excellent manner [6, 7]. Viewing the documents is of course important but today an ability to search the documents is even more important. It follows that metadata need to be added to each document. The manual generation of metadata is very time-consuming and error-prone for many documents. For this reason current research is looking at ways to do this automatically. For historic printed fonts or handwritten text in particular, keyword searching, word spotting, or even text recognition are still a challenge and a matter for further research [1, 2].

Figure 1 shows part of a historic page containing handwritten text as a color image and after preprocessing and binarization. The next operation is line segmentation and the final step is word recognition. Figure 2 shows one approach to the segmentation of touching lines. All of the words in a line are separated by line segments of the same color both above and below the text. Often characters in adjacent lines are touching each other and a cut is needed to separate them. An example of this is shown in Figure 2.

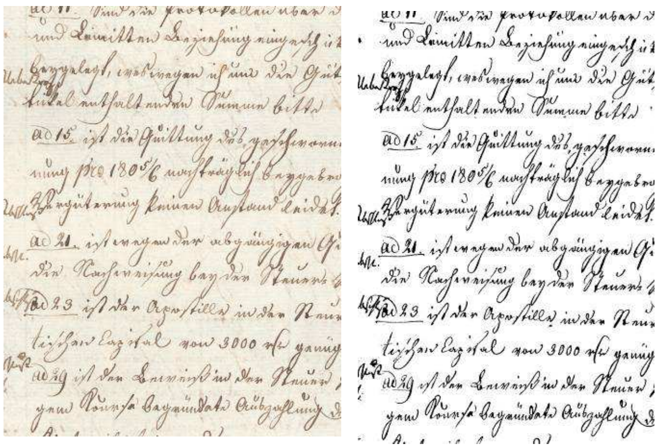

***Figure 1**. Image of a scanned handwritten document (left) and the preprocessed and binarized image of that document (right)*

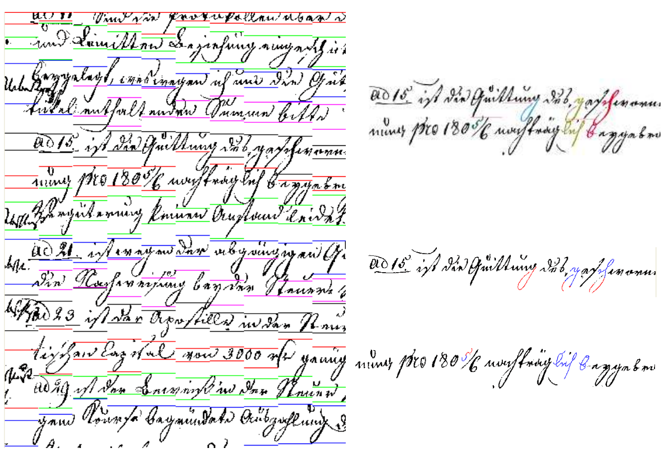

***Figure 2**. Image of the segmented lines (left), an example of two touching lines and segmentation into two single lines (right, from top to bottom)*

Based on preprocessed page images and the segmentation of a page into lines or even words (for many historic text pages a segmentation of lines into words is impossible due to the handwriting style), pattern recognition techniques are under development for the recognition of text or words to help researchers to work on text information without needing to type the text manually. Even though automatic recognition of handwritten historic text is far from being solved, promising research results have been presented recently for word retrieval or word spotting within historic document images [3].

In addition to the information in the text itself historians often need the watermark as further information helping to provide an indication about when the document was produced or about the atelier where a drawing was made. In many cases simple backlight is not sufficient to make a watermark in a sheet of paper visible, especially in cases where either text or graphics interfere with the watermark or where the paper thickness or an underlying ground coating of crayon prevents light transmission. Spectral imaging can help to solve this problem.

Near- and Mid-Infrared Image Acquisition for Watermark Archiving

In a first experiment, the historic paper was placed between a CCD camera sensitive to light between 400 and 900 nm and a source of homogenous white light. A narrow bandpass filter with a center wavelength at 780 or 830 nm was placed in front of the camera to reduce the interference of the handwriting (in black iron-gall ink) in watermark detection.

Figure 3 shows the corresponding photographs taken at 780 nm (left) and 830 nm (right).

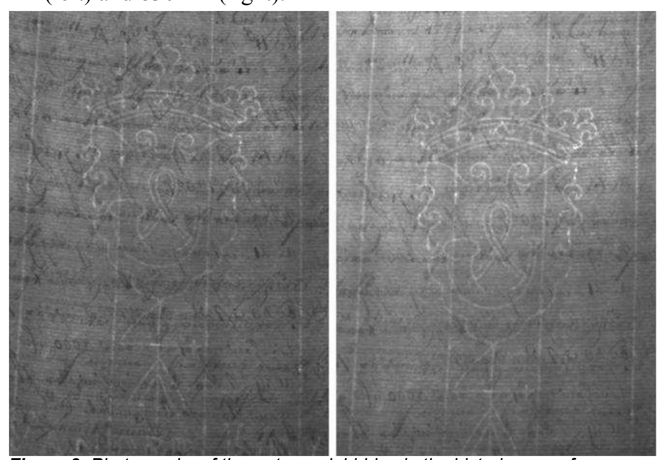

***Figure 3.** Photographs of the watermark hidden in the historic paper from Figure 1 taken with a bandpass filter at 780 nm (left) and at 830 nm (right) in front of the camera.*

From Figure 3 it follows that near-infrared light at 830 nm is more suitable for suppressing the handwriting written with black iron-gall ink than is 780 nm or the visible light. Consequently quite some time ago some scientists were already working with near-infrared cameras or even before this with near-infrared film in order to suppress the image of the iron-gall ink [4]. Unfortunately a kind of ghost image which could not be suppressed in the near infrared always remained.

In contrast to the near-infrared light which ranges from 700 to 3000 nm and which is simply an extension of visible light, the thermal mid-infrared light extends from 3000 nm to 7000 nm (MWIR).

The MWIR cameras are highly sensitive to temperature differences of as little as two hundredths of a degree centigrade (< 0.02 °C). The temperature information is usually supplied as a black and white image with a dynamic range of 12 to 16 bits (the number of gray values varies from 4096 to 65,536). Often, these gray values are converted into false colors, with blue and black representing cold and reds representing hot temperatures in accordance with natural human perception.

The newly developed technique for visualizing watermarks using the MWIR technique was invented in close collaboration between the Fraunhofer Institute for Wood Research (WKI) and the Institute for Communications Technology (IfN) of the Technical University of Braunschweig.

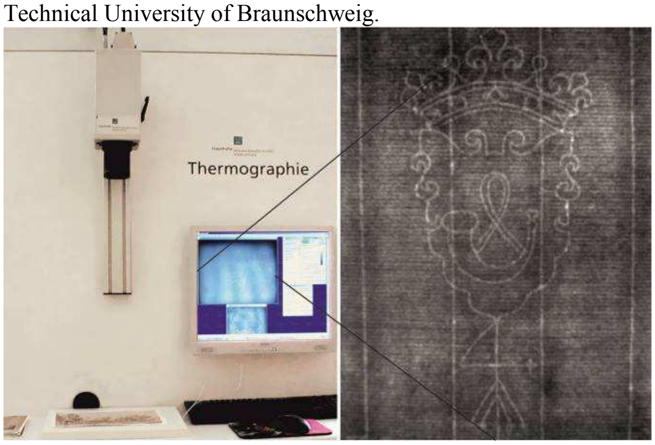

Figure 4. *Photograph (left) of a set-up with an infrared camera and a flat heating plate. The paper watermark shown in Figure 3 can be seen clearly as bright lines on the screen with virtually no interference from the drawing (right).*

This technique is based on the fact that most of the different types of iron-gall ink are invisible or transparent in the infrared and the fact that transmitted infrared radiation is scattered or absorbed differently by the paper and the watermark.

For initial experiments an infrared camera with a sensitivity to infrared reaching from 1500 nm to 5000 nm is positioned above a flat heating plate covered with a thick frame of passepartout on which the object under investigation is placed (Figure 4). When the plate is heated to about 40 °C the infrared radiation emitted from the 'black body' is transmitted through the paper. The radiation will be absorbed differently depending on the thickness and the density of the paper and watermark. This variation in radiation can be captured by the infrared camera and archived on a computer [5].

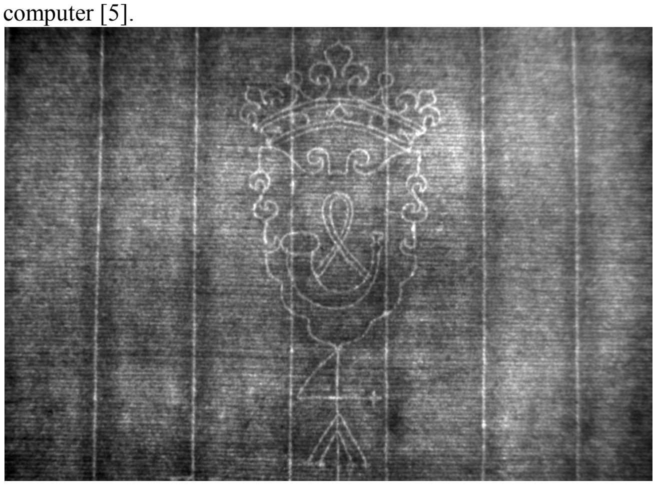

Figure 5. *Thermogram of the watermark hidden in the historic paper from Figure 1 taken with an MWIR-camera and an additional longpass filter with a passband from 3000 to 5000 nm.*

Tests with an MWIR infrared camera show that the ink on the paper is invisible or transparent and only the watermark structure of the historic paper can be seen (Figure 4, right). For additional experiments a longpass filter and four narrowband pass filters with center wavelengths at 3360 nm, 3577 nm, 4240 nm and 4470 nm were used.

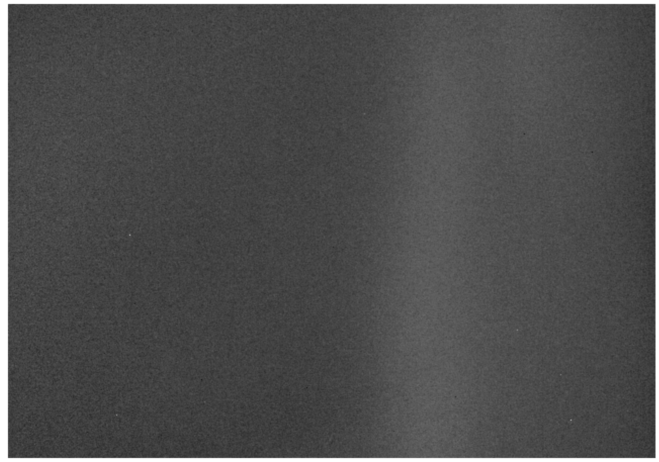

Figure. 6. *Thermogram of the historic paper from Figure 1 taken with an MWIR camera and an additional bandpass filter with the center wavelength at 3360 nm.*

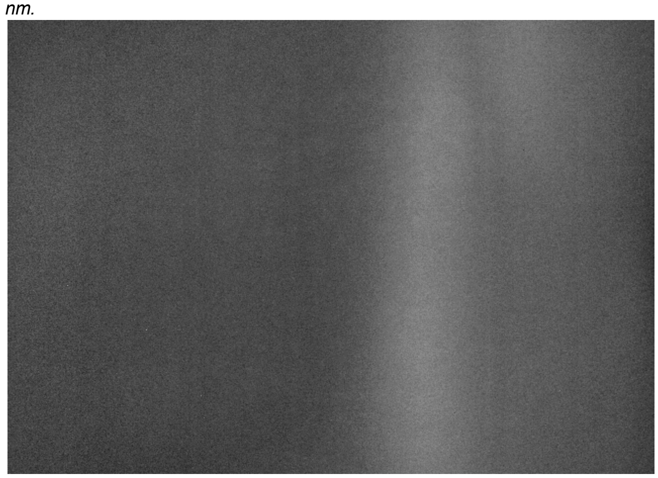

Figure 7. *Thermogram of the historic paper from Figure 1 taken with an MWIR camera and an additional bandpass filter with the center wavelength at 3577 nm.*

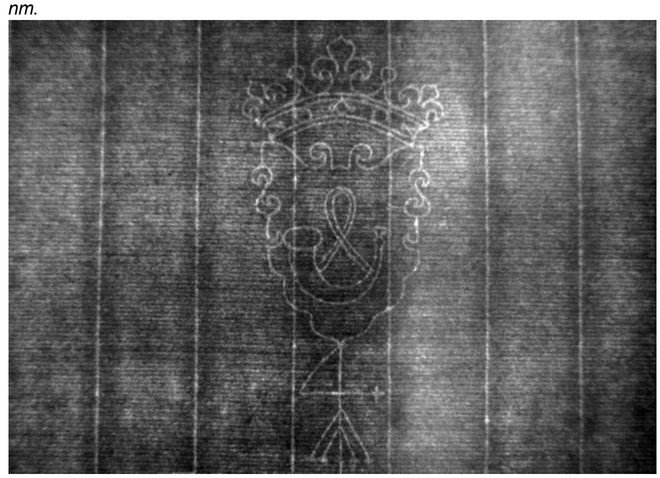

Figure 8. *Thermogram of the historic paper from Figure 1 taken with an MWIR camera and an additional bandpass filter with the center wavelength at 4240 nm.*

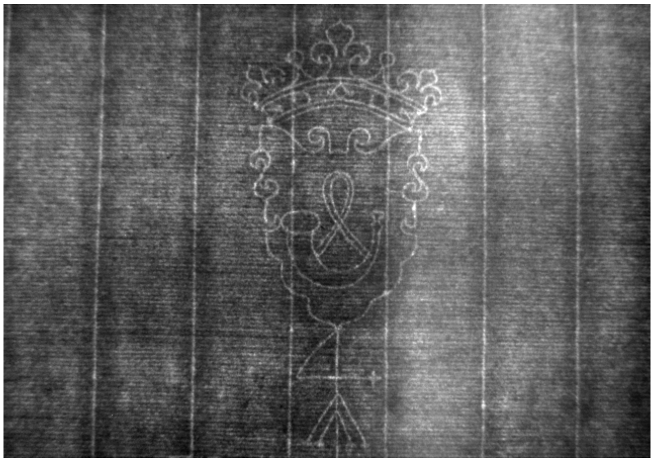

Figure 9. *Thermogram of the historic paper from Figure 1 taken with an MWIR camera and an additional bandpass filter with the center wavelength at 4470 nm.*

The longpass filter has a lower wavelengths limit at 3000 nm and an upper cut-off wavelength at 5000 nm. The thermogram from this test can be seen in Figure 5 and shows the watermark without any interference from the iron-gall ink in its best contrast.

Further tests using narrowband pass filters with center frequencies at 3360 nm and 3577 nm show no information about the watermark, as can be seen from Figures 6 and 7. The images in Figures 8 and 9 obtained by using the narrowband pass filters with center wavelengths at 4240 nm and 4470 nm once again show detailed information for the paper structure and the watermark. From these results we can conclude that there is low transmission or scattering of infrared light between 3360 nm and 3577 nm.

The image with the most detail and best contrast was the one using the longpass filter which is shown in Figure 5. Due to the cut-off of the near-infrared wavelengths between 3000 nm and 5000 nm this image (Figure 5) is even better than the one taken without any filter (Figure 4, right). These experiments show that the investigation of historical papers with infrared cameras using selected wavelength bands results in an excellent image without almost any interfering ink.

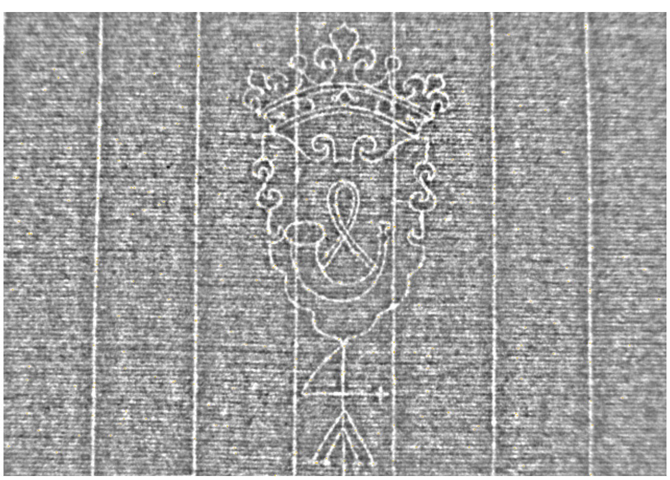

Figure 10. *Thermogram of Figure 5 processed to flatten the shading of the background.*

Figures 10 to 12 show the results from applying certain image-processing methods to the thermogram in Figure 5 in order to make the paper structure and the watermark more clearly visible. Figure 10 shows the thermogram without background shading and with contrast enhancement. The details of the wire screen and the watermark are clearly visible. Figure 11 shows another presentation of this image by relief representation

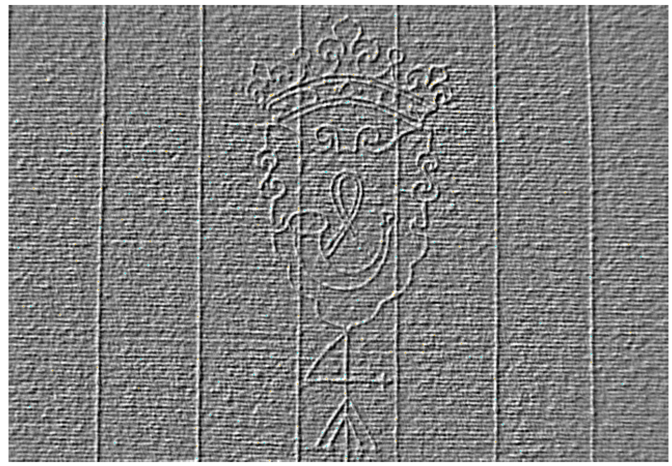

Figure 11. *Image from Figure 10 represented as a relief to make details more visible.*

A binarization as shown in Figure 12 is a possible further image-processing step and may be the starting point of a subsequent automatic recognition and measurement process.

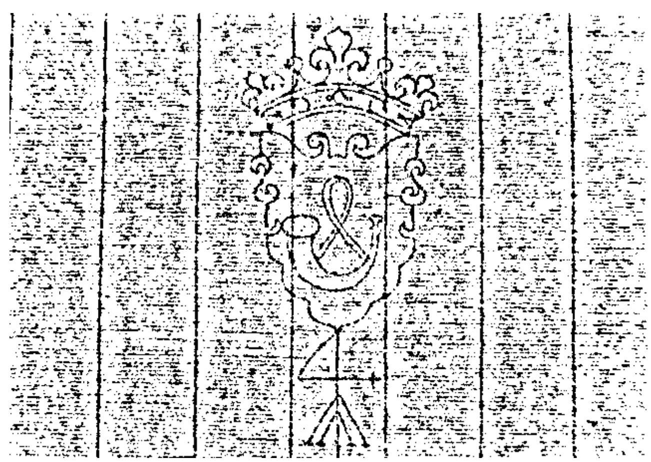

Figure 12. *Binary image of Figure 10 shows the small white lines of Figure 10 as black lines.*

The interesting part of this approach is that the resulting image of the watermark as can be seen in Figure 10 is of such a good quality that image processing and even automatic pattern recognition algorithms can be applied. These techniques can be used to help a researcher to compare features of the paper and the watermark automatically. Measuring the distances of the chain- and laid-lines and the relative position of the watermark within these lines, and the recognition of the watermark itself using pattern recognition methods are now possible. As a simple example for the measurement of the line distances the Fourier transform of the image in Figure 10 shows the periodic structure of the chain and laid lines (Figure 13) and a direct measurement of the distances of these lines is possible.

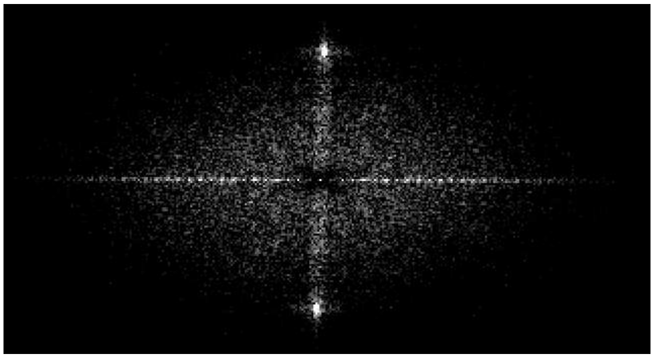

Figure 13. *Detail of the Fourier transform of the thermogram from Figure 10. The white points clearly show the periodic structure of the chain and laid lines.*

Combination of Multipectral Images

While the set-up in Figure 4 (left) is suitable only for single sheets of paper, another configuration is necessary for examining pages attached together in a book (Figure 14). It can be easily extended with an additional CCD camera and light sources so as to enable archiving of the visual image and the paper watermark together in one step.

Figure 14. *Set-up for detecting watermarks in books. The photograph shows the infrared set-up used in the examination of a historic book.*

In Figure 14 the heating plate can be seen as a black sheet behind the page (in the center) in front of the infrared camera (left side). This technique has been used with a first prototype set-up to archive the watermarks in historic sheets of paper at the Bayerische Staatsbibliothek München (Bavarian State Library) in Germany [7, 8]. With this prototype, only the watermarks are digitized using the infrared camera. This means that a second scan for the visible page image has to be done with a scanner and these two data sets have to be merged in a further processing step. It is planned to integrate into the set-up shown in Figure 14 a camera for visible light together with a light source which will enable the operator to create simultaneously an image of the visible page with the writing or drawing together with an image of the paper structure and the watermark. This makes the archiving process simpler and also permits a more detailed analysis of the paper characteristics of the page without the need to examine the paper original.

Conclusion

Images of sheets of paper taken by an infrared camera sensitive to light between 3000 nm and 5000 nm show details of the paper structure without any interference from ink when a low heat-emitting flat source is put behind the paper sheet. The level of interference from the ink depends on the quality and thickness of the paper and ink used. First tests have yielded excellent results from different types of paper and inks. Further research has to be done to learn more about the spectral dependency of ink and paper. Furthermore, the watermark scanner needs to be integrated into a book scanner in order to provide a fast scanning process.

References

[1] H. Balk, Improving Access to Text IMPACT, Periodic Report 2009, FP7 project, http://www.impact-project.eu

[2] A. Juan, V. Romero, J.A. Sánchez, N. Serrano, A.H. Toselli, E. Vidal, Handwritten Text Recognition for Ancient Documents, JMLR: Workshop and Conference Proceedings 11, pp. 58 – 65 (2010)

[3] Y. Leydier, A. Ouji, F. LeBourgeois, H. Emptoz, Towards an omnilingual word retrieval system for ancient manuscripts, Pattern Recognition 42, pp. 2089 - 2105 (2009)

[4] D. A. Kushel, Application of Transmitted Infrared Radiation to the Examination of Artifacts, Studies in Conservations 30, pp. 1 – 10 (1985)

[5] Peter Meinlschmidt, Volker Märgner, Advantages and Disadvantages of Various Techniques for the Visualization of Watermarks, Restaurator 30, 2009, pp. 222 – 243

[6] http://www.bernstein.oeaw.ac.at/twiki/bin/view/Main/ProjectExhibitions#Catalogues (accessed 20. 07. 2009)

[7] http://www.digitale-sammlungen.de/index.html?c=digitalisierung-scanner&l=de

[8] B. Wagner, R. Bacher, Blockbooks in Bavarian Collections: A New Project for the Digitization and Analysis of Early Printed Books, Microform & Digitization Review, MDR, Vol. 40, pp. 18–24, March 2011

Author Biographies

Peter Meinlschmidt studied physics at the universities of Hamburg and Oldenburg (Germany) and at Towson State University (USA). He received his M.Sc. degree in 1991. Since 1996 he has been with the Fraunhofer Institute of Wood Research, Braunschweig, working on NDE applications of thermography to wood and wood-based materials. His current research interests include optical metrology, thermographic techniques and spectral imaging.

Volker Märgner received his diploma and doctorate degrees in electrical engineering from the Technische Universitaet Braunschweig (Germany) in 1974 and 1983 respectively. His main areas of research are image processing and pattern recognition methods using visible light and infrared.