The 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage
VAST (2007)
D. Arnold, F. Niccolucci, A. Chalmers (Editors)

Registration of Multi-Spectral Manuscript Images

M. Diem, M. Lettner and R. Sablatnig

Pattern Recognition and Image Processing Group
Institute for Computer Aided Automation, Vienna University of Technology
diem@prip.tuwien.ac.at

Abstract

Two medieval Slavonic manuscripts are recorded, investigated and analyzed by philologists in collaboration with computer scientists. The aim of the project is to develop algorithms that support the philologists by automatically deriving the description and restoration of the scripts. The parchment partially contains two scripts, where the first script was erased. In order to enhance the erased script, the manuscript pages are imaged in seven bands between 330 and 1000 nm. A registration, aligning the resultant images, is necessary so that further image processing algorithms can combine the information gained by the different spectral bands. Therefore, the images are coarsely aligned using rotationally invariant features and an affine transformation. Afterwards, the similarity of the different images is computed by means of the normalized cross correlation. Finally, the images are accurately mapped to each other by the local weighted mean transformation. The algorithms used for the registration and preliminary results are presented in this paper.

Categories and Subject Descriptors (according to ACM CCS): I.4.3 [Image Processing and Computer Vision]: Registration

1. Introduction

Since ancient times vellum was used to write on, but the laborious manufacturing made vellum a valuable material. This is why the scripts were erased and the parchment used again. The so-called palimpsests (Greek: *palimpsestos* - scraped again) often comprise vestiges of the original text. During the 19th century scientists used chemical means to read palimpsests that were sometimes very destructive, using tincture of gall or later, ammonium hydrosulfate. Modern methods of reading palimpsests using multi-spectral image acquisition are not damaging.

In this paper the results of a multi-spectral image acquisition system and a method for a fully automatic image registration are presented. The aim of multi-spectral imaging is to maximize the contrast between the erased and the second script as well as enhancing and making the erased script visible respectively. Since manual operations such as filter changes or camera changes are performed between the acquisitions, a registration that aligns one spectral image to the other is necessary. If the images are registered, they can be combined, using for instance a principal component analysis.

In a previous approach, manuscript images were registered using solely the cross correlation. In order to get reliable results, the template images, which are details of the so-called reference image, needed to comprise at least one character ($\approx 130 \times 130 px$). An image to which all other images, called sensed images, are aligned to is referred to as reference image. Since the cross correlation needed to be computed over approximately a quarter of the sensed image, two similar characters could be mistaken and the operation was computationally intensive. But the main weakness of this approach is the dependency on rotations between the images. Thus, a modified scale-invariant feature transform, which is rotationally invariant too, aligns the images coarsely. Three consistency checks minimize the error made by estimating the affine transformation matrix. Having aligned the images coarsely, the cross correlation is computed between a $16 \times 16 px$ template image and a $32 \times 32 px$ search window of the sensed image. A local weighted mean mapping function, which is a local polynomial transformation, is estimated by means of the corresponding control points. Thus, non-rigid local distortions caused by different filters or the changing curvature of the parchment are corrected.

The paper is organized as follows. The following section characterizes related work and gives an overview in the range of image acquisition. Section 3 and Section 4 describe the image acquisition and the image registration in more detail. Numerical and visual results of the stated methods are given in Section 5. Finally, the last section gives a conclusion and an outlook.

2. Related Work

There have been efforts in image analysis of historical documents [BPP*03,EKCB03,RB05]. In general, differences between image analysis of ancient versus modern documents result particularly from the aging process of the documents. Some related studies in image analysis of historical documents are covered in this section.

Multi- and hyper-spectral imaging has been used in a wide range of scientific and industrial fields including space exploration like remote sensing for environmental mapping, geological search, medical diagnosis or food quality evaluation. Recently, the technique is getting applied in order to investigate old manuscripts [BPP*03]. Two prominent representatives are the Archimedes Palimpsest [EKCB03] and Tischendorf's Codex Sinaiticus [Eco05]. Easton et al. were the first to capture and enhance the erased writing of the famous Archimedes palimpsest by multi-spectral methods [EKCB03]. The system they propose is modeled on the VASARI illumination system developed at the National Gallery of London [MCSP02]. In that project it turned out that the adoption of spectral imaging produces higher and better readability of the texts than conventional thresholding methods. Balas et al. developed a computer controllable hyper-spectral imaging apparatus, capable of acquiring spectral images of 5nm bandwidth and with 3nm tuning step in the spectral range between 380-1000 nm [BPP*03]. This device was selected as the instrument of choice for the Codex Sinaiticus Digitization Project conducted by the British Library in London [Eco05]. Spectral images of palimpsests and other 'latent' texts have also been enhanced by the Italian company Fotoscientifica Re.co.rd.® (http://www.fotoscientificarecord.com/) which provided for instance the pictures for the EC project *Rinascimento virtuale*, devoted to the decipherment of Greek palimpsest manuscripts [Har02]. The EC project IsyReaDeT developed a system for a virtual restoration and archiving of damaged manuscripts, using a multi-spectral imaging camera, advanced image enhancement and document management software [TBS04].

Two cameras, in contrast to the mentioned imaging systems, are used in our approach. A grayscale camera with an automatic filter wheel takes seven images in different spectral bands. Additionally, color images and UV fluorescence images are taken with a second camera. By aligning the images from both cameras with each other up to ten channels per pixel are available for further processing steps.

Image registration is a fundamental image enhancement task and subsequent image analysis methods depend on the quality of the registration. Due to the diversity of the different image registration tasks, it is impossible to design a method that is applicable for all tasks. When a registration method is designed, the feature detection as well as the transformation have to be selected considering the expected distortions.

3. Image Acquisition

Since photographic techniques in the visible range have proven to be insufficient with palimpsests, spectral imaging has to be applied [BPP*03,RB05]. Images in different wavelengths provide information that is invisible to the human eye [RB05]. Generally, there are narrow spectral bands at which the maximum difference in the reflectance characteristics of each ink exists. The aim of multi-spectral imaging is to provide spectral image cubes, where the third dimension contains spectral information for each pixel. Combining the spectral information enhances the erased script.

For the acquisition of the manuscripts a Hamamatsu C9300-124 camera is used. It records images with a resolution of 4000×2672 pixels and a spectral response between 330 and 1000 nm. A lighting system provides the required UV illumination. Cut-off filters fixed on the lens of the camera select specific spectral ranges. The near UV (320 nm - 440 nm) excites, in conjunction with specific inorganic and organic substances, visible fluorescence light [Mai03]. Up to now the historical manuscripts are recorded with UV fluorescence and UV reflectography. UV fluorescence shows only changes in the upper script layer. In principle, grabbing the visible fluorescence of objects is possible with every camera. UV reflectography is used to visualize retouching, damages and changes through e.g. luminescence. Therefore the visible range of light has to be excluded in order to concentrate on the long wave UV light. This is achieved by applying cut-off filters as well as using exclusively UV light sources.

The image acquisition system is currently improved. Thereby an automatic filter wheel is mounted on the Hamamatsu camera. Seven filters allow the recording of the document pages in seven different bands ranging from 330-1000 nm. Additionally, a RGB color image and a UV fluorescence image of each manuscript page are taken using a Nikon D2Xs. Due to the automatic image acquisition system the registration of the images is solely needed for the correction of the differing distortions caused by the filter changes. Therefore, a simple correlation based approach and a consecutive local transformation could be applied. Since the grayscale images, taken in varying bands with the Hamamatsu camera system, shall be aligned to the color images taken with the Nikon camera, a more extensive registration method needs to be implemented. The image acquisition system is illustrated in Figure 1. The palimpsests are first cap-

tured with the Hamamatsu camera and then moved in order to image them with the Nikon camera.

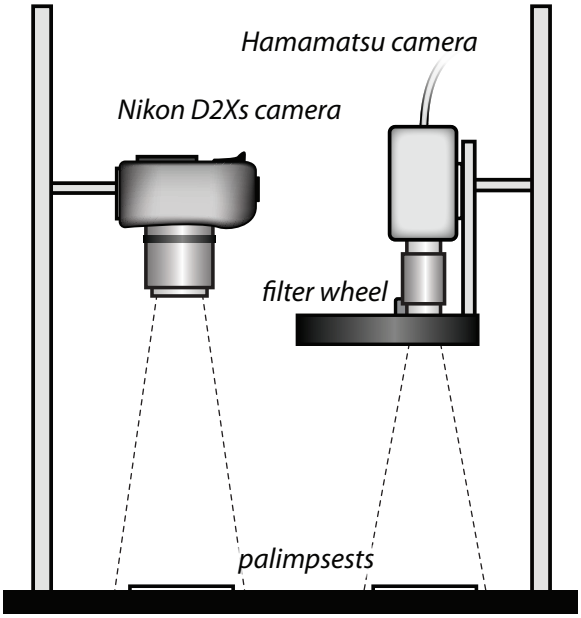

Figure 1: *Illustration of the acquisition system, with the Hamamatsu and the Nikon camera.*

4. Image Registration

Following the acquisition of the manuscripts, the images have to be registered. Image registration is a fundamental task in image processing used to match two or more pictures taken under different conditions.

As stated above image registration is the process of estimating the ideal transformation between two different images of the same scene taken at different times and/or different viewpoints. It geometrically aligns images so that they can be overlaid. There is a wide variety of different methods (especially in remote sensing and medical imaging applications) like the use of corresponding structures or mapping functions [RJ99] which can be adapted for this application. An overview of image registration methods is given by Zitovà and Flusser [ZF03].

4.1. Previous Work

An automatic image registration has been implemented for a previous project. Thereby the control points are localized using an Otsu thresholding approach [Ots79]. In order to reduce low frequency image effects caused by the aging of the manuscripts the images are convolved with a homomorphic filter. Having localized the control points in one image and enhanced both images the correspondence between the images is computed using a normalized cross correlation. The cross correlation needs to be computed between a template image, which has the size of the currently observed character ($\approx 130 \times 130px$), and the whole sensed image. Thus, computing the correspondence as described is computationally more expensive than computing feature descriptors and matching the feature vectors. Certainly, the cross correlation could be computed between the template image and a search window which is an image detail of the sensed image. However, if the translation between the two images is greater than the chosen amplification of the search window, the correlation fails. Another weakness of the cross correlation is that it is neither rotationally nor scale invariant. Hence, the images registered need to have a similar scale and rotation.

4.2. Coarse Registration

Lowe first introduced the Scale-Invariant Feature Transform (SIFT) in 1999 [Low99]. In order to get a scale-invariant feature representation Lowe proposes to compute a scale-space. The control points are detected by computing the local maxima and minima of the Difference-of-Gaussians. Having assigned the orientation to each control point, a local image descriptor is computed which is normalized by the orientation of the control point. Afterwards, the control points are matched using the Best-Bin-First search algorithm.

Since the computation of the scale-space is computationally expensive and the size of the objects is similar in the different images, the scale-space is not computed in our approach. Thus each control point detected has the same scale. In order to detect control points each image is convolved with a gaussian filter kernel having a variance of $\sigma = \sqrt{2}$. The smoothed image is again convolved with the same kernel, which results in an effective smoothing of $\sigma = 2$. The Difference-of-Gaussians is obtained by subtracting the second from the first smoothed image. In order to find local maxima and minima each pixel of the Difference-of-Gaussians image is compared to its 8 neighbors. If its intensity value is greater or lower than the intensities of all neighbors a local extremum is detected and the feature vector is computed. This method detects more than 5000 control points in a $391 \times 493px$ image due to background noise. This is why each pixel is thresholded before it is selected for the local extrema computation. Figure 2 shows a Difference-of-Gaussians image with the local extrema represented by points.

The orientation assigned to each control point is computed similar to Lowe's implementation [Low04]. First the image gradient magnitude $m(x,y)$ and the orientation $\theta(x,y)$ are computed for each pixel of the smoothed image $L(x,y)$.

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = arctan2((L(x,y+1) - L(x,y-1)), (L(x+1,y) - L(x-1,y)))$$

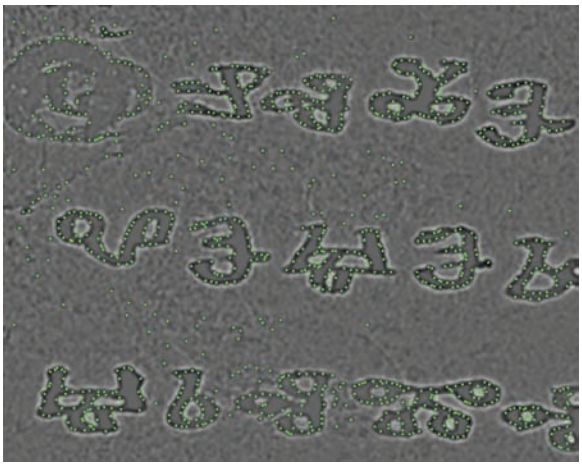

Figure 2: *Detail of a test image after the computation of the Difference-of-Gaussians with $\sigma = \sqrt{2}$. The points represent local extrema.*

An orientation histogram with 36 bins corresponding to 360° is created. Each sample added to the histogram is weighted by its gradient magnitude and a Gaussian weight. Afterwards, the histogram is smoothed with a Gaussian kernel. The maximum of the histogram indicates the dominant direction of local gradients. Figure 3 shows a test image where the white arrows indicate the dominant orientation of each control point.

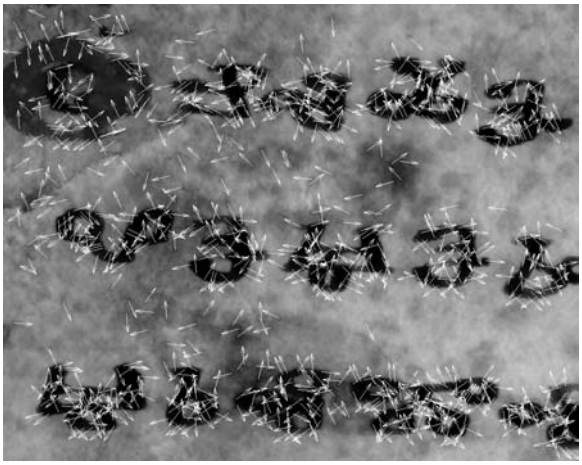

Figure 3: *Detail of a test image. The white arrows illustrate the orientation for each control point. Since the scale-space is not computed, all control points have the same scale.*

In order to compute a local descriptor that characterizes each control point the image gradients $m(x,y)$ and the orientations $\theta(x,y)$ in a $16 \times 16px$ window around each control point are considered. The coordinates of the descriptor and the gradient orientations are rotated relative to the control point orientation so that the features are rotationally invariant. Each gradient is weighted by a Gaussian window of $\sigma = 8$ so that the descriptor does not change significantly with small changes in the position of the window. The control point descriptor consists of eight 4×4 planes where each plane represents the spatial distribution of the gradients for eight different directions. The location of a gradient in the local descriptor depends on the rotated coordinates and the orientation. Each gradient is interpolated to its eight neighbors of the control point descriptor.

After the features are computed for both images, they are matched using the nearest-neighbor algorithm. The Euclidean distance between each control point of the reference and the sensed image is computed. The correspondence of two control points is indicated by the minimal Euclidean distance. Since a control point may be present in only one of the two images, corresponding control points are rejected if their distance to the nearest-neighbor is less than 0.8 times the distance to the second-nearest neighbor. Control points which have more than one correspondence are discarded too. Having discarded the control points according to this scheme ≈ 200 corresponding control points are left for an image with $391 \times 493px$.

An affine transformation matrix is computed using the least squares solution and all corresponding control points. Afterwards, all control points of the sensed image are transformed with the inverse transformation. Their coordinates are subtracted from the coordinates of their corresponding control points in order to get an error estimation. Those control points with an error below the mean error are chosen for the final computation of the affine transformation matrix.

4.3. Cross Correlation

Having aligned the two images coarsely using modified SIFT Features and a global affine mapping function, a normalized cross correlation is computed at the locations of the previously found control points. The aim of the cross correlation and the subsequent local mapping function is to correct non-rigid distortions. The features detected in the images can be matched by means of the image intensity values in their close neighborhood, the feature spatial distribution, or the feature symbolic description [ZF03]. Cross correlation is an area-based method which does not need features of images which have to be registered. The location of the control points are detected exclusively in the reference image in order to avoid false correspondence. Since the images are coarsely aligned, the search windows in the sensed image are set to the same locations as the template images.

The cross correlation calculates the difference of two image details by means of a modified Euclidean distance. The

size of the template image is $16 \times 16px$. Since the images are coarsely aligned by an affine transformation, the search window needs not to be larger than twice the template image. Having defined the image details which need to be compared, the template image is shifted over the entire detail of the sensed image. For each shift the correlation between the template and the search window is computed:

$$c(m,n) = \sum_x \sum_y f(x,y)\, t(x-m, y-n)$$

where $f(x,y)$ denotes the gray values of a detail of the sensed image and $t(x,y)$ the template image. Varying m and n shifts the template over the search window. The resultant function $c(m,n)$ indicates the strongest correspondence of the template image and the search window by the absolute maximum. Hence the control point of the sensed image is placed at the coordinates of the absolute maximum.

The cross correlation is variant to changes in the image amplitude caused, e.g., by changing lighting conditions. Consequently, the correlation coefficient normalizing the template as well as the search window of the sensed image is computed. The dynamic range of the normalized cross correlation $\gamma(m,n)$ moves, independently to changes in the image amplitude, between -1 and 1.

According to the templates magnitude and the proportion of the template and the search window, the computation performs better in the frequency domain than in the spatial domain.

4.4. Local Transformation

Having determined the control points, the parameters of the mapping function are computed. Images which possess only global distortions (e.g. rotation) may be registered with a global mapping function. Likar and Pernuš mentioned that the global rigid, affine and projective transformations are most frequently used [LP01]. As a consequence of non-rigid distortions such as the changing lenses, illumination or curvature of a single page, the images have to be registered using a curved transformation.

Using a global mapping function is practicable only when a low number of parameters defining the transformation are needed (e.g. for rigid or affine transformations). Transformations using polynomials of order n are defined by at least $n+1$ parameters, which results in a complex similarity functional that has many local optima. To overcome this problem a local mapping function is applied. The local weighted mean method [Gos88] is a local sensitive interpolation method. It requires at least 6 control points which should be spread uniformly over the entire image. Polynomials are computed by means of the control points. Thus, the transformation of an arbitrary point is computed by the weighted mean of all passing polynomials. Besides, a weighting function is defined which guarantees that solely polynomials near an arbitrary point influence its transformation.

5. Results

Having discussed the implemented methods, their results are presented in this section. Both the modified scale-invariant features and the normalized cross correlation have been tested on real manuscript images. In contrast to the cross correlation which was tested with palimpsests, imaged as UV fluorescence and UV reflectography, the SIFT features were tested with color images. For testing purposes, one image has been transformed with a randomly generated transformation matrix. Since the affine transformation is a global non-rigid transformation method, it is possible to compute the error of the transformation matrix.

Resulting images of the methods are presented in Figure 4 - 8. Figure 4 shows the corresponding control points of the SIFT features before the previously described consistency checks. The corresponding control points, which are not discarded, can be seen in Figure 5. Due to the fact that more than 50% of the control points matched before the consistency check was performed, there are no false correspondences left. Additionally, the second image was transformed after the consistency check. The Figures 6a and 6b show the sensed image subtracted from the reference image. White pixels indicate high differences between both images. In order to visualize the differences with more contrast, the subtraction image was thresholded in Figure 6b.

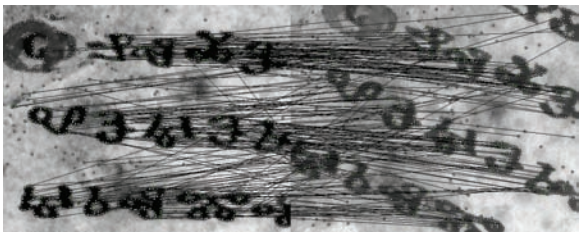

Figure 4: *The corresponding control points between two manuscript images before the consistency check. As can be seen, there are still wrong correspondences. White boxes represent control points without a corresponding one.*

Figure 7a shows the template image with a control point located in its center. The search window with the corresponding control point of the UV reflectography image is shown in Figure 7b. Computing the normalized cross correlation of these two images results in the third image (see Figure 7c) where the strongest peak shows the maximum. The second peak in Figure 7c refers to the high similarity between the template image and the second 'e' in the upper right corner of Figure 7b. Both, the template image ($54 \times 67px$) and the seek window ($310 \times 349px$) of the sensed image are scaled for a better visualization.

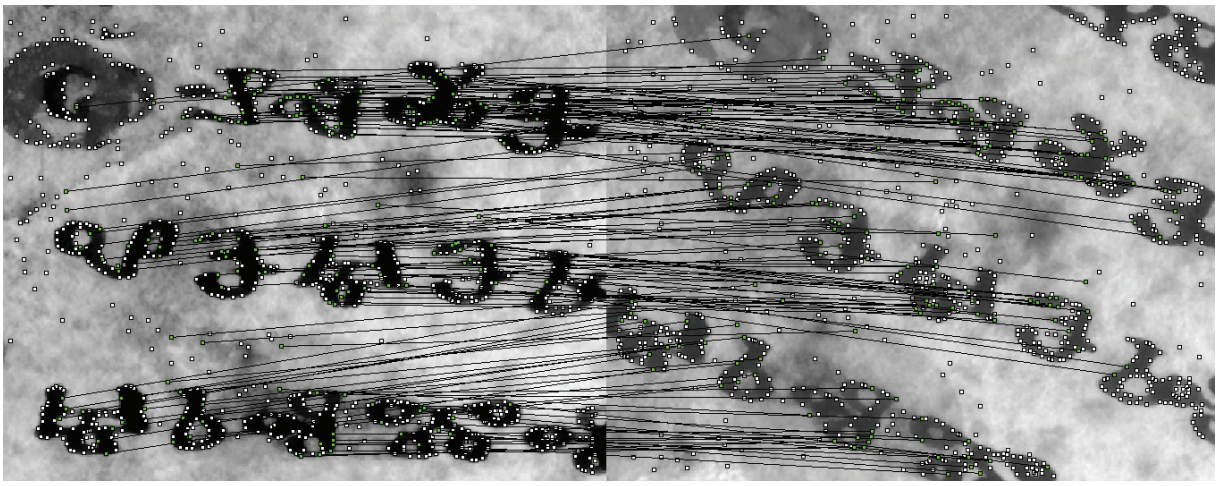

Figure 5: *The corresponding control points after the consistency check. The sensed image was rotated about* 21°*. In contrast to Figure 4, this image contains no wrong correspondences. After the consistency check,* 67.6% *of the corresponding control points remain.*

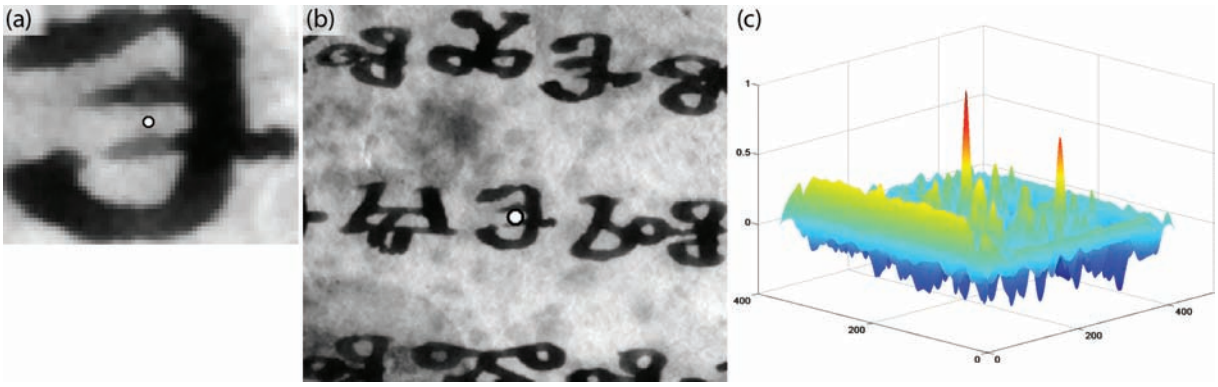

Figure 7: *A control point at the center of the template image (a), which is* 54×67*px for a better demonstration. The search window of the sensed image with the corresponding control point (b). The normalized cross correlation where the strongest peak indicates the location with the strongest correlation (c). The second peak with a lower magnitude corresponds to the 'E' in the upper right corner.*

The local weighted mean method was compared to a global interpolation method (see Figure 8). Since the images possess local non-linear distortions, the registered images correspond only to certain parts of the images if an affine mapping function is applied. Hence, the farther the points are away from the corresponding area, the more they differ. That is why a local sensitive mapping function is applied to compute the transformation after the images are coarsely aligned with an affine transformation.

Additionally, the accuracy of the cross correlation was tested. This was done by calculating the mapping function using 50% of the control points. Then the excluded control points were mapped one to each other by this model. Table 1 shows the matching error computed by the absolute difference in pixel between the corresponding control point and the transformed point.

6. Conclusion and Outlook

This paper introduces a multi-spectral image acquisition system for ancient manuscripts. Multi-spectral imaging allows philologists to analyze ancient manuscripts contactless. In Addition, supplementary information is gained, visualizing

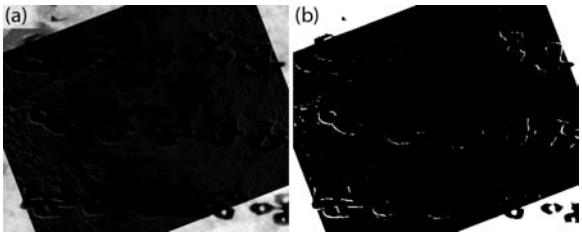

Figure 6: *Image (a) shows the subtraction of the sensed image with the reference image. Bright pixels indicate poorly matching areas. Image (b) shows the same subtraction image thresholded for a better visualization.*

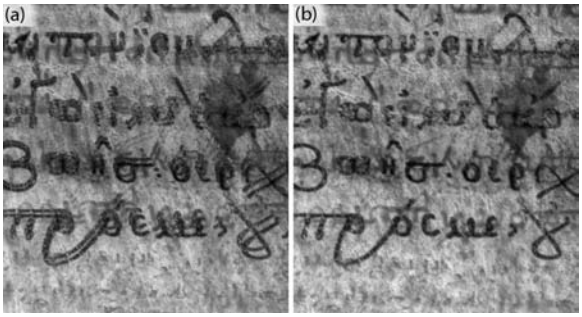

Figure 8: *UV fluorescence image registered to the UV reflectography image using an affine mapping function (a). Registration with the same control points using a local weighted mean transformation (b).*

characters of the erased script that cannot be seen by the human eye.

Furthermore a fully automatic registration, aligning two different images with each other, was depicted. The described approach was compared to a previous registration method. Besides discussing the proposed registration approach, the methods were tested on real images. Additionally, numerical results of the cross correlation's accuracy were given in Section 5.

The registration method proposed is planned to be compared to some others (e.g. [LP01]) and evaluated by applying it to synthetic images. The aim of the current project is a combination of the acquired images by means of a principal component analysis or comparable methods, in order to enhance the erased script. Furthermore, the front side shall be registered with the reverse. Consequently, philologists can distinguish between translucent characters and characters of the erased script.

Acknowledgment This work was supported by the Austrian Science Foundation (FWF) under grant P19608-G12.

IMAGE	N	MEAN ERROR		MAX ERROR	
		x	y	x	y
image 1	26	0.43	0.60	1.59	1.81
image 2	28	0.51	0.36	1.19	0.85
image 3	24	0.51	0.48	1.24	1.43
image 4	26	0.32	0.60	0.88	1.48
image 5	30	0.46	0.64	1.43	1.89
image 6	24	0.43	0.55	1.05	1.79
image 7	30	0.47	0.56	1.18	1.69
image 8	27	0.43	0.47	1.09	1.27

Table 1: *Matching error of the Cross Correlation, where* N *indicates the number of control points.*

References

[BPP*03] BALAS C., PAPADAKIS V., PAPADAKIS N., PAPADAKIS A., VAZGIOURAKI E., THEMELIS G.: A novel hyper-spectral imaging apparatus for the non-destructive analysis of objects of artistic and historic value. *Journal of Cultural Heritage 4* (January 2003), 330–337.

[Eco05] And the word was made flash. *The Economist*, March 23rd 2005.

[EKCB03] EASTON R. L., KNOX K. T., CHRISTENS-BARRY W. A.: Multispectral imaging of the archimedes palimpsest. In *32nd Applied Image Pattern Recognition Workshop, AIPR 2003* (Washington, DC, October 2003), IEEE Computer Society, pp. 111–118.

[Gos88] GOSHTASBY A.: Image registration by local approximation methods. *Image and Vision Computing 6* (1988), 255–261.

[Har02] HARLFINGER D.: Rediscovering written records of a hidden european cultural heritage. In *Berichtband der Konferenz des Netzwerks Rinascimento virtuale zur digitalen Palimpsestforschung* (2002), pp. 28–29.

[Low99] LOWE D. G.: Object recognition from local scale-invariant features. In *International Conference on Computer Vision* (Korfu, 1999), pp. 1150–1157.

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110.

[LP01] LIKAR B., PERNUŠ F.: A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing 19* (2001), 33–34.

[Mai03] MAIRINGER F.: *Strahlenuntersuchung an Kunstwerken*. E. A. Seemann Verlag, 2003.

[MCSP02] MARTINEZ K., CUPITT J., SAUNDERS D., PILLAY R.: Ten years of art imaging research. *Proceedings of the IEEE 90* (2002), 28–41.

[Ots79] OTSU N.: A threshold selection method from grey

level histograms. *IEEE Transactions on Systems, Man, and Cybernetics 9* (1979), 62–66.

[RB05] RAPANTZIKOS K., BALAS C.: Hyperspectral imaging: potential in non-destructive analysis of palimpsests. *International Conference on Image Processing, ICIP 2005 2* (2005), 618–621.

[RJ99] RICHARDS J. A., JIA X.: *Remote Sensing Digital Image Analysis: An Introduction*. Springer, 1999.

[TBS04] TONAZZINI A., BEDINI L., SALERNO E.: Independent component analysis for document restoration. *International Journal on Document Analysis and Recognition 7* (March 2004), 17–27.

[ZF03] ZITOVÁ B., FLUSSER J.: Image registration methods: a survey. *Image and Vision Computing 21* (2003), 977–1000.