

A Method for Fast Composing of Images Captured from Deformed Documents by a Quadocular Scanner Setup

Marco Körner Herbert Süße Wolfgang Ortmann Joachim Denzler
 Chair for Computer Vision, Friedrich Schiller University of Jena
 Ernst-Abbe-Platz 2, 07743 Jena, Germany

{marco.koerner, herbert.suesse, wolfgang.ortmann, joachim.denzler}@uni-jena.de

Abstract

We present a fast and accurate method to compose images delivered by a quadocular document scanner into a global consistent image. The *Shift Detection by Restoration (SDR)* approach is used to overcome document deformations. Our method outperforms state of the art panoramic image approaches and is able to work within one second on standard hardware.

1 Introduction

The today's increasing mobility and the raised safety consciousness demands new solutions for identifying persons like air plane passengers quickly. For this reason novel document scanner devices are requested which are capable to process this high amount of data reliably as well as fast. The FP500 (*cf.* Fig. 1a, [1]) is a recent developed passport scanner for applications at security-relevant places like airports or public authorities. Due to price reduction, headway limitation and avoiding inappropriate distortions, this device is built up by four single low-cost cameras. A previous processing step provides precalibrated and undistorted images assuming an ideally planar document.

Since this assumption does not hold for real documents, we describe a method to compose these partial images of naturally bended documents to a global consistent overall image. This composing step is designed using an estimation of local optical flow via *Shift Detection by Restoration (SDR)* [3, 8]. The main attention of this approach lies on short computation times and high accuracies even under suboptimal conditions. Furthermore, the composing result should be free of non-linear deformations to achieve the machine-readability of the document. We show that the composing step is able to provide sufficient results within less than one second on standard hardware.

The remainder is structured as follows: after a short review of the related work in section 2 and a brief problem statement in section 3, our approach will be depicted in section 4. Some experimental results and a short summary and outlook will follow in sections 5 and 6, respectively.

2 Related Work

The problem of image stitching or panoramic image creation is concerned in a high amount of literature. There are two main approaches: on the one hand, searching for curves in overlapping image regions to find an optimal seam by minimizing the error between these images, as presented in [5, 6, 7]. On the other hand, the approaches published in [2, 9] try to minimize

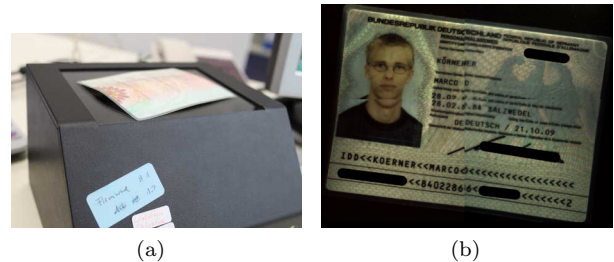


Figure 1. (a) A pathologically bended document lying on the Desko FP500 Passport Scanner [1] and (b) the result of composing according to the initial calibration.

the seam artefacts by smoothing the transitions at the stitching contour. Furthermore, Brown and Lowe published an approach for multi-row stitching in [4]. Nearly all of these panoramic image approaches premise an approximately identical optical center for all cameras recording the images to be stitched.

3 Problem Statement

In the given scenario, documents are captured by a scanner device containing four low-resolution CCD cameras. While the extrinsic calibration of these cameras assumes that the scanned document is exactly planar, this premise does not hold in real applications, *e.g.* due to surface deformations caused by storage and handling (*cf.* Fig. 1a). Even putting pressure onto the document and the glass plate bends its surface and provokes this effect. Hence the composing of the single images solely according to the calibration data fails in nearly all cases, as shown in Fig. 1b. In this section we present our approach for transforming and composing all images in order to obtain a global consistent image.

The composition of the single images using any of the known panorama image algorithms induces hard local nonlinearities and deformations, which constrict further processing, *e.g.* optical character recognition (OCR). Common panorama image approaches premise a mutual optical center for all images to be stitched, from where the images are projected onto spherical or cylindrical surfaces. Fig. 2 shows these effects, where subfigures (a) and (b) display the result of stitching both the uncalibrated and the calibrated images with Brown's and Lowe's approach [4] for panorama image stitching. Note that the stitching process fails when the loop has to be closed, because the errors due to the transformations cumulate. This effect complicates further information extraction algorithms and yield undesirable results.



Figure 2. Panorama images generated by AUTO-STITCH [4] based on (a) the uncalibrated and (b) the calibrated images.

4 Composing Approach

For composing the images, corresponding points have to be found in the image overlap areas. Based on these correspondences, transformations are estimated for all the images. To increase the accuracy, outliers are rejected and all transformations are reestimated. The final transformations are applied to the images to obtain a global consistent image. These steps will be explained in more detail in the following sections.

Local optical flow estimation The transformation estimation is based on the determination of optical flow in the overlapping image regions. The *Shift Detection by Restoration* (SDR) approach [3, 8], which can be seen as a regularized correlation similar to the phase correlation, has been shown to give good results for point correspondence identification, especially under presence of partly periodically structured background. In contrast to the usage of spatial features, this method determines the displacement of two equal- and square-sized images in a signal-based fashion by creating a peak image, where each peak corresponds to a shift hypothesis \mathbf{d} . In Fig. 3 some exemplary input patches are shown in context with their SDR results. As discriminating these peaks would state a strong heuristic problem, each peak hypothesis \mathbf{d} is evaluated by the normalized cross correlation coefficient

$$\rho(\mathbf{d}) = \frac{\sum_{(\mathbf{p}+\mathbf{d}) \in I_1} (I_0(\mathbf{p}) - \bar{I}_0)^2 \cdot (I_1(\mathbf{p} + \mathbf{d}) - \bar{I}_1)^2}{\sqrt{\sum_{\mathbf{p} \in I_0} (I_0(\mathbf{p}) - \bar{I}_0)^2 \cdot \sum_{(\mathbf{p}+\mathbf{d}) \in I_1} (I_1(\mathbf{p} + \mathbf{d}) - \bar{I}_1)^2}} \quad (1)$$

of the original image patch I_0 and the hypothetical shifted image patch I_1 , where \bar{I}_k denotes the mean gray value of image I_k . The shift hypothesis with the highest value of $\rho(\mathbf{d})$ is added to the set

$$\mathbf{c}_{\mathcal{X}} = \{(\mathbf{p}_1^{\mathcal{X}}, \mathbf{q}_1^{\mathcal{X}}), \dots, (\mathbf{p}_{n_{\mathcal{X}}}^{\mathcal{X}}, \mathbf{q}_{n_{\mathcal{X}}}^{\mathcal{X}})\}, \quad (2)$$

where the tuple $(\mathbf{p}_i^{\mathcal{X}}, \mathbf{q}_i^{\mathcal{X}})$ denotes a correspondence between the points $\mathbf{p}_i^{\mathcal{X}} \in I_0^{\mathcal{X}}$ and $\mathbf{q}_i^{\mathcal{X}} \in I_1^{\mathcal{X}}$ in the overlapping image area \mathcal{X} (cf. Fig. 4).

Transformation estimation Once the optical flow fields in the overlapping regions are determined, these information are used to estimate transformations for each image. Translations and special affine transformations are used for this application. As depicted

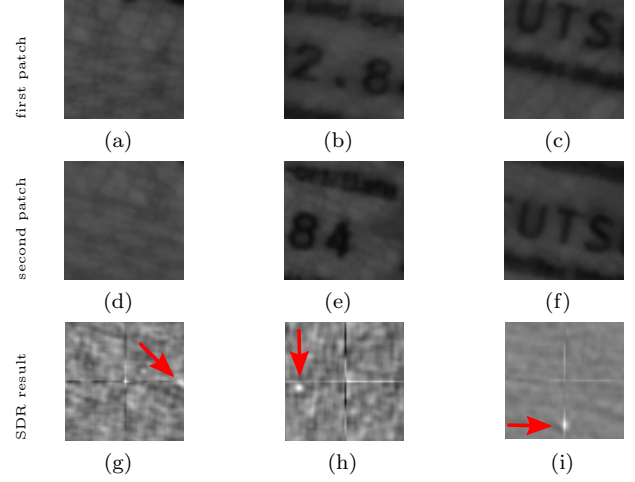


Figure 3. Image patches with (a, d) structured background, (b, e) foreground shifted horizontally and (c, f) vertically and (g-i) their SDR images, where white peaks are indicating the shift hypotheses. Note the suitability of this method for the partial periodically structured background of the images.

in Fig. 4, each transformation is assembled by opposing shear mappings and translations. Using these parameters, the transformation matrices

$$\mathbf{M} = \left\{ \overbrace{\begin{pmatrix} 1 & s_x^0 \\ s_y^0 & 1 \end{pmatrix}}^{\mathbf{M}_0}, \overbrace{\begin{pmatrix} 1 & -s_x^0 \\ s_y^1 & 1 \end{pmatrix}}^{\mathbf{M}_1}, \overbrace{\begin{pmatrix} 1 & -s_x^3 \\ -s_y^1 & 1 \end{pmatrix}}^{\mathbf{M}_2}, \overbrace{\begin{pmatrix} 1 & s_x^3 \\ -s_y^0 & 1 \end{pmatrix}}^{\mathbf{M}_3} \right\} \quad (3)$$

and translation vectors

$$\mathbf{t} = \left\{ \underbrace{(a_x^0, a_y^0)}_{\mathbf{t}_0} = \mathbf{0}, \underbrace{(a_x^1, a_y^1)}_{\mathbf{t}_1}, \underbrace{(a_x^2, a_y^2)}_{\mathbf{t}_2}, \underbrace{(a_x^3, a_y^3)}_{\mathbf{t}_3} \right\} \quad (4)$$

can be built. To enforce global consistency, the transformation matrices are circularly linked by mutual parameters. With respect to these transformations, each target pixel $\hat{\mathbf{p}}$ is originated from input pixel \mathbf{p} according to

$$\hat{\mathbf{p}} = \mathbf{M}_i^{-1} \cdot (\mathbf{p} - \mathbf{t}_i), \quad i \in \{0, \dots, 3\}, \mathbf{p} \in I_i. \quad (5)$$

The free parameters are estimated by solving the linear equation system stated in App. A in the least-squares sense. The first translation vector $\mathbf{t}_0 = (a_x^0, a_y^0)$ is fixed in order to obtain a unique result, while the weighting factor $\omega_{\mathcal{X}} = (n_t + n_r + n_b + n_i) / n_{\mathcal{X}}$ stated in App. A ensures that no overlapping area is favored over the others.

Outlier rejection As a consequence of the presence of homogeneous areas and the poor camera resolution, the displacement vector fields are likely to contain some outliers disrupting the transformation estimation. For this reason, an outlier rejection is used to achieve an outlier-free displacement vector field.

Given the complete shift vector field, initial transformations are estimated. According to these transformations, the correspondences with the highest backprojection errors are removed. These steps are repeated recursively until some predefined abort criterion, e.g. the highest tolerable error, is satisfied.

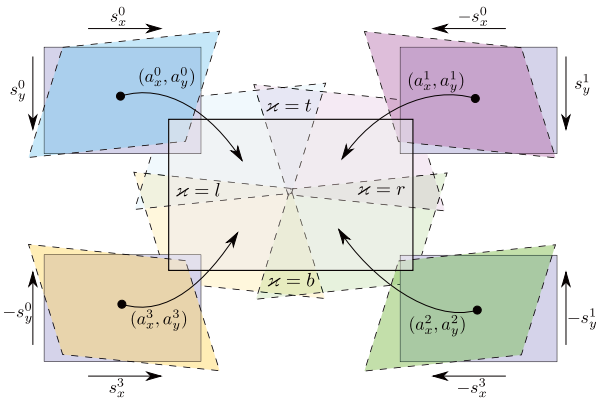


Figure 4. The transformations to be estimated and their parameters.

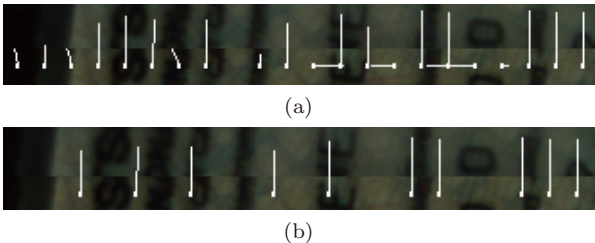


Figure 5. An optical flow vector field within the upper overlapping area (a) before and (b) after outlier rejection.

Fig. 5 shows an exemplary displacement vector field of the upper overlapping region before and after the outlier rejection.

5 Examples and Experimental Evaluation

Since this approach optimizes the visual quality of the composed image, just some exemplary results are shown in Fig. 7 in place of quantitative measures. Practical experiences show that OCR performed on the resulting images is working correctly in about 95% the cases.

For evaluating the runtimes of this method, we performed tests on a desktop computer equipped with an INTEL[®] CORE[™]2 QUAD Q9300 CPU at 2.50 GHz and 8 GB of RAM as well as a notebook computer with an INTEL[®] SU4100 CPU at 1.3 GHz and 2 GB of RAM.

For our time measurements we used the bended document shown in Fig. 7a and tried to compose it both with interpolation during the initial precalibration as well as the full document composing and without any interpolation. We implemented our methods in C++ disregarding any further optimization techniques like parallelization or specialized data structures.

In Fig. 6 one can see, as expected, that the runtime of our approach is strongly correlated with the size of the patches used for the shift detection. The abrupt declinations were caused by the SDR-internal Fourier transforms benefiting from patch sizes of the powers of two. Obviously, the interpolations were responsible just for a small amount of the runtimes, which can be dropped without a noticeable loss of quality. On desktop computers, documents with displacements up to 55 pixels can be composed within one second. For

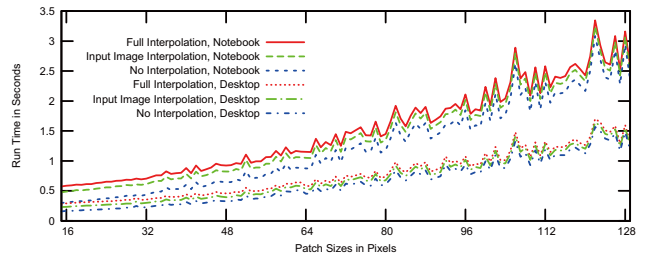


Figure 6. Runtimes obtained for composing the document shown in Fig. 7a on different hardware environments.

displacements of about 40 pixels, composing times of about half a second are possible. Reproducing the experiments on the notebook setup solely doubled the runtime.

6 Conclusion and Outlook

We presented an approach to compose subimages taken of a bended document using a quadocular camera setup. Our method used a signal-based correspondence detection via SDR. Experiments showed that our proposal yields accurate results as well as moderate computation time.

For further speedup, the parallelism of the independent image processing steps should be exploited. Furthermore the Fourier transforms used by SDR could also be implemented on parallel hardware architectures like GPUs in order to decrease the runtime.

References

- [1] Desko FP 500 Passport Scanner/Verifier. Data sheet, DESKO GmbH, Bayreuth, Germany, 2007.
- [2] E. H. Adelson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6):33–41, 1984.
- [3] T. Baumbach and W. Ortmann. Shift detection by restoration demonstrated by signal based point pattern matching. In *ICIAP*, pages 310–315, Washington, DC, USA, 1999. IEEE Computer Society.
- [4] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.
- [5] J. Davis. Mosaics of scenes with moving objects. In *CVPR*, pages 354–360, Washington, DC, USA, 1998. IEEE Computer Society.
- [6] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, pages 341–346, New York, NY, USA, 2001. ACM.
- [7] D. L. Milgram. Computer methods for creating photomosaics. *IEEE Transactions on Computers*, 24(11):1113–1119, 1975.
- [8] H. Süße, K. Voss, W. Ortmann, and T. Baumbach. Shift detection by restoration. In *CAIP*, pages 33–40, London, UK, 1999. Springer-Verlag.
- [9] M. Uyttendaele, A. Eden, and R. Szeliski. Eliminating ghosting and exposure artifacts in image mosaics. *CVPR*, 2:509, 2001.

A Equation System used for Parameter Estimation

$$\underbrace{\begin{pmatrix} \omega_t & 0 & 0 & 0 & 0 & 0 & -(p_{1,y}^t + q_{1,y}^t) \cdot \omega_t & 0 & 0 & 0 \\ 0 & \omega_t & 0 & 0 & 0 & 0 & 0 & -p_{1,x}^t \cdot \omega_t & q_{1,x}^t \cdot \omega_t & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_t & 0 & 0 & 0 & 0 & 0 & -(p_{n_t,y}^t + q_{n_t,y}^t) \cdot \omega_t & 0 & 0 & 0 \\ 0 & \omega_t & 0 & 0 & 0 & 0 & 0 & -p_{n_t,x}^t \cdot \omega_t & q_{n_t,x}^t \cdot \omega_t & 0 \\ \omega_r & 0 & -\omega_r & 0 & 0 & 0 & -p_{1,y}^r \cdot \omega_r & 0 & 0 & q_{1,y}^r \cdot \omega_r \\ 0 & \omega_r & 0 & -\omega_r & 0 & 0 & 0 & 0 & (p_{1,x}^r + q_{1,x}^r) \cdot \omega_r & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_r & 0 & -\omega_r & 0 & 0 & 0 & -p_{n_r,y}^r \cdot \omega_r & 0 & 0 & q_{n_r,y}^r \cdot \omega_r \\ 0 & \omega_r & 0 & -\omega_r & 0 & 0 & 0 & 0 & (p_{n_r,x}^r + q_{n_r,x}^r) \cdot \omega_r & 0 \\ 0 & 0 & \omega_b & 0 & -\omega_b & 0 & 0 & 0 & 0 & -(p_{1,y}^b + q_{1,y}^b) \cdot \omega_b \\ 0 & 0 & 0 & \omega_b & 0 & -\omega_b & 0 & p_{1,x}^b \cdot \omega_b & q_{1,x}^b \cdot \omega_b & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \omega_b & 0 & -\omega_b & 0 & 0 & 0 & 0 & -(p_{n_b,y}^b + q_{n_b,y}^b) \cdot \omega_b \\ 0 & 0 & 0 & \omega_b & 0 & -\omega_b & 0 & p_{n_b,x}^b \cdot \omega_b & q_{n_b,x}^b \cdot \omega_b & 0 \\ 0 & 0 & 0 & 0 & \omega_l & 0 & -q_{1,y}^l \cdot \omega_l & 0 & 0 & p_{1,y}^l \cdot \omega_l \\ 0 & 0 & 0 & 0 & 0 & \omega_l & 0 & -(q_{1,x}^l + p_{1,x}^l) \cdot \omega_l & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \omega_l & 0 & -q_{n_l,y}^l \cdot \omega_l & 0 & 0 & p_{n_l,y}^l \cdot \omega_l \\ 0 & 0 & 0 & 0 & 0 & \omega_l & 0 & -(q_{n_l,x}^l + p_{n_l,x}^l) \cdot \omega_l & 0 & 0 \end{pmatrix}}_{\mathbf{A} \in \mathbb{R}^{10 \times 2 \cdot (n_t + n_r + n_b + n_l)}} \cdot \underbrace{\begin{pmatrix} a_x^1 \\ a_x^2 \\ a_x^3 \\ a_y^1 \\ a_y^2 \\ a_y^3 \\ s_x^0 \\ s_x^1 \\ s_x^2 \\ s_x^3 \end{pmatrix}}_{\mathbf{x} \in \mathbb{R}^{10}} = \underbrace{\begin{pmatrix} (p_{1,x}^t - q_{1,x}^t) \cdot \omega_t \\ (p_{1,y}^t - q_{1,y}^t) \cdot \omega_t \\ \vdots \\ (p_{n_t,x}^t - q_{n_t,x}^t) \cdot \omega_t \\ (p_{n_t,y}^t - q_{n_t,y}^t) \cdot \omega_t \\ (q_{1,x}^r - p_{1,x}^r) \cdot \omega_r \\ (q_{1,y}^r - p_{1,y}^r) \cdot \omega_r \\ \vdots \\ (q_{n_r,x}^r - p_{n_r,x}^r) \cdot \omega_r \\ (q_{n_r,y}^r - p_{n_r,y}^r) \cdot \omega_r \\ (q_{1,x}^b - p_{1,x}^b) \cdot \omega_b \\ (q_{1,y}^b - p_{1,y}^b) \cdot \omega_b \\ \vdots \\ (q_{n_b,x}^b - p_{n_b,x}^b) \cdot \omega_b \\ (q_{n_b,y}^b - p_{n_b,y}^b) \cdot \omega_b \\ (q_{1,x}^l - p_{1,x}^l) \cdot \omega_l \\ (q_{1,y}^l - p_{1,y}^l) \cdot \omega_l \\ \vdots \\ (q_{n_l,x}^l - p_{n_l,x}^l) \cdot \omega_l \\ (q_{n_l,y}^l - p_{n_l,y}^l) \cdot \omega_l \end{pmatrix}}_{\mathbf{b} \in \mathbb{R}^{2 \cdot (n_t + n_r + n_b + n_l)}}$$

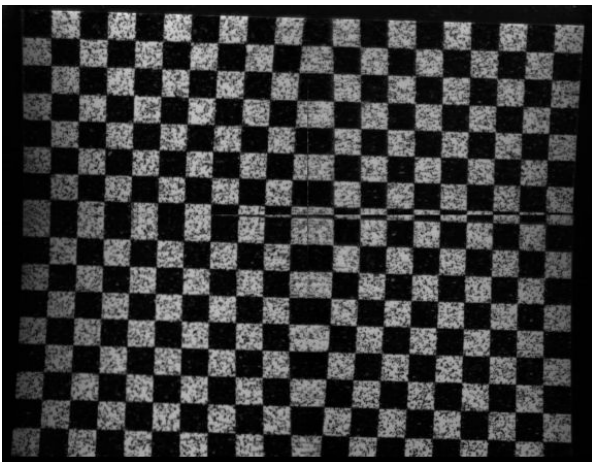
B Exemplary Results



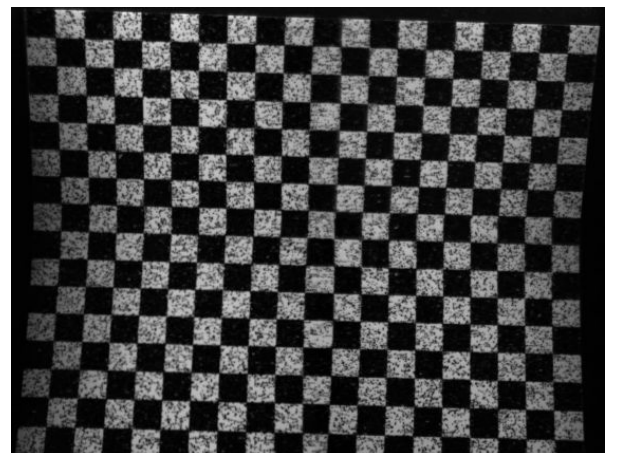
(a)



(b)



(c)



(d)

Figure 7. These images show (a) an extremely bended document and (c) a checkerboard image both composed according to the initial calibration and the results of our approach in Figures (b) and (d) respectively. Note the extent of the affine transformations visualized by the straight lines highlighted in image (b) as well as the checkerboard structure in image (d).