Digitalisation of Warped Documents Supported by 3D Surface Reconstruction

Erik Lilienblum, Bernd Michaelis

Institute for Electronics, Signal Processing and Communications
Otto-von-Guericke University Magdeburg
P.O.Box 4120, 39106 Magdeburg, Germany
erik.lilienblum@e-technik.uni-magdeburg.de
http://iesk.et.uni-magdeburg.de/TI/TI.html

Abstract. The high quality digitalisation of warped documents is still a big problem for most scanner technologies. The presented work is a contribution to develop a new technique handling this problem for so-called top view scanners. Basic principle of the proposed method is a special kind of light section to calculate 3D surface measurements. It works with the comparatively broad stripe lighting of the scanner hardware by using one additional matrix camera. We can reconstruct the 3D surface of the document by simple capturing an image sequence of the stripe lighting during the scanning process. Based on a surface model we transform the warped document in a plane. Result is the two-dimensional output being a nearly distortion-free digital copy of the original document.

1 Introduction

Digital archiving of hard-back literature becomes increasingly an essential part of the work of libraries and museums. Although for this purpose the modern computer technology accomplishes already major premises, the status quo of the scanner technology is not satisfying. In particular it is hardly possible to get distortion-free copies from thick books without damaging them. This represents currently for digitalisation of valuable historical books a large problem. In addition, a full automated character recognition (e.g. OCR) in the area of book crease is often impossible.

In the last years on the market a special kind of book scanner, which is called top view scanner, became generally accepted. To use the books in their natural way those scanners capture a copy of the page from above and from a certain distance. Through the use of a line camera with an appropriate stripe lighting it is possible to receive an evenly sharp and well illuminated two-dimensional image. However, we get inevitably a distorted copy in consequence of the projective geometry of the scanner and the warped surface of the page. An example can be seen in figure 1.a.

The basic idea of our proposed method is the extension of the scanner construction by an additional matrix camera. The matrix camera captures during

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

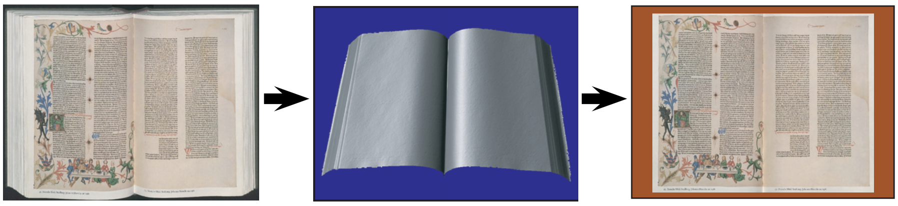

Fig. 1. a) original image captured by a book scanner, b) 3D reconstruction of the page, c) distortion-free copy

the scanning process an image sequence of the moving stripe lighting of the scanner. From this image sequence we calculate a 3D surface reconstruction of the warped page, which is represented in figure 1.b. By combining the surface reconstruction and the original scanner image we can calculate a nearly distortion-free copy of the page as it is shown in figure 1.c.

2 Related Work

Removing the projective distortion in a two-dimensional image of a three-dimensional warped surface is a very common problem. A possibility without the use of additional information or instruments offers the "shape from shading" principle. In [1, 2] the shape of the surface is drawn directly from the distribution of brightness. Because of strict conditions the exactness and reliability of this approach are not particularly high.

Another approach without an extension of the scanner hardware is to recognize the lines of text in a page [3, 4]. Supposed that the lines must be straight, the original scanner image can be corrected accordingly. But this method works only with text pages. Pages with figures or insular text blocks can not be dewarped.

A quite similar technique is used in systems recognizing the margins. One approach is given by Rebmann et al. in [5]. Another approach and an example of its application can be found in the software tool of the company Zeutschel for the top view scanner OS10000, which is based on the patent [6]. In most cases the software solution works well, but there are problems, if the book is not aligned well or if there are notes at the edges. These problems were a main motivation for us to develop an alternative solution.

Book scanner systems with additional hardware support are based mostly on the construction of a vision system and the calculation of a surface reconstruction. There are constructions with one matrix camera and one projector in [7], two matrix cameras in [8] or laser triangulation in [9]. These techniques are often costly concerning the additional hardware and provide low resolutions in the copy of documents due to the exclusive use of matrix cameras. Additionally, there are a lot of other techniques to get a surface reconstruction, which we do not describe here.

If we have got the 3D surface reconstruction of a book page, then we have to solve the dewarping problem to get distortion-free 2D copies. Here the works

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

from Brown [10] and Liang [11] should be named. The results of these papers can be used alternatively to dewarp pages after reconstructing the surface by our proposed method.

3 Concept and Technical Realisation

A concrete technical realization of the method we implemented on the basis of the top view scanner OS10000 in cooperation with the companies Zeutschel and Chromasens, which were partners of a research project. The concept realized thereby is based on three presumptions:

1. By using a line camera a high resolution image is captured.
2. The stripe lighting of the scanner is moved in a repeatable speed and synchronously with the line camera.
3. The moved stripe lighting is captured by a fixed matrix camera in an image sequence with a constant frame rate.

Because presumptions 1 and 2 are already achieved by the OS10000 book scanner, we could realize a concrete setup by assembling an additional fixed matrix camera, as shown in figure 2.

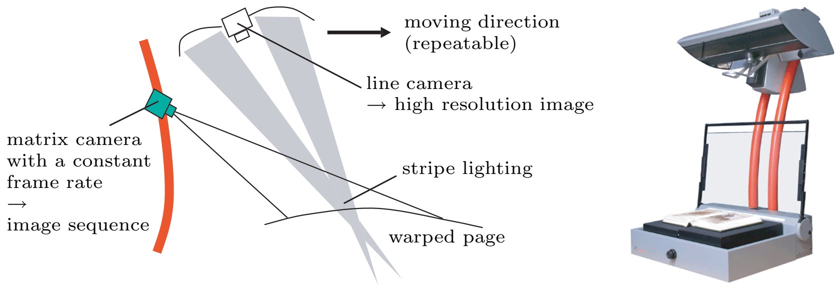

Fig. 2. Left: schematic representation of our setup, right: illustration of an original OS10000 top view scanner

The calculation of our surface reconstruction is based on a further development of classical light sectioning [12]. According to the well known principle the position of the light stripe in an image of the matrix camera depends on the surface shape of the page we are scanning. Thereby from the captured image sequence we can infer the form of the whole surface.

To calculate 3D coordinates we have to calibrate the setup whereas we define an unique device coordinate system. The calibration includes both the matrix camera and the line camera. That provide us with a relation between 3D surface points and 2D image points of the high resolution image we captured by the line camera. If we transform our warped surface into a plane, than this relation

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

remains preserved. Finally, this relation is the basis for outputting the plane as a high resolution nearly distortion-free copy of the page.

The preliminary design of the setup and the basic concept permit a modular implementation in the application. Line camera and matrix camera work synchronized, but independently. Thus the actual application can be divided into the tasks 2D-scan (image of the line camera), 3D-scan (surface reconstruction) and dewarping (output). This partitioning has positive effects on the algorithmic and software-technical implementation. The proposed method can thereby be integrated into practice as a pure extension of an existing scanner architecture in form of an auxiliary module. In 2006 we announced on the new method a patent [13].

4 Surface Reconstruction

In classical light sectioning one determines in an image of the matrix camera the local pixel position of a projected light stripe. Usually, this only works if the light stripe is very small like a projected laser line. Because the stripe lighting of a top view scanner generates a broad light stripe we have to apply an alternative technique.

In our version of light sectioning for a pixel of the matrix camera we determine from the image sequence the point in time, on which the light stripe was passing the pixel. Independent from the width of the stripe lighting we can calculate the point in time with high accuracy. However, by triangulation of the optic beam of the pixel and the plane of the stripe lighting we get in each case the position of a surface point. A schematic comparison of time-based and position-based analysis is represented in figure 3.

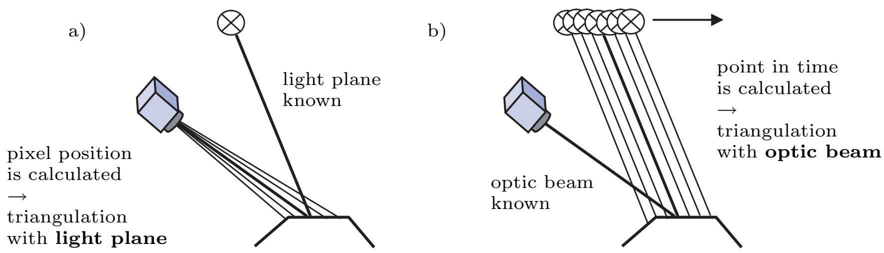

Fig. 3. Position-based (a) and time-based (b) analysis of a light beam

For the time-based analysis we need on each surface point seen by a pixel of the matrix camera a measurable change of brightness over several frames of the matrix camera. Therefore, broadness and speed of the stripe lighting have to be tuned with the frame rate of the matrix camera. The point in time calculated from the brightness distribution on a pixel we call time value.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

If we suppose an ideal light stripe then our time value corresponds to a light plane. By sectioning the optic beam of the pixel with the plane of the stripe lighting we get a surface point. However, this works only in theory. A stripe lighting of a top view scanner is normally not designed for light sectioning and surface reconstruction. There are usually inhomogeneities in form and width of the light stripe. Using the scanner head of the OS10000 is especially difficult, because it does not create a light plane but a light cone. So there is no way to use the classical idea of light sectioning for our propose.

But the lighting behaviour of the scanner stays in each scanning process the same. This means, that the time value is a repeatable measurement for each surface point being seen by a pixel. Because of this repeatability the calculation of time values can be the basis for reconstructing surface points.

Our time value calculation method is based on a template matching. Let $g_{i,j}(t)$ be the measured brightness function on pixel (i,j) over the frames $t \in \mathbb{N}$ of the matrix camera. Then the time value $\tau(i,j)$ is calculated by matching $g_{i,j}(t)$ with a static discrete template function $f(k)$ with $k \in \mathbb{N}$ and $1 \leq k \leq m$. Through interpolation, in the simplest case linear, we get from the discrete brightness function a continuous brightness function with $t \in \mathbb{R}$. The time value we are looking for arises with $\tau(i,j) = c \in \mathbb{R}$ from the solution of the following optimization problem

$$\sum_{k=1}^{m} \bigl(a \cdot f(k) + b - g_{i,j}(c+k)\bigr)^2 \rightarrow \min, \tag{1}$$

where $a, b \in \mathbb{R}$ are arbitrary factors to the multiplicative and additive adaptation. The computed time value corresponds to an optimal shift of the template function along the time axis. Figure 4 shows an schematic example for the calculation of a time value, whereby the template function and the brightness function are represented.

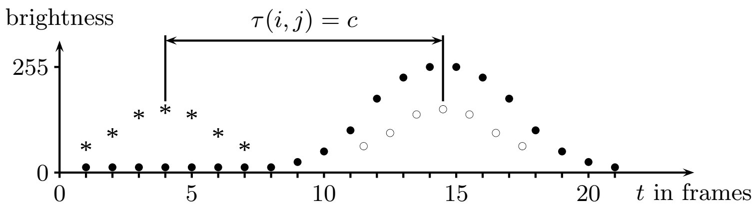

Fig. 4. Calculation of a time value of a pixel: $\bullet$ measured brightness function $g_{i,j}(t)$, $*$ template function $f(k)$, $\circ$ shifted template function $f(t-c)$

Now we have got our time values, that are reproducible measurements. Because we cannot use a light section due to the above mentioned inhomogeneities in the stripe lighting we apply a node based calibration instead of the usual parametric calibration. Basis of this calibration method is a simple lookup table, that yields for each pixel (i,j) of the matrix camera an ordered set of n

nodes (t_l, z_l), which consist of a time value t_l and its corresponding z-coordinate z_l. An illustration of the lookup table data structure is presented in figure 5.

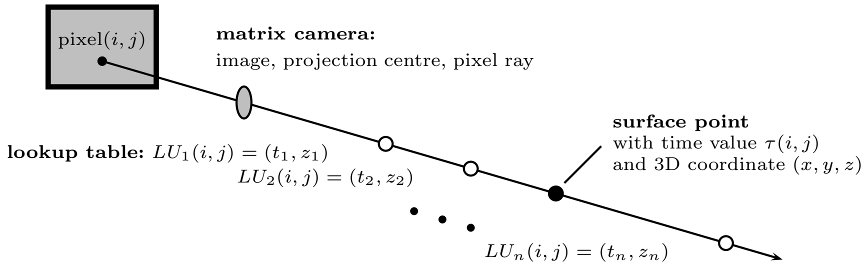

Fig. 5. Data structure of the lookup table and its geometrical meaning

Based on the lookup table the surface reconstruction is quite simple if we use a linear interpolation between the nodes. First we calculate for a pixel (i,j) the time value $\tau(i,j)$. Then we search in the lookup table of that pixel $LU(i,j) = \{(t_1, z_1), (t_2, z_2), \cdots, (t_n, z_n)\}$ an entry l, so that it applies

$$t_l \leq \tau(i,j) < t_{l+1}. \tag{2}$$

The z-coordinate of the surface point being seen by the pixel (i,j) we get with the equation

$$z(i,j) = \frac{\big(t_{l+1} - \tau(i,j)\big)z_l + \big(\tau(i,j) - t_l\big)z_{l+1}}{t_{l+1} - t_l} \quad , \tag{3}$$

which is a linear interpolation between the upper and the lower time value of equation 2. The complete 3D coordinate of the surface point we obtain by the spatial position of the pixel ray, that we know from the calibration of the matrix camera.

5 Dewarping

The calculation of the surface reconstruction is exclusively based on measurements captured by the matrix camera. To execute the dewarping of the page we now need the relation between image coordinates of the line camera and spatial coordinates. As already mentioned we get this relation by calibrating both cameras in one device coordinate system. The calibration is mostly based on standard algorithms and is a separate problem. We do not describe it here.

The crucial step on the way to a distortion-free copy of the scanned page is the transformation of the warped surface into a flat surface. There are two conditions which we must consider:

1. The distance between two points on the warped surface has to be equal to the distance between the same points after the transformation.

2. The relation between image coordinates of the line camera and spatial coordinates has to be invariant relative to the transformation.

There is a number of different solution types for this complex mathematical problem (e.g. in [10, 11]). For our research project we so far only implemented one simple method in order to be able to demonstrate some practice-relevant results.

We assume a simplified surface model for pages, which permits warping only in one direction. For common books this assumption is not far off from reality at all. After calculating the surface reconstruction, we detect the book crease in the xy-plane of our device coordinate system. Then, based on the 3D surface reconstruction we calculate the arc length from each surface point to the book crease. This arc length we take as a distance in the virtual plane between the transformed surface point and the book crease. Thus the distances between points on the virtual plane correspond to the actually existing distances. Based on the relation to the image of the line camera we output the virtual plane as an 2D image.

6 Results

In the analysis with different books and other materials it turns out that the time values, which we get by our proposed method, supply a stable and reproducible basis for the computation of a surface reconstruction. For illustration, the calculated time values of a test scan are plotted in figure 6. The upper figure presents the light stripe in one image of the image sequence captured by the matrix camera at the time t^*. The lower figure presents the brightness distribution on a pixel row j^* and the graph of the time value $\tau(i, j^*)$ which we computed by our template matching method.

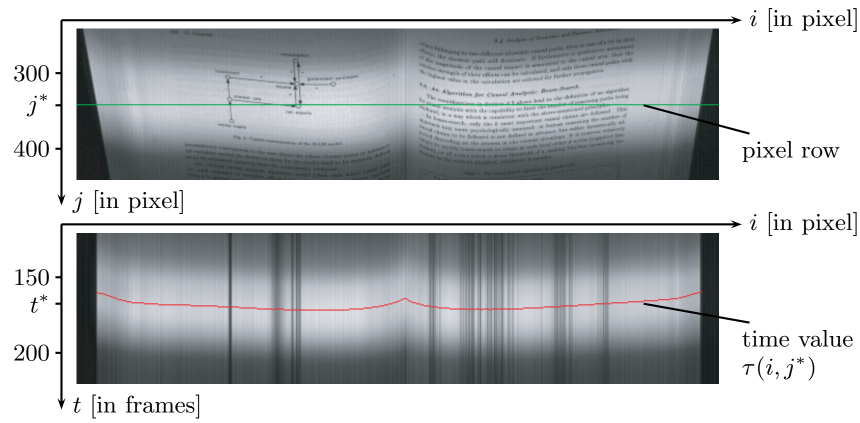

Fig. 6. Illustration of time values in a test scan

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

The shape of time values corresponds directly to the shape of the surface. Using our lookup table and the linear interpolation between their nodes we get quite good 3D coordinates. For further use we transform the xy-coordinates into an equidistant grid. The result is shown in figure 7. A more detailed representation of the measurements can be found in figure 8, which shows a section through the book.

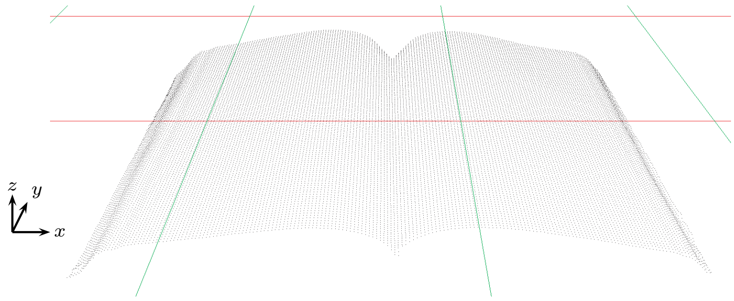

Fig. 7. Calculated surface points of the test page

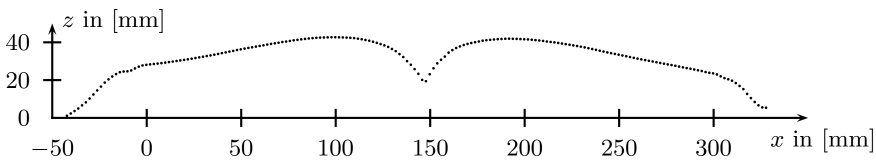

Fig. 8. 3D measurements on a section through the test book at $y = 140$mm

Evaluating the measurements partially unveils some large deviations up to ± 1 mm between measured values and actual values. These measurement errors are not due to incorrectly computed time values or errors in the calibration of the system. They are probably caused by inhomogeneous paper properties and reflections. This appears to be a general problem of the 3D surface reconstruction in our proposed method. However, in future work we still have to investigate this point more exactly .

But despite the 3D measurement errors we can present good results concerning the dewarping. The difference between our result and the original image taken by the line camera of the scanner is quite remarkable. Figure 9 shows an example of the dewarping with our method.

In figure 10 we extract an image detail from figure 9. In particular the metrically accurate shape of the letters in their original size should be emphasized. This generally leads to an improved automatic character recognition in the proximity of the book crease. However, the visual quality, which is important for

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

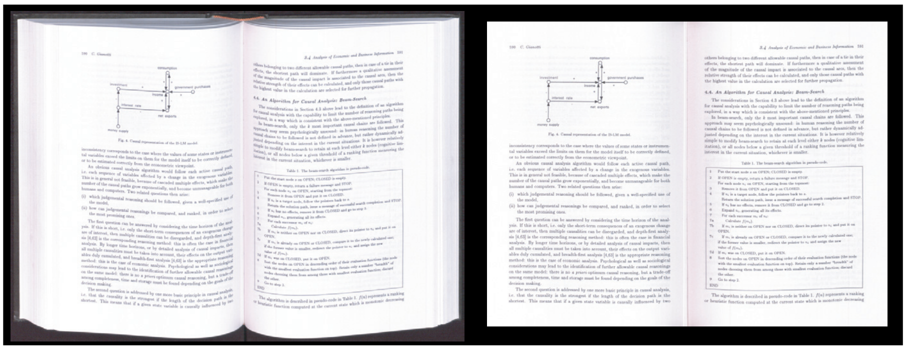

Fig. 9. Example of the dewarping by our method

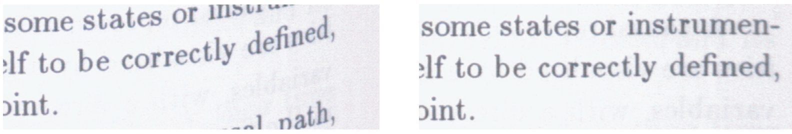

Fig. 10. Image detail of dewarping

digitalisation of art-historical books, particularly depends on the uniformity. Therefore, we have to accomplish a smoothing of the 3D measurements, a correction of shading and a compensation of differences concerning the sharpness. But this should be the concern of further work.

7 Conclusion and Further Work

We have presented a new method to upgrade top view scanners for the digitalisation of warped documents. By setting up only one additional matrix camera we are able to compute the surface shape of a page through analysing the stripe lighting during the scanning process. The surface reconstruction is the basis for correcting distortions occurring naturally due to the warped surface of the page. The result of the scanning improved by our proposed method is an output, which yields a nearly distortion-free copy of the scanned document.

However, we can still think of further improvements. To reduce the errors due to the inhomogeneous paper properties and reflections we will develop some extensions. One idea is to use a second matrix camera. This increases the cost of the method, but it brings more stability of the surface reconstruction, because we can use pure stereo photogrammetry instead of light sectioning.

Also we will extent our dewarping algorithm to make it more independent from the shape of the page. The aim is the digitalisation of arbitrary formed documents. This is especially important for old historical books with unknown rigid deformations.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

Acknowledgment

This work was supported by AiF/Germany grants (FKZ: KF0056101SS4). We thank the companies Chromasens and Zeutschel for providing us with the book scanner OS10000 for research purposes. And special thanks to Roger Klein.

References

1. Wada, T., Ukida, H., Matsuyama, T.: Shape from shading with interreflections under a proximal light source: Distorsion-free copying of an unfolded book. International Journal of Computer Vision **24**(2) (1997) 125–135
2. Zhang, Z., Tan, C.L., Fan, L.: Estimation of 3d shape of warped document surface for image restoration. In: Proc. of 17th International Conference on Pattern Recognition. Volume 1. (2004) 486–489
3. Wu, C., Agam, G.: Document image de-warping for text/graphics recognition. In: Proc. of Joint IAPR and SPR. (2002) 348–357
4. Zhang, Z., Tan, C.L.: Correcting document image warping based on regression of curved text lines. In: Proc. of 7th International Conference on Document Analysis and Recognition. Volume 1. (2003) 589–595
5. Rebmann, R., Michaelis, B., Krell, G., Seiffert, U., Püschel, F.: Improving image processing systems by artificial neural networks. In: Reading and Learning. Volume 2956 of LNCS. Springer (2004) 37–64
6. Frei, B.: Method and device for the correction of a scanned image. Patent US020050053304A1 (2002)
7. Donescu, A., Bouju, A., Quillet, V.: Former books digital processing: image warping. In: Proc. Workshop on Document Image Analysis. (1997) 5–9
8. Yamashita, A., A.Kawarago, Kaneko, T., Miura, K.: Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system. In: Proc. of 17th International Conference on Pattern Recognition. Volume 1. (2004) 486–489
9. Chu, K.B., Zhang, L., Zhang, Y., Tan, C.L.: A fast and stable approach for restoration of warped document images. In: Proceedings of the Eighth International Conference on Document Analysis and Recognition. (2005) 384–388
10. Brown, M., Seales, W.: Image restoration of arbitrarily warped documents. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(10) (2004) 1295–1306
11. Liang, J., DeMenthon, D., Doermann, D.: Flattening curved documents in images. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2005) 338–345
12. Shirai, Y.: Recognition of polyhedrons with a range finder. Pattern Recognition **4** (1972) 243–250
13. Lilienblum, E., Michaelis, B., Schnitzlein, M.: Verfahren zum Entzerren einer mittels einer Scanneinrichtung eingescannten zweidimensionalen Abbildung einer nicht ebenflächigen Vorlage. Pat.Reg.: DE 10 2006 032 533.8-31 (2006)