Reconstruction of 3D Surface and Restoration of Flat Document Image from Monocular Image Sequence

Hiroki Shibayama, Yoshihiro Watanabe, and Masatoshi Ishikawa

Graduate School of Information Science and Technology, University of Tokyo, Japan
{Hiroki_Shibayama,Yoshihiro_Watanabe,
Masatoshi_Ishikawa}@ipc.i.u-tokyo.ac.jp

Abstract. There is a strong demand for the digitization of books. To meet this demand, camera-based scanning systems are considered to be effective because they could work with the cameras built into mobile terminals. One promising technique proposed to speed up book digitization involves scanning a book while the user flips the pages. In this type of camera-based document image analysis, it is extremely important to rectify distorted images. In this paper, we propose a new method of reconstructing the 3D deformation and restoring a flat document image by utilizing a unique planar development property of a sheet of paper from a monocular image sequence captured while the paper is deformed. Our approach uses multiple input images and is based on the natural condition that a sheet of paper is a developable surface, enabling high-quality restoration without relying on the document structure. In the experiments, we tested the proposed method for the target application using images of different documents and different deformations, and demonstrated its effectiveness.

1 Introduction

There is a growing demand for camera-based document analysis and recognition. Book digitization, a relatively new application of camera-based document capturing which images all of the pages in a book, has a rapidly expanding market globally and is attracting various types of potential users, including libraries, corporations, and general users of books, official documents, and notes. However, the conventional technology cannot meet the demands for ease-of-use and high-speed book digitization. One emerging solution that meets these demands is *Book Flipping Scanning* [1]. This is a new style of scanning in which all pages of a book are captured while a user continuously flips through the pages without stopping at each page. Although this new technology has had a tremendous impact in the field of book digitization, a compact, expensive, high-speed 3D sensing module would be required to allow a massive number of potential users to try this scanning style with portable terminals such as mobile phones.

Focusing on this promising new approach, we attempted to implement *Book Flipping Scanning* with a single camera. Fig. 1 shows our concept. With this

K.M. Lee et al. (Eds.): ACCV 2012, Part IV, LNCS 7727, pp. 350–364, 2013.

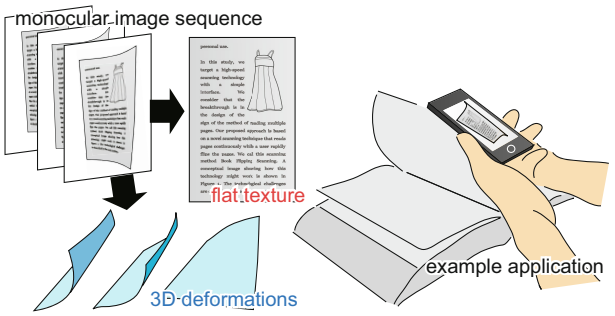

Fig. 1. Our method reconstructs surface deformations and flat document images using a monocular image sequence

system, the user can hold a camera and merely flip through a document held in front of it, allowing all information to be captured rapidly and in high definition.

For this purpose, the rectification problem described above is also essential. In our target application, multiple images are assumed as the input. This assumption enables a new possibility of allowing a flat document to be restored even if the document structure is unknown. However, because paper is a non-rigid surface, a deformation model is needed to realize this approach. A sheet of paper exhibits the characteristics of a so-called developable surface [2–5], an interesting property that can be exploited to meet this requirement.

Based on the design concept, this paper describes a new rectification technique for reconstructing 3D deformations and restoring a flat document image from multiple monocular images captured from a deformed sheet of paper. The deformations and restoration of a flat document image are not independent; however, the complementary relationships described by the developable property of a sheet of paper enables us to solve this problem. The proposed method employs a different approach from the conventional ones where the document structure in a single captured image is utilized. In particular, this technique is suitable for high-speed book digitization with a single camera, an application which is expected to show explosive growth.

2 Related Work

Book Digitization Using a Camera. Obtaining the 3D surface shape of a document simplifies unwarping of the captured image because this operation can be described as a planar development. For example, methods employing a stereo-camera system [6] and the so-called Shape from Shading technique [7, 8] are proposed. As for the scanning styles employed, these are typically the same as those used in flatbed scanners, requiring the user to scan page-by-page. In order to speed up book digitization, *Book Flipping Scanning* was developed to scan all of the pages in a book while a user flips through the pages, by obtaining the 3D surface shape using a high-speed, structured-light-based sensing system [1]. In addition, distorted images with various deformations can be restored to a

flat document image using a model-based estimation involving the developable-surface condition [3]. These systems realize camera-based book scanning based on 3D measurement techniques. In contrast, our challenge here is to develop a new book digitization technique similar to *Book Flipping Scanning* but based on a single-camera configuration.

Rectification of Warped Document Images. Recently, optical character recognition (OCR) has found new potential in a vast array of applications involving capturing document images with mobile phone cameras. However, OCR is less accurate when the characters are distorted. Although new OCR techniques have been proposed recently [9, 10], basically a rectification step is required. The typical methods utilize lines of texts. The rectification can be described using the assumption that the lines of text are horizontal in the flat document. However, extraction of lines of text is still challenging in various page layouts in which text and figures are mixed [4, 11, 12]. Also, there are two types of rectification techniques, describing the estimation as 2D warping [4, 13] and 3D reconstruction [12]. These methods solve the rectification problem under the assumption that the document contents and its structure are known. In addition, although there is a method utilizing the known bounding contour [14], it is difficult to detect it in case of book and there are typical cases where only insensible differences of contours between images can be found in spite of the large differences in 3D deformations. Instead of those approach, we assume the use of multiple images of deformed sheets of paper and describe the deformation as a fundamental property of paper in terms of a developable surface.

Non-rigid Surface Estimation. There have been some studies dealing with non-rigid objects, such as registration techniques between a reference image on a flat surface and a warped image on a deformed surface [15–17]. In addition, there are methods of reconstructing 3D surfaces from monocular video by using this kind of image registration, such as a learning-based approach [18] and a model-based approach [19, 20]. In particular, the model-based approach utilizes the inextensibility of the target material as a constraint. The limitation of those methods is that they require a template image captured in advance with known deformation of the target. In contrast, a method that requires no reference image has been proposed [21]. Although the preconditions for the methods meet our target application, high-quality restoration of a flat image is not the main target of that work. Also, the surface deformation is modeled based on a mesh structure, which increases the complexity of the inverse problem depending on the resolution of the mesh.

3 Reconstruction of Deformation of 3D Surface and Its Flat Texture

3.1 Overview of the Proposed Method

In this paper, the restoration of a single page is considered. An entire book can be digitized by applying this method to each captured page. In our method,

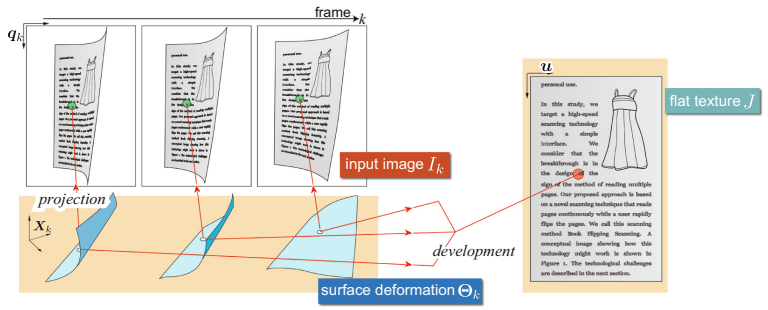

Fig. 2. Problem structure illustrating the correspondence mappings

the inputs are N images $\{I_k\}_{k=1}^N$ of a deformed page captured by a monocular camera. The outputs are N deformation shapes $\{\boldsymbol{\Theta}_k\}_{k=1}^N$ and a single restored flat document image J. Each shape $\boldsymbol{\Theta}_k$ is observed in each input image I_k.

The assumptions for this data are as follows. A single camera whose intrinsic parameters are known in advance captures the input images. The output restored image J is assumed to be a texture on the surface when it is developed into a plane. The target object is assumed to be developed into a plane without any extension, contraction, or tearing, which means that the Gaussian curvature is identically zero over the entire surface [2]. Also, in order to improve the accuracy, the aspect ratio of the paper is known in advance. Considering actual applications such as book digitization, these assumptions are considered to be acceptable.

The observation situation is explained as follows. At every frame, the shape of the target object continues to be deformed. The camera captures the texture deformation depending on the shape deformation in each input image I_k. The captured texture deformation contains clues to reconstruct the shape $\boldsymbol{\Theta}_k$ and the original flat texture J. This is illustrated in Fig. 2.

We model the transformation between two coordinate systems, those of the k-th camera image and the flat document image, as follows:

$$\boldsymbol{q}_k = \boldsymbol{g}\left(\boldsymbol{u}; \boldsymbol{\Theta}_k\right) = \boldsymbol{g}_k(\boldsymbol{u}) \tag{1}$$

In this equation, the point $\boldsymbol{u}$ on the flat image is projected onto the point $\boldsymbol{q}_k$ on the input image plane through process $\boldsymbol{g}_k$ involving the deformation and observation process, which is controlled by the deformation parameter $\boldsymbol{\Theta}_k$ in our task. Using Equation (1), all of the pixels can be projected onto the captured image. Therefore, each camera image I_k can be formulated as follows:

$$I_k\left(\boldsymbol{q}\right) = \sum_{\boldsymbol{u} \in \{\boldsymbol{u} \mid \|\boldsymbol{g}_k(\boldsymbol{u}) - \boldsymbol{q}\| \leq \epsilon\}} B\left(\|\boldsymbol{g}_k(\boldsymbol{u}) - \boldsymbol{q}\|\right) J(\boldsymbol{u}) \tag{2}$$

Here, $I_k(\boldsymbol{q})$ and $J(\boldsymbol{u})$ represent the brightness value at each point $\boldsymbol{q}$ and $\boldsymbol{u}$, respectively. Also, $B\left(d\right)$ is the lens blurring effect (the Point Spread Function, or PSF). In this paper, we define the PSF by using a Gaussian operation.

The target can deform its shape freely under the developable conditions. Although our method allows such free deformation, the developed flat image J is

unique, as shown in Fig. 2. In this way, introducing the idea of development gives temporal consistency between the captured images. Therefore, increasing the number of input images increases the number of conditions, and this framework enables us to construct an environment for solving the inverse estimation problem for the surface deformations and the developed flat document image. Also, as the number of input images increases, the synthesized information for estimating a single restored, unwarped image becomes more rich, so that a high-resolution image can be recovered. In addition, the high-quality developed flat image J can recursively contribute to improvement of the accuracy in surface estimation.

Based on this approach, we propose estimating the surface deformations and a developed image from multiple images captured by a monocular camera based on the following framework:

$$\min_{\{\tilde{\boldsymbol{\Theta}}_k\}_{k=1}^N, \tilde{\boldsymbol{J}}} \sum_k \sum_{\boldsymbol{q}} \|I_k(\boldsymbol{q}) - \tilde{I}_k(\boldsymbol{q}; \tilde{\boldsymbol{\Theta}}_k, \tilde{\boldsymbol{J}})\|^2 + f(\tilde{\boldsymbol{J}}) \tag{3}$$

$\tilde{\boldsymbol{J}}$ is the vector aligning the brightness value of the image $\tilde{J}$. Other vector notations have similar definitions. $\tilde{I}_k$ is generated from the estimated shape $\tilde{\boldsymbol{\Theta}}_k$ and the restored image $\tilde{\boldsymbol{J}}$. The second term is a regularization term, and the function f constrains the restored image J. In our experiments, the square norm of the Laplacian image of the restored image was used.

In order to formulate this framework concretely, it is crucially important to describe the developable non-rigid body with a small parameter set so as not to increase the problem scale. In this paper, we introduce the concept of a developable surface used in the field of differential geometry. If another typical model is used to represent the deformation, uniqueness is no longer ensured, and it is difficult to maintain stable estimation accuracy. Details of the developable surface modeled in this paper are shown in Section 3.2.

Also, in the problem represented in Equation (3), simultaneous estimation of both the deformation $\{\tilde{\boldsymbol{\Theta}}_k\}_{k=1}^N$ and the restored image $\tilde{J}$ is required. However, such simultaneous estimation complicates the inverse problem, making the solution difficult to reach. Therefore, in our method, the surface and the developed texture are estimated alternately. In addition, our method estimates the solution based on nonlinear optimization, so that initial estimation parameters need to be given. Based on these aspects, our method is organized as follows.

1. **Initial deformation estimation.** In this step, the initial deformation parameters are estimated. The estimation is allowed to be rough.
2. **Reconstruction of the 3D deformation.** Using the input image I_k and the restored image J, each 3D surface deformation $\boldsymbol{\Theta}_k$ is estimated.
3. **Reconstruction of the flat texture.** Using all of the input images $\{I_k\}_{k=1}^N$ and the estimated deformations $\{\boldsymbol{\Theta}_k\}_{k=1}^N$, a developed flat image J is restored.
4. **Iteration** Steps 2 and 3 are iterated until convergence.

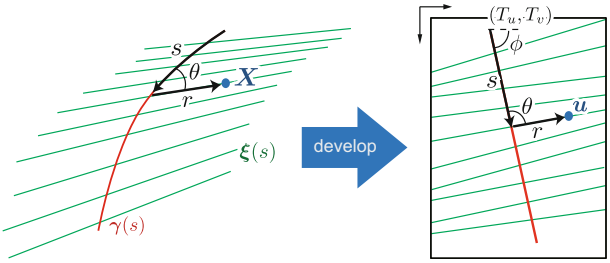

Fig. 3. Developable surface and its planar development

3.2 Developable Surface for Monocular Reconstruction

The developable surface is parameterized by the family of lines $(\boldsymbol{\gamma}(s), \boldsymbol{\xi}(s))$. Here, $\boldsymbol{\gamma}$ is a curve called the *directrix*. Also, $\boldsymbol{\xi}$ is a vector called the *ruling*. The Gaussian curvature is identically zero over the entire surface [2]. This condition makes it possible to develop the surface to a plane. An example developable surface is shown in Fig. 3.

A 3D point $\boldsymbol{X}$ on the surface is described by two parameters (s, r), as in the following equations:

$$\boldsymbol{X}(s,r) = \boldsymbol{\gamma}(s) + r\boldsymbol{\xi}(s), \quad \boldsymbol{\xi}(s) = \frac{\boldsymbol{\gamma}''(s) \times \boldsymbol{\gamma}'''(s)}{\|\boldsymbol{\gamma}''(s) \times \boldsymbol{\gamma}'''(s)\|} \tag{4}$$

Parameters (s, r) mean the lengths of the *directrix* and *ruling*, respectively. In other words, the 3D point $\boldsymbol{X}$ on the surface can be reached after moving by s along the *directrix* and moving by r along the *ruling*.

In this definition, once the curve $\boldsymbol{\gamma}$ is defined, the entire surface is determined. This means that in the surface reconstruction, we need to estimate only the parameters describing the *directrix*. This is a powerful advantage because the introduction of the developable surface model does not make the solution complex.

In addition, in the developable surface, on the planar development, *directrix* and *ruling* are transformed to straight lines. Therefore, as shown in Fig. 3, the coordinates on the flat image are as follows:

$$\boldsymbol{u} = \begin{bmatrix} T_u \\ T_v \end{bmatrix} + s \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} + r \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{5}$$

Here, the registration parameters are set to $\boldsymbol{\Phi} = (T_u, T_v, \phi)$, and θ is the angle between the transformed *directrix* and *ruling*, which can be obtained from the surface model.

As the curve $\boldsymbol{\gamma}$, we use the following Bezier curve:

$$\boldsymbol{B}(t) = \sum_{i=0}^{n-1} \boldsymbol{p}_i \binom{n-1}{i} t^i (1-t)^{n-1-i} \tag{6}$$

Here, the points $\boldsymbol{P} = (\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n-1})$ are the control points of the Bezier curve. The developable surface deformation can be defined uniquely only if the control points are provided. The parameter t in the Bezier curve can be transformed to the arc length s. In this case, a position on the surface is defined using the 2D coordinates $\boldsymbol{\eta} = (t, r)$.

Based on these definitions, the deformation parameter $\boldsymbol{\Theta}$ is defined as follows:

$$\boldsymbol{\Theta} = \{\boldsymbol{\Phi}, \boldsymbol{P}\} \tag{7}$$

3.3 Initial Estimation for Recovering Rough Surfaces

This section describes the initial estimation step. Our method estimates each deformation by using two input images. First, we get the corresponding feature points in two images. Also, in the steps of our deformation recovery, we utilize parts of the page outline, including the four corners and some points on the edges. The numbers and positions of the points on the edges are arbitrary. In the experiments, we set the image pair as two successive images in the captured video. Also, in order to remove the outliers caused by the same characters appeared in a page, we applied adaptive patch segmentation in the images and found the corresponding points only between the patches located in the same position in two images. As feature descriptors, SURF features were used [22].

The corresponding positions on the two input images I_A and I_B are $\boldsymbol{q}_A$ and $\boldsymbol{q}_B$, respectively. Similarly, the coordinates on the surfaces are $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_B$, and the coordinates on the restored flat image are $\boldsymbol{u}_A$ and $\boldsymbol{u}_B$. The values with the tildes mean estimated ones.

If the two points $\boldsymbol{q}_A$ and $\boldsymbol{q}_B$ are corresponding points, the estimated points $\tilde{\boldsymbol{u}}_A$ and $\tilde{\boldsymbol{u}}_B$ also are required to be located at the same position. The estimation errors for the feature points, the corners, and the edges are defined as Equations (8), (9) and (10), respectively:

$$E_{fp} = \sum_{i=1}^{N_{fp}} v_{fp}(\|\boldsymbol{q}_A^i - \tilde{\boldsymbol{q}}_A^i\|^2 + \|\boldsymbol{q}_B^i - \tilde{\boldsymbol{q}}_B^i\|^2) + w_{fp}\|\tilde{\boldsymbol{u}}_A^i - \tilde{\boldsymbol{u}}_B^i\|^2 \tag{8}$$

$$E_c = \sum_{I=A,B} \sum_{j_I=1}^{N_{cI}} v_c\|\boldsymbol{q}_I^{j_I} - \tilde{\boldsymbol{q}}_I^{j_I}\|^2 + w_c\|\boldsymbol{u}_I^{j_I} - \tilde{\boldsymbol{u}}_I^{j_I}\|^2 \tag{9}$$

$$E_e = \sum_{I=A,B} \sum_{k_I=1}^{N_{eI}} v_e\|\boldsymbol{q}_I^{k_I} - \tilde{\boldsymbol{q}}_I^{k_I}\|^2 + w_e\|\boldsymbol{u}_I^{k_I} - \tilde{\boldsymbol{u}}_I^{k_I}\|_{(x,y)}^2 \tag{10}$$

Here, N_{fp}, N_{cI}, and N_{eI} are the total numbers of each point; i, j_I, and k_I are indices; and v_{fp}, ω_{fp}, v_c, ω_c, v_e, and ω_e are constant weights. The target position of the flat-image coordinates for the contour points are calculated from the given aspect ratio. In the second term of Equation (10), the distance only in the x or y coordinate is calculated, depending on the edge direction.

We assume a sample with various deformations described using N_{pca} principal vectors based on principal component analysis (PCA), as follows:

$$\boldsymbol{P} = \sum_{i=1}^{N_{pca}} \alpha_i \boldsymbol{\rho}_i \tag{11}$$

This sample deforms its shape at a fixed position. Therefore, in order to fit the sample deformation to the observed one, a coordinate transformation including rotation $\boldsymbol{R}$ and translation $\boldsymbol{T}$ must be estimated.

Based on this configuration, the initial deformations are estimated using two images:

$$\min_{\tilde{\boldsymbol{\alpha}}^A, \tilde{\boldsymbol{\alpha}}^B, \tilde{\boldsymbol{H}}^A, \tilde{\boldsymbol{H}}^B, \tilde{\boldsymbol{\Phi}}^A, \tilde{\boldsymbol{\Phi}}^B, \tilde{\boldsymbol{R}}, \tilde{\boldsymbol{T}}} E_{fp} + E_c + E_e \tag{12}$$

Here, $\tilde{\boldsymbol{H}}^A$ and $\tilde{\boldsymbol{H}}^B$ are the sets of the corresponding points $\tilde{\boldsymbol{\eta}}^A$ and $\tilde{\boldsymbol{\eta}}^B$ on the surface. Normally the estimated points $\tilde{\boldsymbol{u}}$ on the restored flat image can be calculated from the observed points $\boldsymbol{q}$ and the deformation $\boldsymbol{\Theta}$. However, this calculation cannot be analytically defined. Therefore, instead of employing this calculation, we add the points $\boldsymbol{\eta}$ for estimation of the mapping between the observed point and the flat-image point. This problem is solved by using the Levenberg–Marquardt method.

3.4 Reconstruction of 3D Surface Deformation

In the surface reconstruction, the developed flat texture J is fixed. Each deformation $\boldsymbol{\Theta}_k$ for each captured image I_k is individually reconstructed. This reconstruction requires solving the following equation:

$$\min_{\{\tilde{\boldsymbol{\Theta}}_k\}_{k=1}^N} \sum_k \sum_{\boldsymbol{q}} \|I_k(\boldsymbol{q}) - \tilde{I}_k(\boldsymbol{q}; \tilde{\boldsymbol{\Theta}}_k)\|^2 + f(\boldsymbol{J}) \tag{13}$$

In order to search for the solution $\boldsymbol{\Theta}_k$, an iterative technique is essential. However, as the surface parameters change, the correspondence mapping $\boldsymbol{g}_k$ in Equation (1) must also change. In this case, its calculation must be repeated at every iterative update of the parameters. This leads to the critical problem of a long calculation time.

Therefore, in a similar way to the initial deformation estimation, instead of brightness evaluation, we evaluate the distance between the corresponding feature points in the input image I_k and the estimated captured image $\tilde{I}_k$, which is generated based on Equation (2), to reconstruct the surface deformation. Although the evaluation function is different from the reconstruction of a flat image, the meanings are essentially equivalent.

First, using the deformation parameter $\tilde{\boldsymbol{\Theta}}_k$ estimated in the previous iteration and the restored image J, the estimated captured image $\tilde{I}_k$ is generated. Between two images I_k and $\tilde{I}_k$, the corresponding feature points are obtained. Also, using the deformation parameter, the point $\boldsymbol{u}_i$ on the flat image for the point $\boldsymbol{q}_i$ on the observed image is calculated. In addition, corners and edge points are obtained

from the input images. However, if this is the first time for this step after the initial estimation, the deformation recovery is achieved by using only outline information.

The minimization problem for this step is described as follows:

$$\min_{\tilde{\boldsymbol{\Theta}}_k, \tilde{\boldsymbol{H}}} \sum_{i=1}^{N_p} v_i \|\boldsymbol{q}_i - \tilde{\boldsymbol{q}}_i\|^2 + w_i \|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2 \tag{14}$$

Here, N_p is the total number of points, and i is the index, and v_i and w_i are constant weights. Three sets of weights are switched for the feature points, corners, and edge points. For the same reason as described in Section 3.3, the point set $\tilde{\boldsymbol{H}}$ is added for efficient estimation. The second term allows us to fix the flat image J. In the case of edge points, this term calculates the distance only in x or y direction as described in Section 3.3. This problem is solved by using the Levenberg–Marquardt method.

3.5 Flat Document Image Restoration

In this step, the target optimization problem is formulated as follows:

$$\min_{\tilde{\boldsymbol{J}}} \sum_k \sum_{\boldsymbol{q}} \|I_k(\boldsymbol{q}) - \tilde{I}_k(\boldsymbol{q}; \tilde{\boldsymbol{J}})\|^2 + f(\tilde{\boldsymbol{J}}) \tag{15}$$

Based on this minimization problem, the input images can be synthesized. This synthesis requires the function $\boldsymbol{g}_k$ shown in Equation (2). This function establishes the correspondence between the camera coordinates and the flat texture coordinates.

This correspondence process is decomposed into two stages. The first stage is to establish the correspondence between the 2D point $\boldsymbol{q}$ on the camera image and the 3D point $\boldsymbol{X}$ on the surface. The second stage is to establish the correspondence between the 3D point $\boldsymbol{X}$ on the surface and the 2D point $\boldsymbol{u}$ on the developed plane. This relationship is described in Section 3.2.

As a result, the correspondence between the camera coordinates and the flat texture coordinates is established, so that the problem formulated as Equation (15) can be solved. In this paper, we solve this problem by using the conjugate gradient method.

4 Experiments

We captured, in advance, various deformations while flipping through a book and generated sample deformations for the initial estimation. We set the number of control points for the *directrix* described in Equation (6) to three. The weights in the surface estimation were set to focus on the feature points compared with the points on the contour. The repeat count described in Section 3.1 was 5. As a comparison, we used a non-rigid registration technique based on a thin-plate spline (TPS) [23]. The unwarping was tested by giving the corner correspondences between the input image and the restored image.

4.1 Evaluations Using Various Datasets

In the experiment, the method was applied to a synthetic input dataset. We used 10 document images whose resolutions were 1,032 × 1,458 pixels. The original images are shown in Fig. 4. Using each image, the observed image sequences were created by applying deformations and rigid motion. We prepared four sets of sequences (two deformations and two rigid motions). The observed sequence consisted of three images. The input resolutions of the images shown in the figure were 2,048 × 2,048.

The restored results are shown in Fig. 5. We set the resolution at 1,032 × 1,458. In the figure, the input images, the image unwarped by TPS, and the image unwarped by the proposed method are shown. Also the example reconstructed

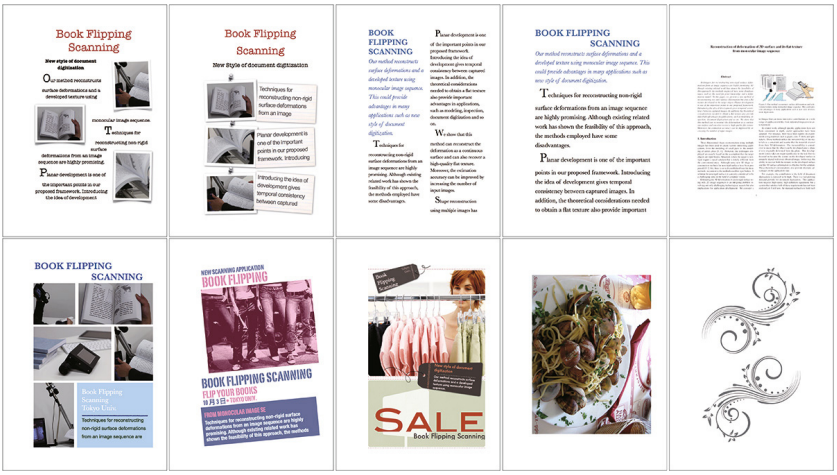

Fig. 4. The used original document images

Fig. 5. The synthesized input images and the restored images. Four example results are shown. In each example, the input image and the enlarged image (left), image unwarped by TPS (center), and image unwarped by the proposed method (right) are shown.

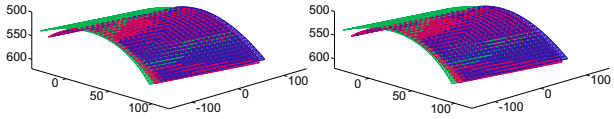

Fig. 6. The example reconstructed deformation. Left: the reconstructed deformation. Right: The true deformation.

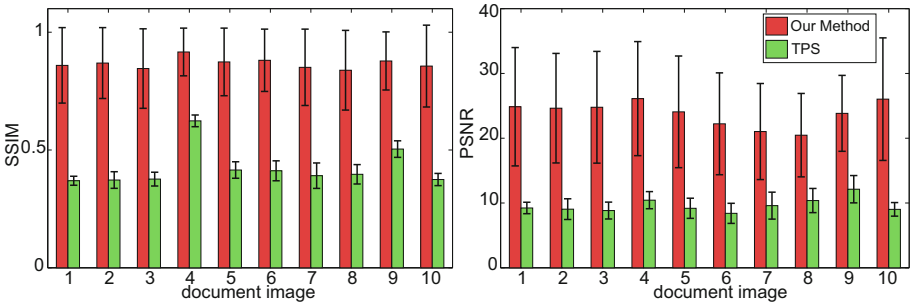

Fig. 7. The evaluated image quality. The label of the document image on the horizontal axis corresponds to the order in Fig. 4 (the upper left to the lower right).

deformation is shown in Fig. 6. In this figure, the reconstructed deformation and the true one are compared. The deformations observed in the three input images are drawn in different color meshes.

We also evaluated the image quality using peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [24]. The evaluated quality is shown in Fig. 7. The 10 document images with four different input set were evaluated. Compared with TPS warping, our method restored high-quality images. Also, our method achieved rectification with various deformations, and changes of the target document did not significantly affect the accuracy. Restoration succeed even when only a part of the document had a small illustration.

4.2 Experiments Using Book Flipping Images

The experimental setup is shown in Fig. 8. The camera captured 2,048 × 2,048-resolution images at 180 fps while the user flipped the pages. In this setup, a pattern projector (multi-line pattern) was also provided to obtain the true deformation based on a structured light technique. The pattern was switched on and off alternately for every captured frame [1]. In this experiment, only the images in which the pattern was not projected were used. The input sequence consisted of three images.

Fig. 9 shows the restored image after the initial estimation. As shown in the figure, the quality in this step was not high enough. Also, Fig. 9 shows an example of feature point correspondence between the input image and the estimated one in the deformation reconstruction step.

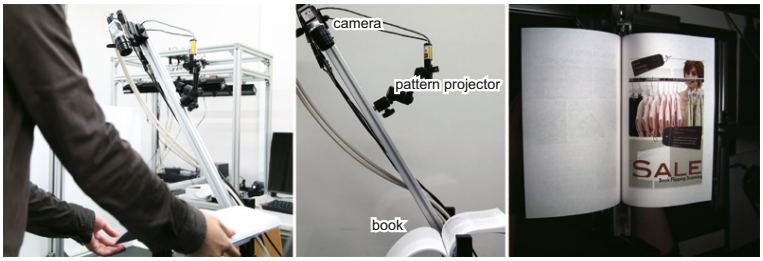

Fig. 8. Experimental setup and the captured images

Fig. 9. Restored images after initial estimation and the example of feature point correspondence in the deformation reconstruction

Fig. 10. Restored images of six example documents. In each example, the input image (left), image unwarped by TPS (center), and the image unwarped by the proposed method (right) are shown.

Fig. 10 shows the restoration results. The resolution of the rectified images was 495 × 740. In the figure, the input images and the images unwarped using TPS are also shown. Fig. 11 shows the examples of the errors between the input images and the estimated ones. The errors in the brightness values are shown.

Fig. 12 shows an example of the reconstructed deformation. In the figure, the reconstructed deformation and the true deformation measured based on a structured light technique are compared. Since the absolute error is difficult to be obtained because the reconstruction is based on a single camera, the quantitative comparison is shown. Fig. 13 shows the image quality evaluating the flipped 11-page documents. As the correct images, those shown in Fig. 4 were used. On the reconstructed image, non-uniform shading effect was left. This reduced the qualitative performance even when we achieved the geometric rectification

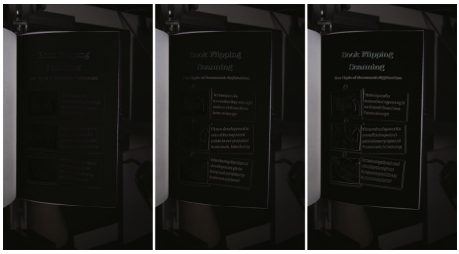

Fig. 11. Example of the error between the input image and the estimated one. The images shown are the difference images obtained by subtracting these images.

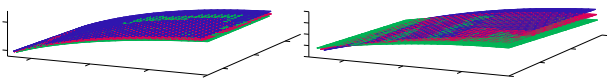

Fig. 12. Example of the reconstructed deformation. Left: the reconstructed deformation. Right: the measured deformation using the active stereo technique.

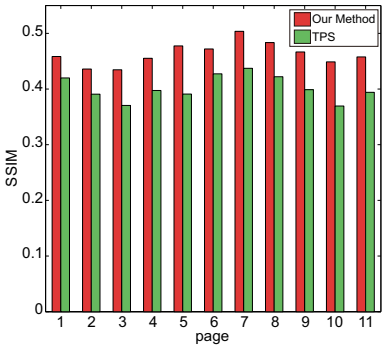

Fig. 13. The evaluated image quality

successfully. Therefore, the visibly good results in Fig. 10 was not reflected to the evaluation perfectly.

5 Conclusion

We have presented a new method for reconstructing 3D deformation and restoring an unwarped flat document image from multiple images. The planar development feature involved in the paper allows us to solve the problem based on temporal consistency between input images. Our method can handle various types of documents without knowing their structure and layout in advance. Experiments showed that our method worked well and reconstructed both the surface and the developed texture. In addition, the image quality was evaluated based on the PSNR and SSIM, and the results were considered to be acceptable. We confirmed that our proposed method was effective in realizing a new book scanning system.

Acknowledgments. This research was funded by Strategic Information and Communications R&D Promotion Programme (SCOPE).

References

1. Nakashima, T., Watanabe, Y., Komuro, T., Ishikawa, M.: Book flipping scanning. In: Adjunct Proceedings of UIST, pp. 79–80 (2009)
2. Carmo, M.P.D.: Differential Geometry of Curves and Surfaces. Prentice Hall (1976)
3. Watanabe, Y., Nakashima, T., Komuro, T., Ishikawa, M.: Estimation of non-rigid surface deformation using developable surface model. In: Proceedings of ICPR, pp. 197–200 (2010)
4. Cao, H., Ding, X., Liu, C.: Rectifying the bound document image captured by the camera: A model based approach. In: Proceedings of ICDAR, pp. 71–75 (2003)
5. Liang, J., DeMenthon, D., Doermann, D.: Unwarping Images of Curved Documents Using Global Shape Optimization. In: Proceedings of CBDAR, pp. 25–29 (2005)
6. Yamashita, A., Kawarago, A., Kaneko, T., Miura, K.T.: Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system. In: Proceedings of ICPR, pp. 482–485 (2004)
7. Zhang, Z., Tan, C.L., Fan, L.: Estimation of 3D shape of warped document surface for image restoration. In: Proceedings of ICPR, pp. 486–489 (2004)
8. Prados, E., Camilli, F.: A unifying and rigorous shape from shading method adapted to realistic data and applications. Journal of Mathematical Imaging and Vision 25, 307–328 (2006)
9. Hase, H., Shinokawa, T., Yoneda, M., Suen, C.: Recognition of rotated characters by eigen-space. In: Proceedings of ICDAR, pp. 731–735 (2003)
10. Narita, R., Ohyama, W., Wakabayashi, T., Kimura, F.: Three dimensional rotation-free recognition of characters. In: Proceedings of ICDAR, pp. 824–828 (2011)
11. Liang, J., DeMenthon, D., Doermann, D.: Geometric rectification of camera-captured document images. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 591–605 (2008)

12. Tian, Y., Narasimhan, S.G.: Rectification and 3D Reconstruction of Curved Document Images. In: Proceedings of ICCV, pp. 377–384 (2011)
13. Zhang, L., Tan, C.L.: Warped Image Restoration with Applications to Digital Libraries. In: Proceedings of ICDAR, pp. 192–196 (2005)
14. Gumerov, N.A., Zandifar, A., Duraiswami, R., Davis, L.S.: 3d structure recovery and unwarping of surfaces applicable to planes. International Journal of Computer Vision 66, 261–281 (2006)
15. Crum, W.R., Hartkens, T., Hill, D.L.G.: Non-rigid image registration: theory and practice. The British Journal of Radiology 77, 140–153 (2004)
16. Bartoli, A., Zisserman, A.: Direct estimation of non-rigid registrations. In: Proceedings of BMVC (2004)
17. Gay-Bellile, V., Bartoli, A., Sayd, P.: Direct estimation of nonrigid registrations with image-based self-occlusion reasoning. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 87–104 (2010)
18. Salzmann, M., Urtasun, R., Fua, P.: Local deformation models for monocular 3D shape recovery. In: Proceedings of CVPR (2008)
19. Salzmann, M., Pilet, J., Ilic, S., Fua, P.: Surface deformation models for nonrigid 3D shape recovery. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 1481–1487 (2007)
20. Brunet, F., Hartley, R., Bartoli, A., Navab, N., Malgouyres, R.: Monocular Template-Based Reconstruction of Smooth and Inextensible Surfaces. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part III. LNCS, vol. 6494, pp. 52–66. Springer, Heidelberg (2011)
21. Varol, A., Salzmann, M., Tola, E., Fua, P.: Template-free monocular reconstruction of deformable surfaces. In: Proceedings of the ICCV, pp. 1811–1818 (2009)
22. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
23. Sprengel, R., Rohr, K., Stiehl, H.S.: Thin-plate spline approximation for image registration. In: Proceedings of the IEEE Engineering in Medicine and Biology Society, pp. 1190–1191 (1996)
24. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: Proceedings of ICPR, pp. 2366–2369 (2010)