ECP 2005 CULT 038097/Bernstein

BERNSTEIN

Deliverable no. 19, ref. D2.4
Numerical Paper Description Standards

Deliverable number	*D2.4*
Dissemination level	*Public*
Delivery date	*31 August 2008*
Status	*Final*
Author(s)	*The Bernstein Consortium*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

Introduction

Papers containing no watermarks and/or paper with not accurately assignable watermark motifs cannot be found in the repertories and the watermark databases, since the search is based on an accurate motif assignment. Especially single sheets of paper which are cut or not bound together are hard to identify by watermark analysis, e.g. 600 early Dutch drawings (1400-1630) from the Dresdner *Kupferstichkabinett*. Two thirds of them contain no watermark or only an unidentifiable part of a watermark. In such cases numeric paper parameters could be helpful to identify the paper and to reduce the number of hits in strongly represented watermark motif groups (e.g. "*gothic P*" or "*bulls head*").

Therefore it is a goal of the Bernstein project to find and select numerical paper parameters that allow to compare and to date paper sheets also without watermarks. Direct dating of a paper, only based on its structure seems impossible, since no given paper structure can be assigned only to a certain decade or a certain region. Instead of this fact it was necessary to find paper parameters which are sufficient to allow the identification between dated and undated paper.

Such investigations are summarized by the term "*numeric description of paper*" (WP2, T 2.1b) in the Bernstein project proposal. Some sieve parameters were already measured for years by all databases involved. New parameters were investigated.

Since there are no resources available to insert new parameters for the approx. 100,000 records of the existing databases, the only way is the usage of digital image processing (tools). In the frame of the project the Bernstein Consortium has to restrict to the adaptation and integration of software already existing.

Watermark parameters

The watermark dimensions "width" and "height" are very effective parameters in order to reduce the hit quantity of a watermark search for a certain motif. These parameters are determined by the bounding box of the vertically oriented watermark. The watermark height for complete watermarks is present in all four watermark databases and is one basic search parameter in the Bernstein portal. The dissemination toolkit offers an integrated user-friendly interactive version for the determination of the bounding box and the calculation of the watermark parameters height and width. The width of watermarks is not available in all databases but is a searchable parameter in the Bernstein-portal (beta version: http://bernstein.iicm.tugraz.at:8080/BernsteinPortal).

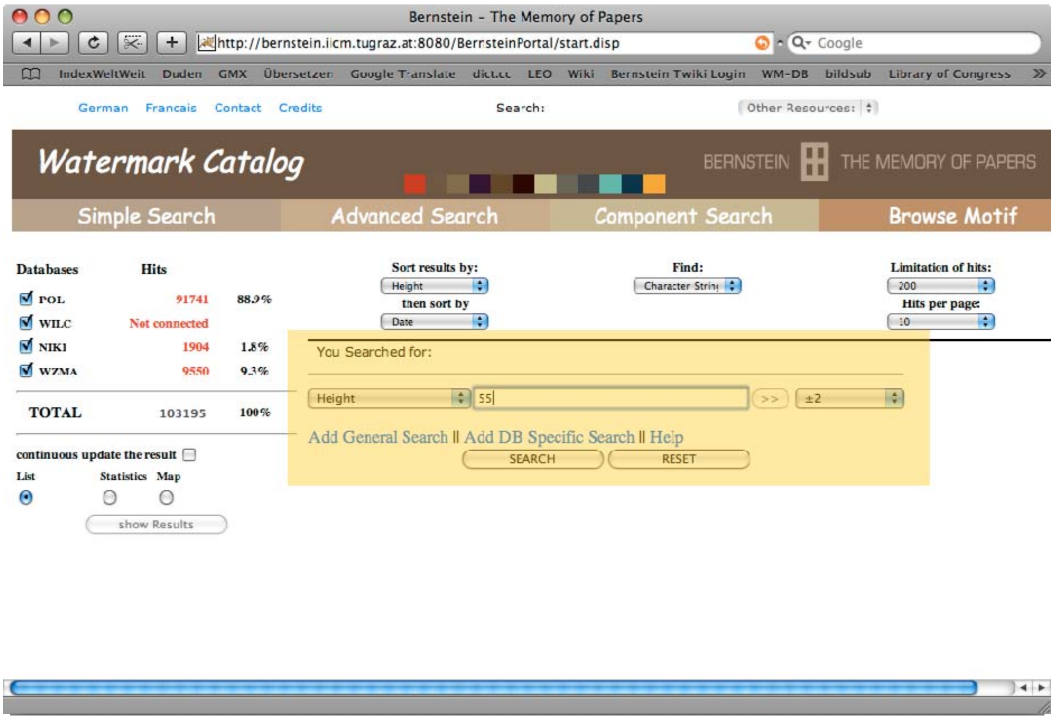

Figure 1: Screen shot of the user surface of the beta version of the Bernstein portal. The search mask – containing the parameter "height" – is highlighted in yellow.

Software for the technical description and comparison of watermarks

In the frame of Bernstein a numerical paper description software, provided by the project partner at Delft University of Technology, has been successfully tested for the watermark collections of Piccard (Stuttgart). The goal was to compare the watermarks from Piccard-Online with those of the printed volumes and the Piccard file cards in order to find the unpublished watermarks as well as doubly published ones. The first step was to scan the file cards and pages of the printed Piccard, afterwards the watermarks were separated automatically piece by piece.

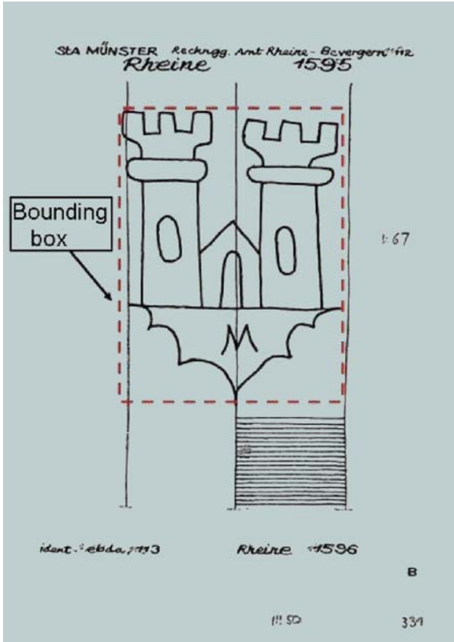

Figure 2: The illustration shows a watermark of the Piccard-Online Collection, which is cut out along the red dashed lines for an automatic picture comparison done by software from the Bernstein-partner of Delft University of Technology.

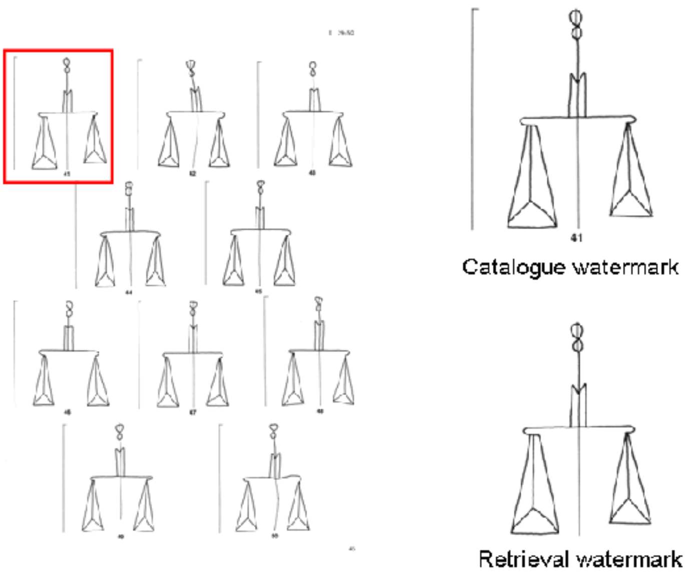

Figure 3: A page from the printed Piccard. The red box contains a single isolated watermark which is forwarded for automatic comparison with the others. On the top of the right hand side the automatically extracted watermark can be seen, below an almost identical watermark was found.

The software which is implemented in MatLab divides the area into equal zones, in which the respective pixels of each single watermark are counted and stored as vector. These vectors from both, the printed volumes and Piccard-Online, are compared and verified in reference to identity.

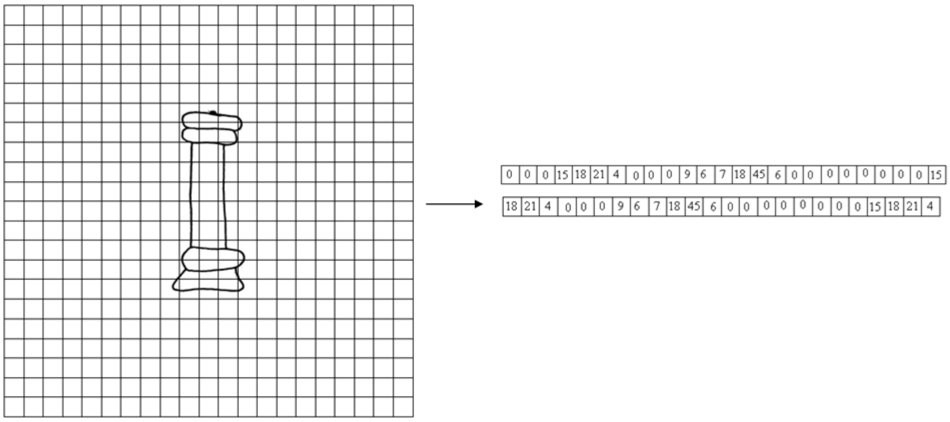

Figure 4: Zone vectors for the comparison of watermarks.

This method is very promising and needs to be developed further in a potential subsequent project. In the current version the software works well with binary images but it does not work with colour or grayscale images. The automatic determination of further numerical parameters like the wire length, or the covered are of the watermark could be another interesting enhancement.

Chain lines distance

The sieve marks, especially the chain lines, are important hints for an allocation of a sheet of paper to a certain region and time. The parameter "chain line distances" is already included in all four databases of the Bernstein project (Piccard-Online, WZMA, NIKI, WILC), but is not part of the search masks in all of them. Nevertheless the chain line distance is one of the major search criterions in the Bernstein portal.

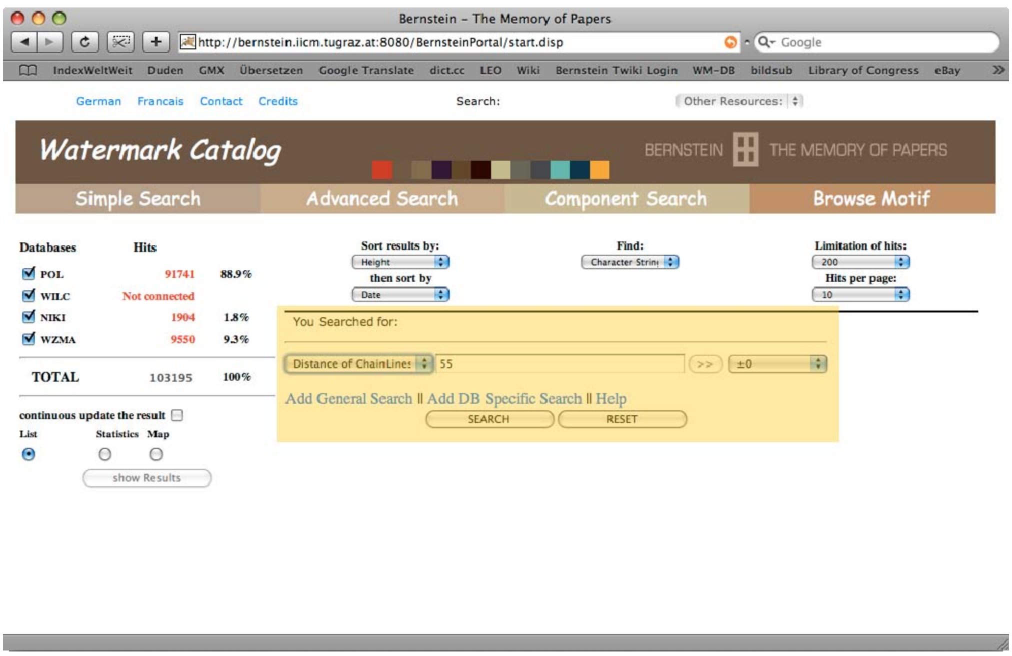

Figure 5: Screen shot of the user surface of the beta version of the Bernstein portal. The search mask with the parameter "distance of chain lines" is highlighted in yellow.

One inconsistency concerning the chain line distance in the databases arises in those cases where one chain line crosses the watermark. In these cases the distance between the two surrounding chain lines is approximately the double distance of two neighbour chain lines (e.g. AT5000-663_198 [WZMA], WM I 00895 [WILC], POLE 85221 [POL]).

Automatic determination of chain line and laid line distances

Besides the watermark, chain lines and laid lines are the most important sieve marks. Both, chain lines and laid lines, are relatively well visible as periodic structures in radiographs and digital subtraction photos even if they contain a lot of noise. Therefore, software for the analysis and determination of chain lines and laid lines can already be used since a couple of time. The distinction between chain lines and laid lines is easy because the distance of chain lines is essentially bigger than the distance between the laid lines i.e. both are in different wave length areas in the FFT-analyses.

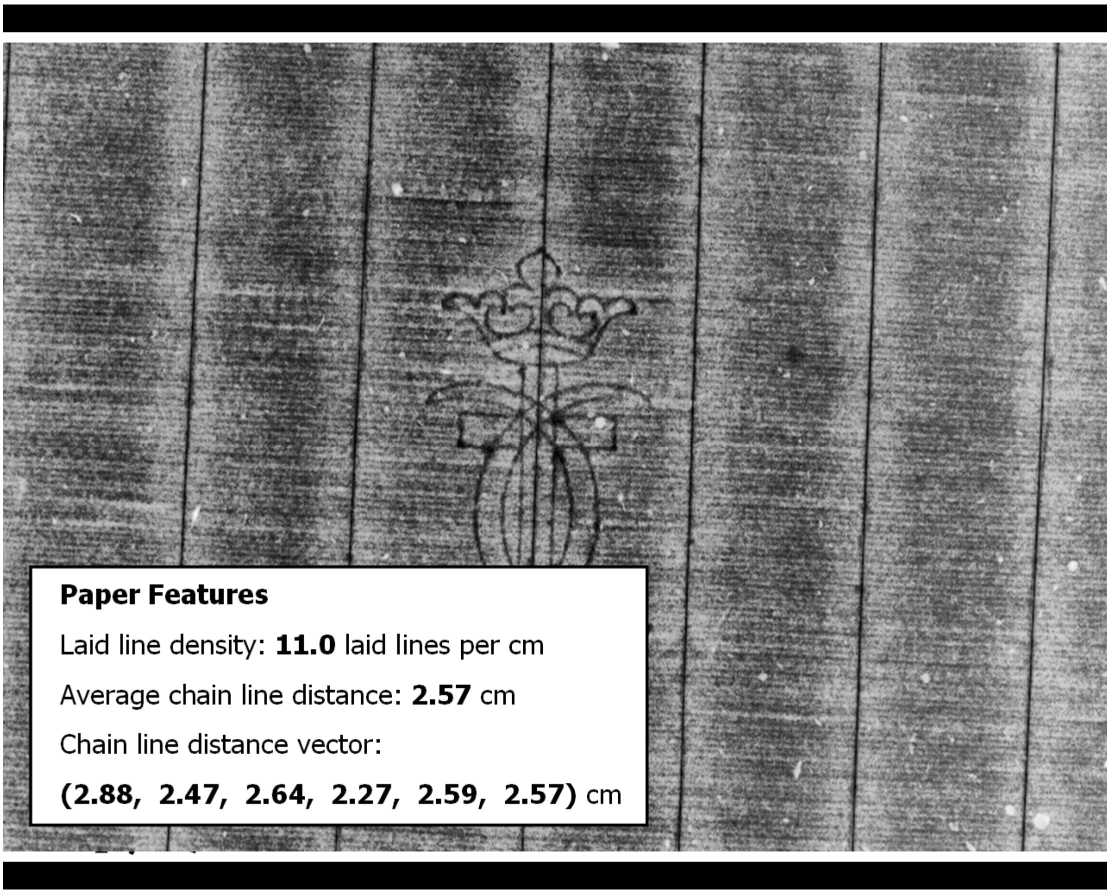

Figure 6: Results of the detection of the chain lines and laid lines in an electron radiography. The detected chain lines are emphasized with black lines.

The results of the chain and laid lines analyses are stored as vectors which can be used for the search and comparison in databases.

Integration of AD751

AD751 is a powerful software tool for the analysis of laid lines. It was developed by Vlad Atanasiu and will be integrated into the Bernstein portal as downloadable software package. Although this software is available and the distance of laid lines is an important parameter, this parameter is not included among the search parameters in the Bernstein portal, because the biggest database Piccard-Online does not provide sufficient information about laid lines. Piccard's drawings focus on the watermarks and the chain lines. Many drawings do not contain laid lines at all and if they do, the chain lines are often only hinted insufficient for a software based analysis.

So far, the software AD751 has been used successfully for the database WILC. It appeared that measuring the laid line distance can limit search results very effectively. Therefore this software is integrated into the Bernstein workspace and its usage is recommended for databases with images of the following types: rubbing, radiography, digital subtraction picture.

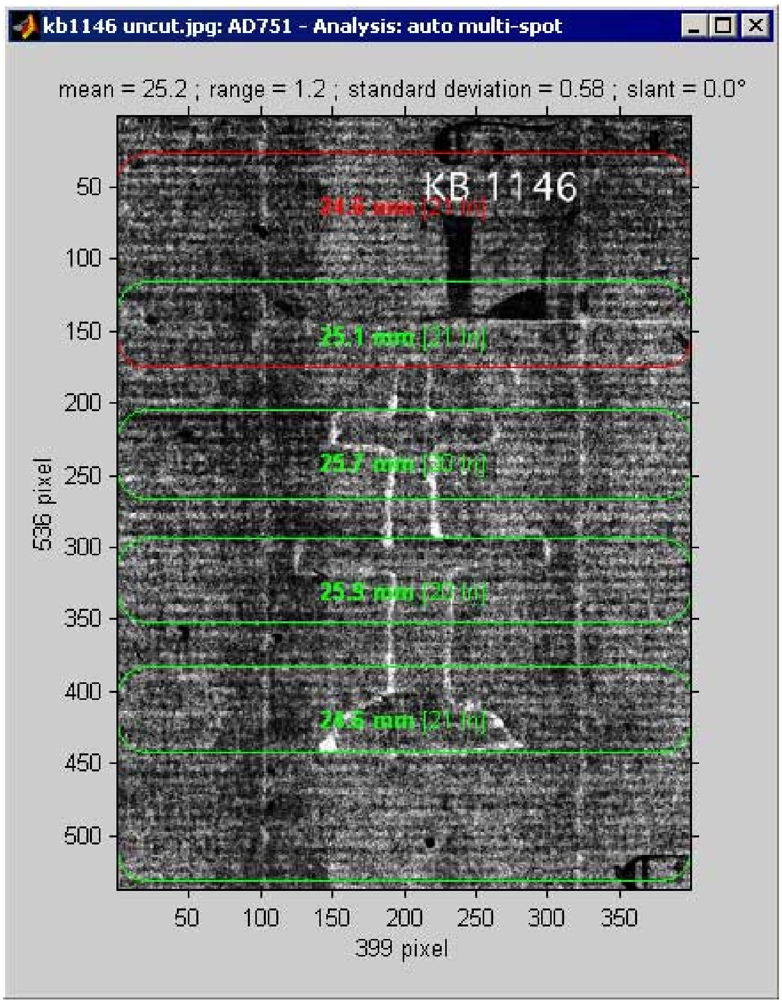

Figure 7: Screen shot of the result of an AD751 analysis.
The laid lines were recognized automatically and the distances measured.

Further paper parameters

There are several numeric parameters which represent further features of the paper mould. These have not been included in any database so far but integration could be of advantage for a better identification of paper. Relevant parameters could be the following:

(1) wire thicknesses,
(2) open sieve areas,
(3) variations in the connections between the wires and position of possible connecting wires at the watermark,
(4) presence and/or width of the shade underneath the chain lines,
(5) distance of "*Bewindedrähte*".

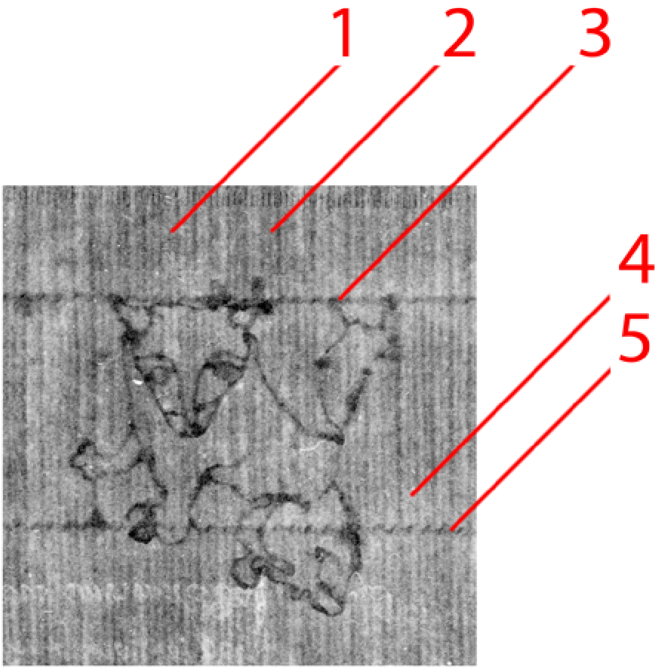

Figure 8: **Part of DE-SLUB-Msc.Dred.A50_fol.046**

Parameter paper quality

A further numerical parameter could be a mathematically accurate defined index for the paper quality. It is quite probable that the best solution to achieve this parameter are digital transmission pictures (see figures 9 and 10).

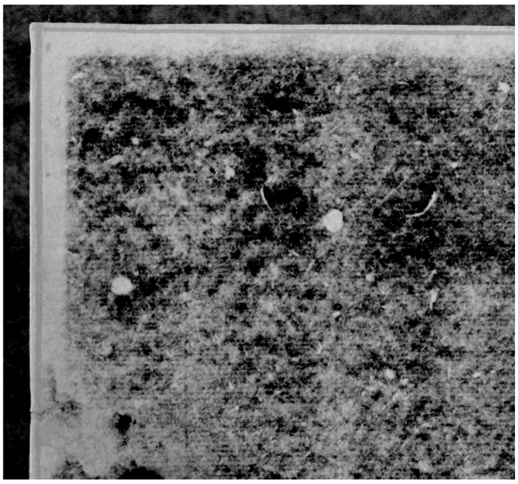

Figure 9: Part of DE-KKDD-C1685. Example of a rather inferior paper quality. Fibrous material lumps, contaminants (big white spots) and folds are clearly visible.

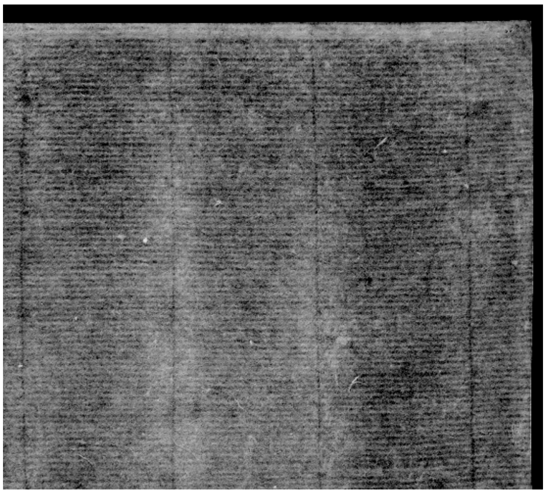

Figure 10: Part of DE-KKDD-C4147. Example of high quality paper. The paper has a very homogeneous structure with even transparency and without contaminants and "*Gautschfalten*".

Sufficient reference data are necessary in order to be able to make a reliable classification of the paper quality. For this reason an automated picture subtraction procedure was tested very extensively and is available for collecting further data. In future, an FFT based noise filter could be developed, which can be connected to a paper quality index.

Radiography methods are not appropriate for visualizing the quality of paper. The fibrous material knots and "*Gautschfalten*" (both signs for paper of inferior quality) are penetrated nearly undamped by x-ray. Only metallic impurities or inscriptions with metal reserve parts appear on the radiographic film. Therefore this picture recording procedure is only very reduced suitable for statements about the paper quality. Furthermore, drawings which are the basis for the watermark collection of Piccard and rubbings (partial in WILC) do not include any information about paper quality.

UV comparison

An illumination and recording with UV (ultra violet) light can reveal paper quality differences even if the paper samples do not show any significant differences in the visible optical spectrum. There is an official paper test and comparison procedure which is used in paper test laboratories for industrially manufactured papers.

Figure 11: UV absorption of different modern office papers.

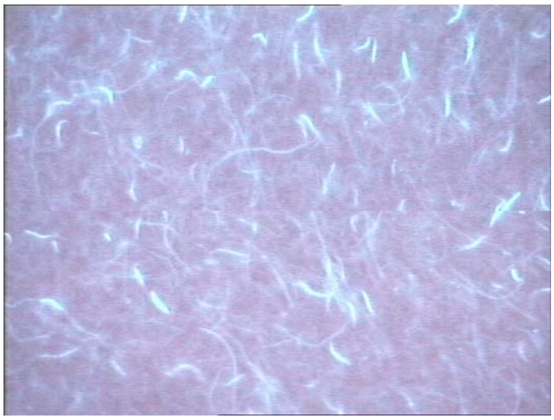

Figure 12: UV absorption of a paper with bleached fibres.

White industrial papers exhibit different absorption behaviour within the UV range due to their different contents materials. It was shown however that not all materials react to the UV stimulation. The fibres must have been bleached for this respectively the paper must have been processed with further additives. This was not the case in the time paper was produced by hand however and therefore the UV comparison might hardly be worthwhile for the data we have in the Bernstein databases.

A further disadvantage is that the UV lamps of the equipment provided for these investigations (‚Docucenter' manufactured by Projectina, used in the paper test laboratory of Proost&Brandt) cannot be calibrated. The results are not reproducible since the intensity of the UV source in the course of its life span diminishes and differs among the devices. Therefore this UV-procedure is suitable for a direct comparison of two or more samples, but not suitable for the purpose of a sample database.

However, for a follow-up project it could be considered whether such a picture-giving procedure could be appropriate to determine the types of the used fibres as well as the fibre lengths, widths, curvatures distributions, entanglements, a.s.o. Hereby, further comparison criteria could be generated.

Medium-term goals

In the near future, numeric paper parameters should make it possible to identify and date papers, no matter whether they contain watermarks or not. A paper identification system without an obligatory watermark motif classification would increase the quantity of users and successful retrievals significantly and so far not datable papers (without watermarks or containing incomplete watermarks) could be dated.

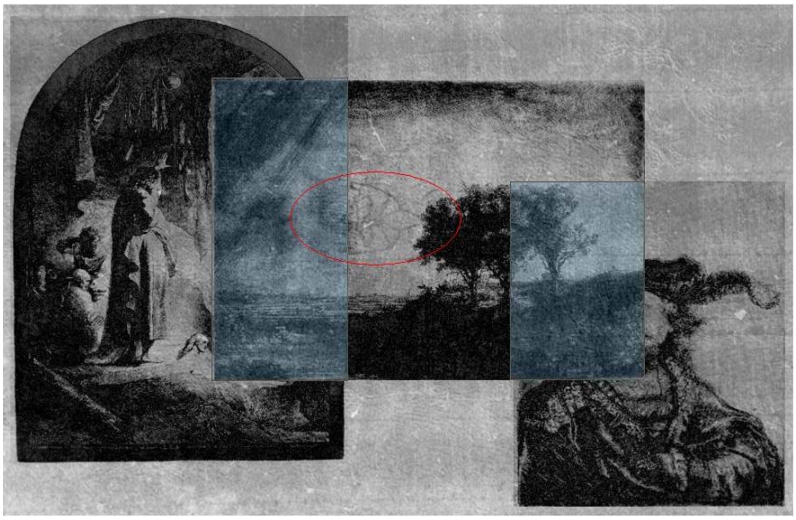

Figure 13: Schematic collage of the goals. Individual papers – also without watermarks – could be assigned to one mould with a virtual mould reconstruction, based on numeric paper structure parameters.

Such a trend has far-ranging consequences regarding the picture recording of papers in future. Especially the full area of not bound objects (single paper sheets) should be recorded instead of recording only an area around the watermark.

As tests revealed, neither rubbings nor radiographs can deal with large objects. Only the digital picture subtraction method can record sheets of a size up to A0. The quality of this method is comparable to radiography methods, which was proven by extensive tests with the collections of the Dresdner *Kupferstichkabinett*.

Conclusions

All watermark databases as well as the Bernstein portal are optimized to identify and date papers with watermarks. Papers without watermarks or with not accurately assignable watermark motifs are hardly to identify by the databases available. Art historians, archivists, and musicologist for example are occupied with papers which predominantly do not contain watermarks or they contain only parts of a watermark.

Therefore it is one of the goals of the Bernstein project to find additional paper parameters, which allow the comparison of papers with and without watermarks if necessary. Direct dating of a paper only on basis of its structure seems impossible. Instead, it appears to be necessary to find sufficient corresponding paper parameters in order to relate undated papers with dated ones. Such investigations were summarized by the term *‚numeric description of paper’* (WP 2.1b) in the Bernstein project. In this context already existing software for numeric paper description has been adapted and tested with selected examples. Here are the results and conclusions:

1.) Already existing numerical parameters in the data bases were integrated as search criteria into the Bernstein portal.
2.) Already existing image analysis software, which calculates numeric paper parameters has been adapted for tests within the Bernstein project and will be downloadable from the portal. There is no software which can handle all types of recordings within the Bernstein databases.
3.) The tests were accomplished very successful on selected data sets. At the moment most of the software is implemented with MatLab. An implementation in a high-level programming language (e.g. C++) would be preferable.
4.) The tested image processing software works with the file formats TIF and PNG and thus confirms with the resolutions and the data of the Bernstein consortium.
5.) In the frame of Bernstein, further possible developments in respect to a numerical paper description could be demonstrated and studied. The future will show, in what respect these can be developed in subsequent projects to achieve effective paper comparison parameters.
6.) To support future development of virtual mould reconstructions and data mining it is recommended to record already the full area of paper and not only the watermark in current projects.