World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:3, No:11, 2009

A Stereo Vision System for Top View Book Scanners

Erik Lilienblum and Robert Niese and Bernd Michaelis

Abstract—This paper proposes a novel stereo vision technique for top view book scanners which provide us with dense 3d point clouds of page surfaces. This is a precondition to dewarp bound volumes independent of 2d information on the page. Our method is based on algorithms, which normally require the projection of pattern sequences with structured light. We use image sequences of the moving stripe lighting of the top view scanner instead of an additional light projection. Thus the stereo vision setup is simplified without losing measurement accuracy. Furthermore we improve a surface model dewarping method through introducing a difference vector based on real measurements. Although our proposed method is hardly expensive neither in calculation time nor in hardware requirements we present good dewarping results even for difficult examples.

Keywords—stereo vision, 3d surface reconstruction, dewarping documents, book scanner

I. Introduction

The issue of getting page copies from bound volumes has been existing since document copies were available at all. There are serious problems dealing with flat bed scanners. Commonly the user has to press the book on a glass plate, which may lead to damage of the book spine. And in general, the copy results are not convincing. For this reason a special kind of book scanner, which is called top view scanner, became generally accepted in the last decade. 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.

The main problem using top view book scanners is the dewarping problem, because we inevitably get a distorted copy in consequence of the projective geometry of the scanner and the warped surface of the page. There are many different kinds of dewarping techniques. The most powerful but also most expensive technique uses stereo vision methods to reconstruct the 3d surface of the page [1], [2], [3]. The surface reconstruction is the best basis for dewarping an 2d image of the page. Normally it works for arbitrary documents [4]. Other approaches work without a stereo vision setup. Well-known examples base upon shape from shading [5] or fitting text lines [6]. Further methods and substantial references are given by Liang et al. in [7].

In this paper we propose a stereo vision system with two matrix cameras that work together with a line camera. The line camera is part of a top view book scanner and provides a high resolution good quality 2d image. The matrix cameras may be seen as an additional option of the book scanner for a 3d-supported image processing. The difference to other stereo vision approaches specified above is the active use of the book scanner lighting for 3d data acquisition. It serves as a kind of structured light to solve the correspondence problem between both matrix cameras. Thus we can disclaim both a separate projection of structured light and a separate scanning process to get the surface reconstruction.

In earlier works we used the stripe lighting of the scanner to apply a special kind of light sectioning [8]. This approach works with only one matrix camera and supplies rather good results. Since 2007 it is used in a commercial scanner system. The new approach in this paper supplies much better 3d results and can serve as the better basis for the complete dewarping system.

II. Stereo Vision Setup

The theory of stereo vision goes back on spatial seeing. In the technical conversion we are using at least two cameras to reconstruct 3d surfaces. Due to the fast development of computer technique and the high relevance for industrial applications a multitude of different methods were investigated in the last years. The method which is used for a specific application depends on the requirements concerning the measurement accuracy and the resolution in time and space. An overview is given in [9].

Stereo vision systems need distinctive texture on the surface to compute a 3d reconstruction. Usually the self-texture on book pages is for this well suitable, but not in each case. On areas without text there is no usable image content. Without self-texture information a dense and high accuracy 3d reconstruction of surfaces requires methods with projection of structured light [10]. Moreover, if non-moving objects are measured, the accuracy can be substantially increased by the use of pattern sequences [11]. But the technical facilities make applications within the field of the book scanners difficult. First, for using a structured light projection in daylight the projector has to be very powerful. And second, the structured light projection cannot be accomplished, while the book scanner is in process. That means that two processing steps are necessary.

The basic idea in this paper is to use the stripe lighting of the book scanner like a projected texture information. Thus, we don't need an additional projector for our setup. The technique works in daylight trouble-free, because the stripe lighting is very bright. And, the acquisition of stereo images takes place during the scanning process. These are three important physical characteristics of our approach.

Erik Lilienblum, Robert Niese and Bernd Michaelis are with Institute for Electronics, Signal Processing and Communications, Otto-von-Guericke University Magdeburg, P.O.Box 4120, 39106 Magdeburg, Germany, e-mail: erik.lilienblum@ovgu.de, web: http://iesk.et.uni-magdeburg.de/TI/TI.html

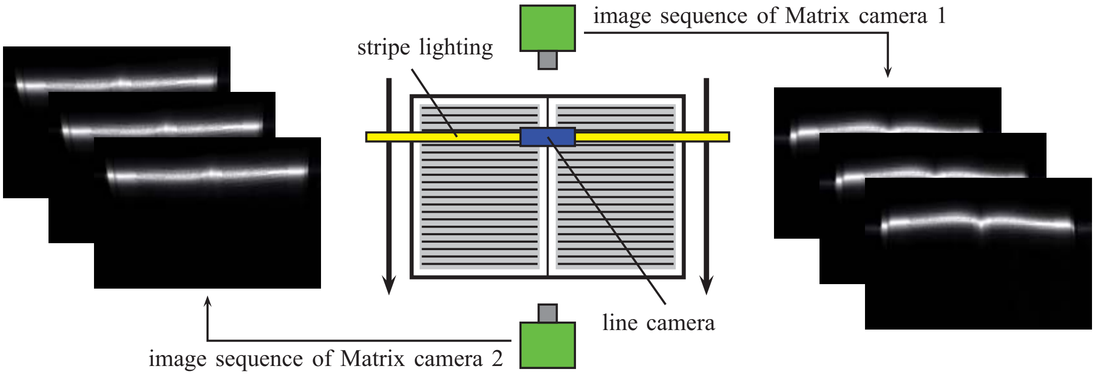


Fig. 1. Schematic stereo vision setup from above

The setup consists of two cameras, which are positioned orthogonal to the line camera and the stripe lighting above the scanning area. The matrix cameras capture during the scanning process a triggered image sequence of the moving stripe lighting. A schematic illustration of this seen from above and examples of original images are given in figure 1. The frame rate of the cameras has to be matched to width and speed of the stripe lighting. It means that a signal of the stripe lighting should be seen on each pixel in at least five images. The stereo vision system is calibrated by state of the art methods. They are well described by Luhmann in [12].

III. Surface Measurement

For optical 3D measurement of surfaces many different procedures exist. In this paper we limit our treatment to photogrammetry based on a pinhole camera model in consideration of optical aberration and geometric triangulation. Fundamentals can be likewise found in [12]. For the solution of the correspondence problem we developed a special method with so called time values. In the last subsection we propose a method to handle pages with areas of different brightness, which normally lead to overexposure or underexposure.

A. Triangulation Method

In stereo vision triangulation means the computation of a 3d point from two different view points. The view points are determined by the location of the projection center of the cameras and the corresponding pixels in the stereo images. We call two pixels p_1 from camera 1 and p_2 from camera 2 corresponding, if they "see" the same point of surface. To decide wether two pixels are corresponding or not we have to formulate a correspondence criterion, normally a correlation function. As a stereo vision system, which uses projection of pattern sequences, our correspondence criterion refers to the gray value sequences $G(p_1)$ and $G(p_2)$ of the two regarded pixel. In our work we use the SSD (sum of squared differences) approach to formulate the correlation function

$$SSD(p_1, p_2) = \sum_{i=1}^{n} \left[G_i(p_1) - G_i(p_2)\right]^2 \tag{1}$$

whereby G_i is gray value number i of sequence G. The correlation function is also defined for so called subpixel $p = (u, v)$ with $(u, v) \in \mathbb{R}^2$. In case of subpixel the gray value sequence $G(p)$ is calculated by bilinear interpolation. Other correlation functions than the SSD we do not investigate here.

To find the corresponding points there are different searching strategies. We use the epipolar geometry, which reduces the search problem to one dimension. For a pixel p in one camera image we compute in the other camera image a line segment corresponding to the epipolar line in a certain depth range. For each pixel on the line segment $l(p, l)$, whereby l is a free parameter with $l_{\min} \leq l \leq l_{\max}$, we get a 3d coordinate $(x, y, z) = r(p, l)$ by triangulation. By solving the optimization problem

$$SSD\left(p, l(p, t)\right) \rightarrow \min \tag{2}$$

we get a surface point $r(p, l)$. Doing that for all pixels in both images we provide a dense surface reconstruction of the page. We call this an image based method.

B. Calculation of Time Values

There are many different possibilities to solve the optimization problem from equation 2. To get an efficient calculation method we chose an approach similar to phase shifting [13]. In phase shifting we calculate from the gray value sequence of a pixel a phase value. To solve the optimization problem the phase values of two pixels have to be just equal. Based on the moving stripe lighting, which can be seen as a projected pattern, we can define the calculation of time values. A time value depends on the brightness distribution in the gray value sequence of two pixels. The method to calculate time values is based on a matching of the two gray value sequences.

Let $g(t)$ and $f(t)$ be the gray value sequences over the frames $t \in \mathbb{N}$ with $t_0 \leq t \leq t_{\max}$. Then the time value $\tau(g, f)$ is calculated by matching $g(t)$ with $f(k)$ for $k \in \mathbb{N}$ and $t_0 \leq k_0 \leq k \leq k_n \leq t_{\max}$ with $k_n = k_0 + n$. If we modify by interpolation the discrete gray value sequences $g(t)$ into a continuous function with $t \in \mathbb{R}$, then the time value we are looking for arises with $\tau(g, f) = c \in \mathbb{R}$ from the solution of the following optimization problem

$$\sum_{k=k_0}^{k_n} \left(a \cdot f(k) + b - g(c + k)\right)^2 \rightarrow \min, \tag{3}$$

with the condition $k_0 - t_0 \leq c \leq t_{\max} - k_0 - n$ and where $a, b \in \mathbb{R}$ are arbitrary factors to the multiplicative and additive adaptation. The computation corresponds to an optimal shift of the gray value sequence $g(t)$ along the time axis (frames) by the time value c, getting a minimal SSD according to equation 1. Figure 2 shows an schematic example for the calculation of a time value, whereby the two gray value sequences are represented.

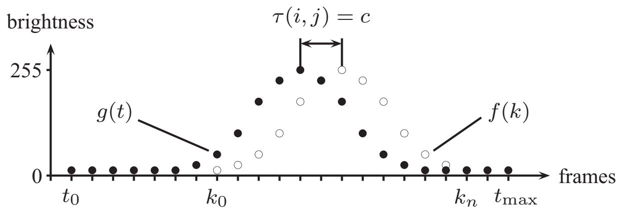


Fig. 2. Calculation of a time value of a pixel by matching $g(t)$ with $f(k)$

The optimization problem can very efficiently be solved by using an over-determined system of linear equations. In the following we derive the solution. As starting point we define an approximate value $\tilde{c} \in \mathbb{N}$ with $\tilde{c} \leq c < \tilde{c} + 1$. However, the approximate value can be computed for example over the difference of the maximum positions in both gray value sequences. Furthermore we define for the discrete functions g and f the notation $g_k = g(\tilde{c} + k)$ and $f_k = f(k)$. The continuous function g we compute through a linear interpolation, which we define with

$$g(c + k) = g_k + (c - \tilde{c})(g_{k+1} - g_k). \quad (4)$$

If we apply this equation in the optimization problem above, then we get in equation 3 summands of the form

$$\big(af_k + b - g_k - c(g_{k+1} - g_k) + \tilde{c}(g_{k+1} - g_k)\big)^2. \quad (5)$$

Let's assume the factors a, b, c are known and the error in each summand is zero, then the equation

$$af_k + b - g_k - c(g_{k+1} - g_k) + \tilde{c}(g_{k+1} - g_k) = 0 \quad (6)$$

is valid. This equation can be transformed into

$$af_k + b - c(g_{k+1} - g_k) = g_k - \tilde{c}(g_{k+1} - g_k) \quad (7)$$

which leads for $k = 1, \ldots, m$ to an system of linear equations with

$$\begin{bmatrix} f_1 & 1 & g_2 - g_1 \\ f_2 & 1 & g_3 - g_2 \\ \vdots & \vdots & \vdots \\ f_m & 1 & g_{m+1} - g_m \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} g_1 - \tilde{c}(g_2 - g_1) \\ g_2 - \tilde{c}(g_3 - g_2) \\ \vdots \\ g_m - \tilde{c}(g_{m+1} - g_m) \end{bmatrix}. \quad (8)$$

Because the gray value sequences are real measurements this system of linear equations normally is over-determined. An optimal approximation of the factors a, b, c according to the method of least squares leads directly to the solution of the optimization problem in equation 3. It can be realized with

$$\mathbf{x} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} \mathbf{b} \quad (9)$$

whereby $A\mathbf{x} = \mathbf{b}$ is the algebraic notation of equation 8. In case of three unknown variables the computation of equation 9 can be implemented by a single step algorithm. It leads to a well efficiency computation of time values, which are our basis to get 3d surface points by triangulation according subsection III-A.

C. Overexposed Image Sequences

For well exposed image sequences the proposed method yields dense and quite accurate 3d surface reconstructions. But there are serious problems if there is a high dynamic of brightness on a page. On dark parts we get underexposed light signals and on bright parts we get overexposed light signals. While underexposed image sequences cannot be improved (because in worst case there is no signal at all) our method allows us to handle overexposed image sequences easily.

Assuming the function g is overexposed in the interval $[t_1, t_2]$ with $t_0 \leq t_1$ and $t_2 \leq t_{\max}$. Then the function has no maximum peak, but a maximum plateau between t_1 and t_2. Even though g is overexposed the function f may be normally exposed, because the two matrix cameras have different geometric positions. In addition, the reverse case may happen. However, we have to assume that function f is overexposed too. The interval here is $[k_1, k_2]$ with $k_0 \leq k_1$ and $k_2 \leq k_n$. Both assumptions mean that the optimization problem from equation 3 to get the time value c does not work correctly, if the size of the intervals is different with $t_2 - t_1 \neq k_2 - k_1$.

To take overexposure into account we have to skip all rows with overexposed values from the system of linear equations in equation 8. Thus the system of linear equations is getting the form

$$\begin{bmatrix} f_1 & 1 & g_2 - g_1 \\ \vdots & \vdots & \vdots \\ f_i & 1 & g_{i+1} - g_i \\ f_j & 1 & g_{j+1} - g_j \\ \vdots & \vdots & \vdots \\ f_m & 1 & g_{m+1} - g_m \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} g_1 - \tilde{c}(g_2 - g_1) \\ \vdots \\ g_i - \tilde{c}(g_{i+1} - g_i) \\ g_j - \tilde{c}(g_{j+1} - g_j) \\ \vdots \\ g_m - \tilde{c}(g_{m+1} - g_m) \end{bmatrix}, \quad (10)$$

whereby the indices i, j are defined on the overexposed intervals with

$$\begin{aligned} i &= \min\{k_1, t_1 - 1\}, \\ j &= \max\{k_2, t_2\}. \end{aligned} \quad (11)$$

It should be noted that equation 10 supplies good results only if enough lines with normal exposed gray values are remaining. The intervals can be calculated very quick by analyzing the gray value sequences before calculating a time value. The approximate value $\tilde{c}$ may be calculated by center of mass algorithms.

IV. The Dewarping Method

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. The success of an approach depends more on the quality of the 3d data than on the quality of the dewarping algorithm. For exact measurements the approach from Brown [4] should provide a good dewarping result even for rigid surface deformations. But computational costs of that approach are very high. For this reason we prefer approaches which are based on a simplified mathematical model of a warped page [14].

In [15] we propose a dewarping method based on a simplified surface model. To compensate weak measurements this model is formulated very stringent. Thus, it works very fast and has been rather stable regarding measurement errors. However, it leads to dewarping errors in case of dents on a page. To take advantage of the powerful 3d measurement method proposed in this paper we have to improve our dewarping approach.

The copying error produced by a top view scanner is caused by perspective distortions. That is the main reason for the strong influence of dents. Let's assume we take an image with a telecentric view. Then the influence of dents is very small compared to the influence of the warped page. Thus, we accomplish our dewarping algorithm in two steps:

1) We apply our simplified surface model to dewarp the page in general.
2) We correct the deviations between the simplified surface model and the real 3d measurements by calculating a difference vector in the 2d scanner image.

Figure 3 shows the first and the second step in a horizontal page cut. It can be seen that the error caused by the dent has less influence in case of a telecentric view from above after dewarping the page. In the camera view the dent causes a shifting in the scanner image which we note by the difference vector Δ.

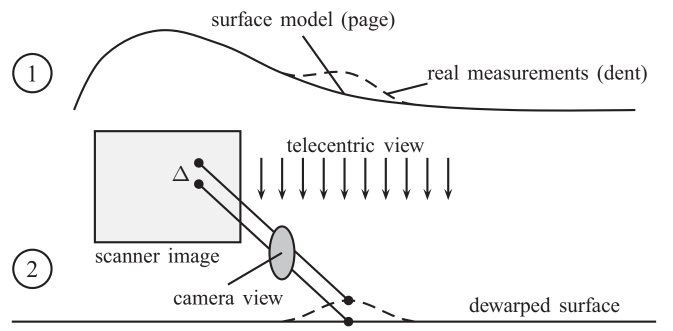


Fig. 3. Dewarping a page cut by a simplified surface model regarding a dent

Let G be the image function of the scanner mapping a 3d point into the 2d image and (x, y, z) a 3d point on the simplified surface model. Assumed that (x', y', z') is a 3d point of the real measurements, which is orthogonal on the model surface above (x, y, z). Then the difference vector in the 2d scanner image is

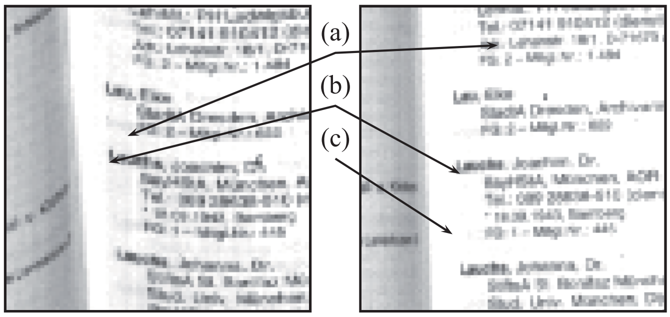
$$\Delta = G(x, y, z) - G(x', y', z'). \tag{12}$$

The difference vector is used to correct the output image.

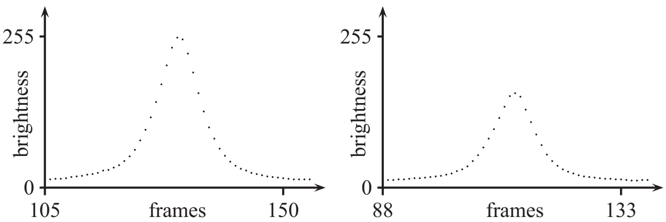
V. Results

The proposed methods were realized in a Zeutschel book scanner OS12000c in December 2008. Normally, this top view book scanner is already supported by one additional matrix camera, which yields a 3d surface reconstruction to dewarp the pages [8], [15]. For our setup we fix two CMOS matrix cameras (1k×1k) around the book scanner without affecting the original system. Thus we got the possibility to compare the results objectively.

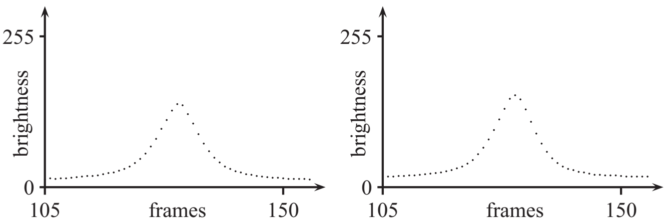
There are three stages to check our results. First, we have to show the correctness of the time value calculation. This is the basis for the second stage – the reconstruction of the 3d surface. Here we have to show that our triangulation method works well and that the time values are a capable feature to solve the correspondence problem. The third stage considers the dewarping method. The results here are still subjective, because we didn't gather enough measurements with the proper test objects (for instance with well known dents on the page). Suitable investigations are planned for the next publications.



(a) time value: -17.3567 frames



(b) time value: -0.0345 frames



(c) time value: 122.4751 frames

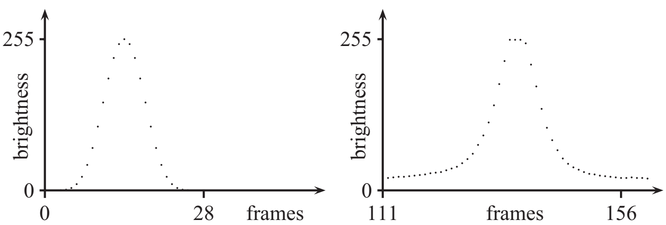
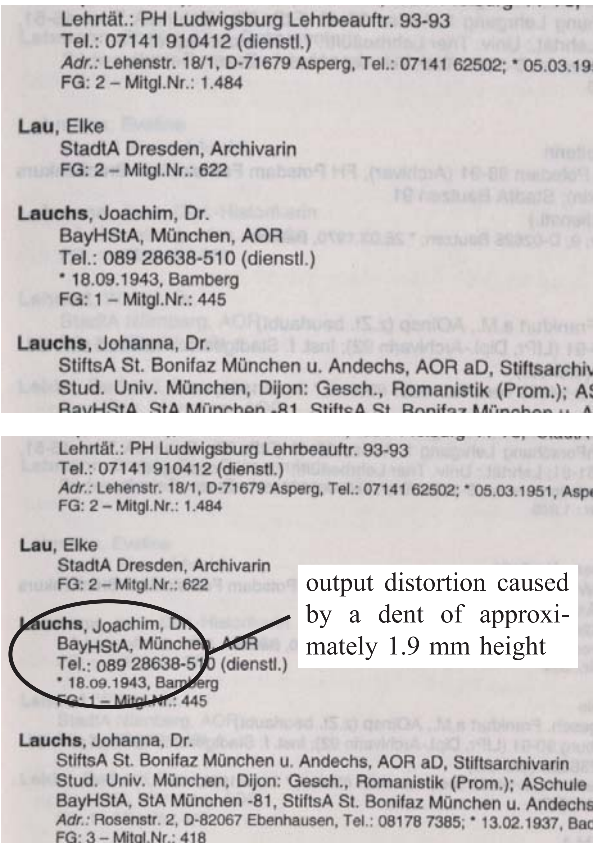


Fig. 4. Results of the time value calculation on different examples

Fig. 5. Surface reconstruction by the one camera and the two camera approach

Figure 4 shows results of the time value calculation for (a) two pixels mapping different 3d surface points, (b) two corresponding pixels and (c) a pixel with overexposure matched with a template function. The upper part of figure 4 shows a special image of the cameras composed from the maximum value on each pixel regarding the image sequences. All examples are calculated within an interval of 30 gray values of brightness function of the particular pixel. A mathematical proof of the calculation method of the optimization problem in equation 3 should follow from a proof of equation 9 and the derivation of the system of linear equations 8.


Lehrtät.: PH Ludwigsburg Lehrbeauftr. 93-93
Tel.: 07141 910412 (dienstl.)
Adr.: Lehenstr. 18/1, D-71679 Asperg, Tel.: 07141 62502; * 05.03.19
FG: 2 – Mitgl.Nr.: 1.484

Lau, Elke
StadtA Dresden, Archivarin
FG: 2 – Mitgl.Nr.: 622

Lauchs, Joachim, Dr.
BayHStA, München, AOR
Tel.: 089 28638-510 (dienstl.)
* 18.09.1943, Bamberg
FG: 1 – Mitgl.Nr.: 445

Lauchs, Johanna, Dr.
StiftsA St. Bonifaz München u. Andechs, AOR aD, Stiftsarchiv
Stud. Univ. München, Dijon: Gesch., Romanistik (Prom.); A

Lehrtät.: PH Ludwigsburg Lehrbeauftr. 93-93
Tel.: 07141 910412 (dienstl.)
Adr.: Lehenstr. 18/1, D-71679 Asperg, Tel.: 07141 62502; * 05.03.1951, Asp
FG: 2 – Mitgl.Nr.: 1.484

Lau, Elke
StadtA Dresden, Archivarin
FG: 2 – Mitgl.Nr.: 622

Lauchs, Joachim, Dr.
BayHStA, München, AOR
Tel.: 089 28638-510 (dienstl.)
* 18.09.1943, Bamberg
FG: 1 – Mitgl.Nr.: 445

Lauchs, Johanna, Dr.
StiftsA St. Bonifaz München u. Andechs, AOR aD, Stiftsarchivarin
Stud. Univ. München, Dijon: Gesch., Romanistik (Prom.); ASchule
BayHStA, StA München -81, StiftsA St. Bonifaz München u. Andechs
Adr.: Rosenstr. 2, D-82067 Ebenhausen, Tel.: 08178 7385; * 13.02.1937, Bac
FG: 3 – Mitgl.Nr.: 418

Fig. 6. Dewarping results on a dent with and without considering the difference vector

The main problems using a one camera solution are weak measurements caused by different texture on the page. Figure 5 shows the comparison between the method with one camera and the new stereo vision approach with two cameras. The significant deviations of the weak measurements could obviously be avoided.

In [15] we could show that a simplified surface model in principle yields good dewarping results even relative large measurement errors occur on the surface reconstruction. There are only dewarping errors on dents or in the area near the book crease. In both cases the new approach described in section 4 supplies much better results. Figure 6 shows a typical problematic image section with a dent. The lower example is dewarped only by a simplified surface model, whereas using the same measurements the upper example is additionally corrected through difference vectors.

VI. Conclusion and Further Work

In this paper we introduce a novel stereo vision technique for a low cost upgrade of modern top view scanners. We realized the stereo vision setup with only two matrix cameras above the scanner and without a projector or other additional equipment. As the computational basis for a 3d surface reconstruction of pages we present an efficient method to calculate time values from image sequences. We could show, that time values are well suitable to solve the correspondence problem, which leads to stable 3d measurements with high accuracy. Furthermore we could improve a model based dewarping method by considering the difference between the surface model and the real measurements. Thus the combination of accurate 3d measurements and the improved dewarping method leads to a significant advancement in the output result.

In further works we want to investigate the options to overexpose the image sequences on purpose to get better results on very dark pages. Thus the 3d feature can be used in nearly all kinds of documents. Also we want to improve the dewarping method toward arbitrary warped documents. The aim is to get a model free dewarping algorithm which is efficient enough to apply in a commercial used book scanner hardware.

ACKNOWLEDGMENT

This work was supported by AiF/Germany grants (FKZ: KF0056101SS4) and LSA/Germany grants (ZWB: 0904/00066). Special thanks to Roger Klein.

REFERENCES

[1] Donescu, A., Bouju, A., Quillet, V.: Former books digital processing: image warping. In: Proc. Workshop on Document Image Analysis. (1997) 5–9

[2] 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

[3] 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

[4] Brown, M., Seales, W.: Image restoration of arbitrarily warped documents. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 1295–1306

[5] Tan, C.L., Zhang, L., Zhang, Z., Xia, T.: Restoring warped document images through 3d shape modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 195–208

[6] Ulges, A., Lampert, C.H., Breuel, T.M.: Document image dewarping using robust estimation of curled text lines. Proc. of 8th International Conference on Document Analysis and Recognition **0** (2005) 1001–1005

[7] Liang, J., Doermann, D., Li, H.: Camera-based analysis of text and documents: a survey. International Journal on Document Analysis and Recognition **7** (2005) 84–104

[8] Lilienblum, E., Michaelis, B.: Digitalisation of warped documents supported by 3d-surface reconstruction. In: The 5th International Conference on Computer Vision Systems Conference Paper. (2007) DOI: 10.2390/biecoll-icvs2007-69.

[9] Blais, F.: A review of 20 years of range sensor development. Journal of Electronic Imaging **13** (2004) 231–243

[10] Valkenburg, R.J., McIvor, A.M.: Accurate 3d measurement using a structured light system. Image and Vision Computing **16** (1998) 99–110

[11] Albrecht, P., Michaelis, B.: Improvement of the spatial resolution of an optical 3-d measurement procedure. In: IEEE Transactions on Instrumentation and Measurement. Volume 47., Brisbane (1998) 158–162

[12] Luhmann, T., Robson, S., Kyle, S., Harley, I.: Close Range Photogrammetry. Whittles Publishing (2006)

[13] Lilienblum, E., Michaelis, B.: Optical 3d surface reconstruction by a multi-period phase shift method. Journal of Computers (JCP) **2** (2007) 73–83

[14] Liang, J., DeMenthon, D., Doermann, D.: Geometric rectification of camera-captured document images. IEEE Transactions on Pattern Analysis and Machine Intelligence (2007) (PrePrint).

[15] Lilienblum, E., Michaelis, B.: Book scanner dewarping with weak 3d measurements and a simplified surface model. In: Discrete Geometry for Computer Imagery - DGCI 2008. Volume 4992 of LNCS. (2008) 529–540