A Digital Image Processing and Database System for Watermarks in Medieval Manuscripts

Emanuel Wenger (*), *Victor N. Karnaukhov* (#), *Alois Haidinger* (§), *and Maria Stieglecker* (§)

(*) Austrian Academy of Sciences, Commission of Scientific Visualization, Vienna, Austria
E-mail: emanuel.wenger@oeaw.ac.at

(#) Russian Academy of Sciences, Institute for Information Transmission Problems, Moscow, Russia
E-mail: victor.karnaukhov@iitp.ru

(§) Austrian Academy of Sciences, Commission of Paleography and Codicology of Medieval Manuscripts, Vienna, Austria
E-mail: alois.haidinger@oeaw.ac.at or maria.stieglecker@oeaw.ac.at

ABSTRACT
Watermarks are the most important tool for dating old, not explicitly dated paper documents. Hence, catalogues and databases of watermarks play an eminent role for the work of medievalist and paper historians. This article presents an integrated software system developed for storage, retrieval, manipulation, digital processing, and identification of watermarks in old manuscripts. The whole processing pipeline beginning with the scanning of watermarks up to their identification is described. Most of these watermarks were captured from manuscripts kept in the Klosterneuburg monastery (Austria).

KEYWORDS: watermark, database, dating, medieval manuscripts, paper, digital image processing

INTRODUCTION

Medieval manuscripts, and incunabula represent an important part of our cultural heritage. Investigation, cataloging, and restoration of these books are necessary in order to preserve our heritage for the future. Many of the old books (documents) are not dated explicitly, although the knowledge of the date of their production would be essential for historical research. The comparison of dated watermarks with undated such is the major method for dating undated medieval handwritten paper documents. Several standard catalogues exist containing thousands of hand-drawn sketches of watermarks [2,6]. The identity of a watermark with one in the standard catalogues is a good indicator for the age of the watermark and document in question.

There are some essential hindrances to the precise dating of the watermarks in a document using the standard catalogues containing hand-drawn sketches of the watermarks. In many cases the watermarks of a document are covered by the written text such that it is impossible to produce good hand-drawn sketches. But even if it is possible to make a perfect sketch of the watermark, it is still time-consuming and tedious to find an identical watermark among the hundreds of similar ones in the catalogues. Furthermore, all catalogues are incomplete and there is no guarantee that the search for an identical watermark will be successful. Additionally, it appears that the proof of the watermark identity is not sufficient for a reliable dating result. Some additional data must be supplied and taken into account. A way for overcoming most of these drawbacks is the use of computers, which are ideal tools for cataloguing, comparing, and retrieving of huge amounts of data. An image and text database [4,7,8] combined with special software tools for tracing, image processing, and comparison can facilitate significantly the identification task.

ACQUISITION AND DIGITAL PROCESSING OF WATERMARK IMAGES

Watermarks are small deviations in the thickness of paper. Prior to inputting the watermark into the computer, a hardcopy has to be produced. Beta radiography and electron radiography are considered the best methods for clear recording of watermarks because of their accuracy and sensitivity. Minute deviations in the density of the paper are recorded sufficiently well by them. However, even these methods produce hardcopies with low contrast, except some very bright or dark spots and areas caused by holes in the paper or special color ink.

The watermark processing starts with the input of the prerecorded watermark hardcopies. A sensitive flatbed scanner with a transparency extension is used for their scanning. In the preprocessing stage the scanned images are enhanced in order to increase the contrasts, to compensate for scanning artifacts and to reduce the noise. *Figure 1* shows a typical watermark image scanned from a beta-radiographic hardcopy. The objects which are used for the watermark identification are: watermark (a), chain lines (b), and laid lines (c).

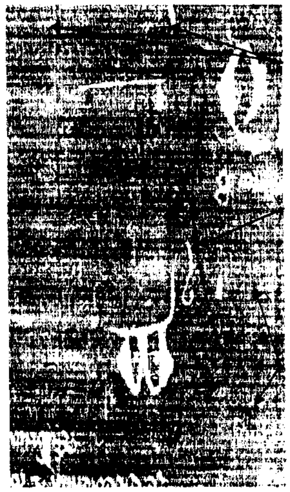


Figure 1: watermark image: watermark (a), chain lines (b), and laid lines (c)

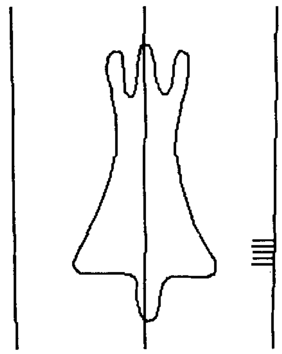
COMPUTER-AIDED EXTRACTION OF WATERMARK CONTOURS

Watermark contours cannot be extracted automatically due to the many artifacts in the images. A semi-automatic procedure was chosen in order to achieve good and fast results. The semi-automatic procedure is designed to minimize user interaction and to optimize the quality of contour tracing. The procedure for watermark contour extraction is implemented as an interactive one. At first, the watermark motif, which defines the overall watermark shape, is determined by user interaction. The watermark motif (e.g. bell, scales, bow and arrow, ox head etc.) with subtypes down to four levels is selected from a hierarchically-structured list and predetermines the number and positions of control points, which have to be finally positioned manually. Their number is kept minimal as far as possible (typically between ten and twenty). Furthermore, the motif also sets geometric restrictions for the variability of the contour.

Figure 2: Watermark image with drawn contours and control points

Then, a semi-automatic procedure based on parametric cubic Bézier curves [1] is used for the approximation of the watermark contours [9]. The control points as chosen by the user divide the watermark contours into sets of adjacent contour segments. The resultant Bézier curve interpolates the two control points of each contour segment and approximates the other two. These additional control points are calculated under the criterion to fit the watermark contour as good as possible. The two main steps of this procedure are demonstrated in *Figure 2* and *Figure 3*. By dragging the control points, the user has the opportunity for fine-tuning of contour segments. The final set of control points is stored in a relational watermark database.

The Bézier curve is affine invariant. This means that any linear



transformation of its control points defines a new Bézier curve, which is just the same transformation of the original curve. So, an extracted contour can be easily adjusted for other watermarks of the same motif.

RELATIONAL DATABASE OF WATERMARKS

A relational data model was chosen for the development of the watermark database. All data are organized in a set of related tables. The database contains two main parts. One deals with the watermark classification according to their textual description. Each watermark motif has a defined place in a hierarchical structure of four levels of watermark subtypes. According to the position in the structure, a unique digital code is generated for each watermark type registered in the database. It is an eleven-digit code consisting of three preceding digits for the watermark type code and four subsequent pairs of digits for subordinated subtype codes. The capacity of this code is enough for coding up to 999 watermark types and up to 99 watermark subtypes in each of four possible subordinated levels. This hierarchical structure is based on the watermark motif classification rules [3].

The second part of the database is used for the registration and management of concrete watermark entities. This part of the database contains a large set of related data required for the classification and complete description of watermarks. All watermark images are stored in this part of the database. Our fast growing database contains currently more than 3,000 watermarks. The main part of these watermarks was captured from manuscripts kept in the library of Klosterneuburg monastery (Austria).

SOFTWARE TOOLS

A large set of software tools is implemented together with the database management in one system. User interaction with this system is realized through a graphical user interface containing a set of menus, buttons, and other controlling items. This graphical user interface was designed following the Microsoft Windows-style.

One module of the system supports the processing and visualization of watermark images and the extraction of contours. The database allows complex textual as well as graphical queries. The software tools include computer-aided measurement of metrical parameters, watermark contour extraction, watermark classification, and automation of other routine jobs. A snapshot from a typical user session is presented in *Figure 4. (pg. 264)*

This snapshot shows three windows: one displaying the watermark image (left side of the screen), one for controlling the watermark contour extraction process (center), and one for the textual watermark description (right side). The system is implemented on a PC platform and runs under Windows 95/98/NT-4.0/2000.

SUMMARY

Methods and tools for digital processing of watermark images and storing them in a relational watermark database were developed and implemented. All methods and tools are integrated in a software system, which allows the user to perform all necessary tasks for watermark classification within one system. The system is fully implemented and has proven its usefulness in a production environment.

ACKNOWLEDGMENTS

This work is supported by the Austrian Science Fund (grant FWF-Projekt P13298-ARS) and INTAS.

REFERENCES

1. Bézier P., *Numerical Control: Mathematics and Applications*, translated by A.R.Forrest and A.F.Pankhurst, John Wiley & Sons, London, 1972

2. Briquet Ch.M. Les Filigranes. *Dictionnaire historique des marques du papier dés leur apparition vers 1282 jusqu'en 1600.* Paris 1907

3. International Association of Paper Historians – IPH, *International Standard for the Registration of Watermarks*, P.F. Tschudin, ed., Rienen, Switzerland, 1997.

4. Karnaukhov V.N., Merzlyakov N.S., Wenger E., Haidinger A., and Lackner F. Digital Analysis of Watermarks of Medieval Manuscripts. In *Computer Optics*, Vol.14-15, 1995, 11-24 (in Russian)

5. Karnaukhov V.N., Wenger E., Merzlyakov N.S., Haidinger A., Lackner F. Thematic processing and retrieving of watermarks. In *Proceedings of SPIE* Vol. 2363, 1995, 32-39

6. Piccard G. *Die Wasserzeichenkartei Piccard im Hauptstaatsarchiv Stuttgart*. Findbuch I-XV, W. Kohlhammer, Stuttgart, 1966-1987

7. Rauber C., Ruanaidh J., Pun T. Secure distribution of watermarked images for a digital library of ancient papers. In *Second ACM international conference on digital libraries*, Philadelphia, ACM, 1997

8. Rauber C., Tschudin P., Startchik S., Pun T. Archival and retrieval of historical watermarks. In *Proceedings of ICIP 1996 - International conference on image processing*, Lausanne. IEEE, 1996

9. Wenger E., Karnaukhov V.N., Haidinger A. Contour Extraction of Watermarks in Old Manuscripts In: Maria Alberta Alberti, Giovanni Gatto and Ivan Jelinek, Eds., *Eurographics '99 Short Papers and Demos*, , Milano, 1999, 236-238

10. Wenger E., Karnaukhov V.N., Haidinger A., Merzlyakov N.S. Digital image analysis for dating of old manuscripts. In: Image Analysis Applications and Computer Graphics. (R. T. Chin, H. H. S. Ip, A. C. Naiman, T.-C. Pong, eds.). In *Lecture Notes in Computer Science 1024*, Springer, Berlin, 1995, 522-523

ABOUT THE AUTHORS

Emanuel Wenger is a researcher at the Commission of Scientific Visualization of the Austrian Academy of Sciences in Vienna. His research interests include computer graphics, digital image processing, and virtual archaeology. He graduated from the University of Technology in Vienna in 1978.
E-mail: *emanuel.wenger@oeaw.ac.at*

Victor N. Karnaukhov is a researcher at the Institute for Information Transmission Problems (IITP) of the Russian Academy of Sciences in Moscow. His research interests include multimedia, digital image processing, and database systems. He graduated at the Moscow State University in 1975. He received his PhD in computer science and image processing from the IITP of the Russian Academy of Sciences in 1986.
E-mail: *victor.karnaukhov@iitp.ru*

Alois Haidinger is a researcher at the Commission of Paleography and Codicology of Medieval Manuscripts in Austria of the Austrian Academy of Sciences. He has published a number of catalogues and articles on medieval manuscripts, mainly from the libraries of Klosterneuburg and Melk, since 1974. He is specially interested in developing methods to use filigranology for dating medieval paper manuscripts.
E-mail: *alois.haidinger@oeaw.ac.at*

Maria Stieglecker is a fellow researcher at the Commission of Paleography of Medieval Manuscripts at the Austrian Academy of Sciences in Vienna. Her research interests include late medieval

manuscripts and their watermarks. She studied history at the University of Vienna and graduated in 1995. In the same year she finished her post-graduate studies at the Institute of Research of Austrian History in Vienna (MAS).
E-mail: *maria.stieglecker@oeaw.ac.at*

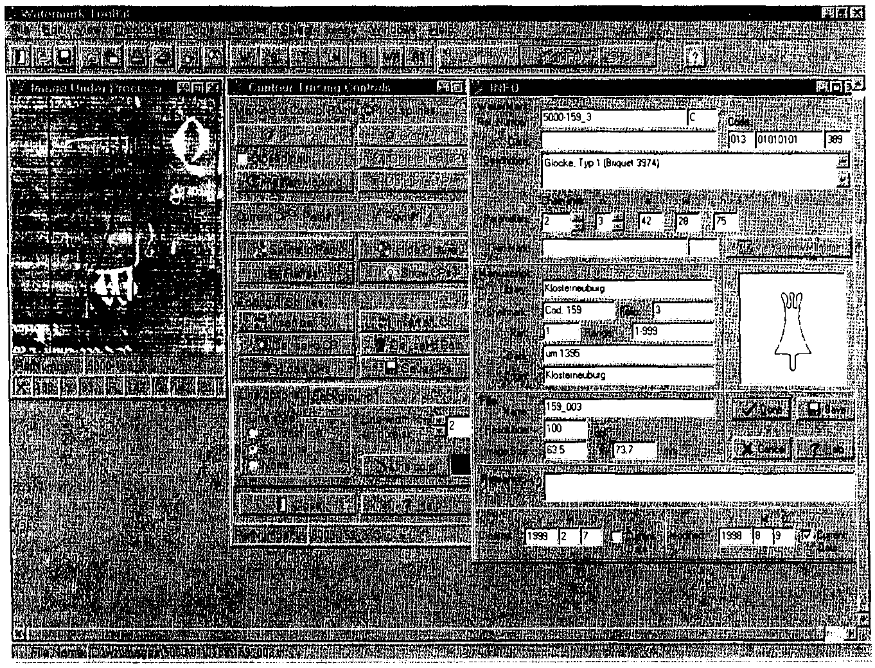


Figure 4: Graphical user interface with three open forms