Video Mosaicing for Document Imaging

Noboru Nakajima† Akihiko Iketani† Tomokazu Sato†‡
Sei Ikeda‡ Masayuki Kanbara†‡ Naokazu Yokoya†‡
†*Common Platform Software Research Laboratories, NEC Corporation,*
8916-47 Takayama, Ikoma 630-0101, Japan
‡*Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0192, Japan*
{n-nakajima@ay,iketani@cp}.jp.nec.com, {tomoka-s,sei-i,kanbara,yokoya}@is.aist-nara.ac.jp

Abstract

This paper describes a real-time video mosaicing method for digitizing paper documents with mobile and low-resolution video cameras. Our proposed method is based on the structure-from-motion technique with real-time processing. Therefore, it can dewarp an image suffering from perspective deformation caused by a slanted and curved document surface. The curved surfaces are reconstructed as an expanded flat image by 3D polynomial surface regression and projection onto a flat mosaic plane. The order of the polynomial is automatically determined according to the geometric AIC. Moreover, the real-time scanning and reconstruction process offers interactive user instruction on the style of user's camera handling, such as appropriateness of camera posture and scanning speed. Manual camera scanning for video mosaicing captures a point on the target multiple times and that allows other merits. They include super-resolution and shading correction. The experiments demonstrated the feasibility of our approach.

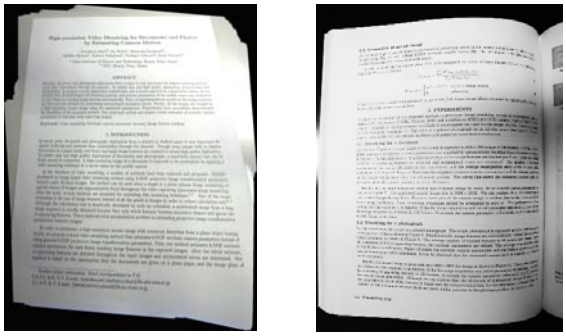
1. Introduction

Digital camera devices are getting increasingly commoditizing. But in contrast to their facility, their quality and assurance of document digitization are not sufficient for document analysis and recognition. The literature in the DAR field describes several technical challenges for their practical realization such as the followings [1-3]:

(1) dissolution for deficiency in resolving power,
(2) dewarping of curved surface suffering perspective deformation,
(3) elimination of illumination irregularity,
(4) target localization and segmentation.

A full A4 page image at 400 dpi requires more than 14.7M pixels, whereas a flatbed scanner enables a few thousands dpi. For the resolution deficiency problem (1), video mosaicing is one of the most promising solutions. In video mosaicing, partial images of a document are captured as a video sequence, and multiple frames are stitched seamlessly into one large, high-resolution image. Conventional video mosaicing methods usually achieve pair-wise registration between two successive images, and construct a mosaic image projecting all the images onto a reference frame, in general, the first frame. Szeliski [4] has developed a homography-based method, which adopts 8-DOF planar projective transformation. After his work, various extensions have been proposed [7-10]. One of the major developments is the use of image features instead of all of the pixels for reducing computational cost [5-6]. All of these methods, however, align the input images to the reference frame, and thus will suffer perspective slanted effect if the reference frame is not accurately parallel to the target surface, even though a target page is rigidly planar, as shown in Figure 1(a).

There actually exists a severe image distortion (2) when capturing a curved target, as shown in Figure 1(b). The homography-based methods are applicable just when a target is a piecewise planar, or the optical



(a) Mosaic image generated by conventional methods (b) Curved page captured with a digital camera

Figure 1. Problems in digitized document image.

center of the camera is approximately fixed (panoramic mosaicing). Thus, if the target is curved, the above assumption no longer holds, and misalignment of images will degrade the resultant image. Although there are some video mosaicing methods that can deal with a curved surface, they require an active camera and a slit light projection device [10].

On the other hand, the usability of the system is also an important factor. Unfortunately, conventional video mosaicing methods have not focused on the usability of the system because interactive video mosaicing systems have not been developed yet.

In order to solve these problems, we employ a structure-from-motion technique with real-time processing. Our method is constructed of two stages as shown in Figure 2. In the first stage, a user captures a target document using a web-cam attached to a mobile computer (Figure 3). In this stage, camera parameters are automatically estimated in real-time and the user is supported interactively through a preview window for displaying a mosaic image on the mobile computer. After that, camera parameters are refined, and a high-resolution and distortion-free mosaic image is generated in a few minutes. The main contributions of our work can be summarized as follows: (a) perspective deformation correction from a mosaic image for a flat target, (b) dewarped mosaic image generation for a curved document, and (c) interactive user support on a mobile system. Camera scanning allows surplus textural information on the target, or a point on the document is captured multiple times in the image sequence. Other benefits to be reaped from our system include (d) super-high-definite image construction by super-resolution [11] for enhancing the solution for the low-resolution problem (1), and (e) shading correction by intrinsic image generation [12] for irregular illumination (3). In the objective of document capturing, the location of the target, at which a user principally points a camera, is straightforward here for the target location problem (4).

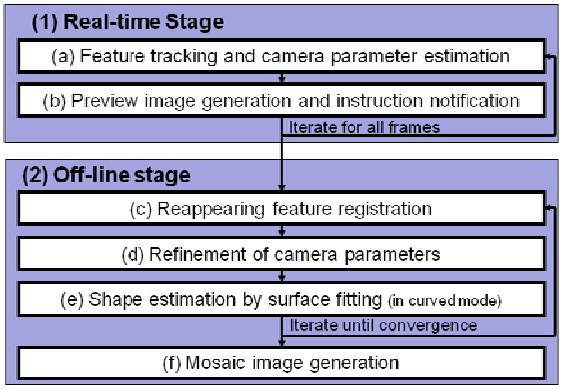


Figure 2. Flow of proposed method.

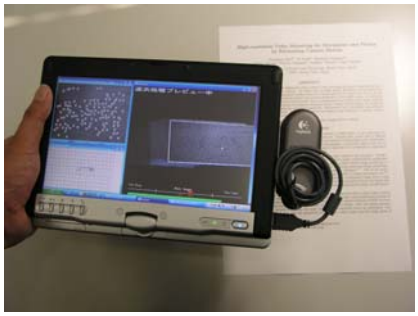


Figure 3. Mobile PC with web-cam.

The assumptions made in the proposed method are that intrinsic camera parameters, including lens distortion, are fixed and calibrated in advance. For curved documents, the curvature must lie along a one-dimensional direction, or be cylindrical, and varies smoothly on a page.

2. Video Mosaicing for Flat and Curved Document

This section describes a method for generating a high-resolution and distortion-free mosaic image from a video sequence. The flow of the proposed method is given in Figure 2. Although our system has two modes, one for a flat target (flat mode) and the other for a curved target (curved mode), the basic flow for these two modes is almost the same. In the real-time stage, the system carries out 3D reconstruction processes frame by frame by tracking image features (a). The preview of a temporal mosaic image is rendered in real-time and updated in every frame to show what part of the document has already been captured (b). After the real-time stage, the off-line stage is automatically started. In this stage, first, re-appearing image features are detected in the stored video sequence (c), and camera parameters and 3D positions of features estimated in the real-time stage are refined (d). If the system works in the curved mode, surface parameters are also estimated by fitting a 3D surface to an estimated 3D point cloud (e). After several iterations, a distortion-free mosaic image of high-resolution is finally generated (f) applying the super-resolution method. In the following sections, first, extrinsic camera parameters and an error function used in the proposed method are defined. Stages (1) and (2) are then described in detail.

2.1. Definition of Extrinsic Camera Parameter and Error Function

In this section, the coordinate system and the error function used for 3D reconstruction are defined. We define the coordinate system such that an arbitrary point $\mathbf{S}_p = (x_p, y_p, z_p)$ in the world coordinate system is projected to the coordinate $\mathbf{x}_{fp} = (u_{fp}, v_{fp})$ in the *f*-th image plane. Defining 6-DOF extrinsic camera parameters of the *f*-th image as a 3x4 matrix $\mathbf{M}_f$ the relationship between 3D coordinate $\mathbf{S}_p$ and 2D coordinate $\mathbf{x}_{fp}$ is expressed as follows:

$$(au_{fp}, av_{fp}, a)^T = \mathbf{M}_f (x_p, y_p, z_p, 1)^T, \quad (1)$$

where a is a normalizing parameter. In the above expression, $\mathbf{x}_{fp}$ is regarded as a coordinate on the ideal camera with the focal length of 1 and of eliminated radial distortion induced by the lens. In practice, however, $\mathbf{x}_{fp}$ is the projected position of $\hat{\mathbf{x}}_{fp} = (\hat{u}_{fp}, \hat{v}_{fp})$ in the real image, which is given transferring according to the known intrinsic camera parameters including focus, aspect, optical center and distortion parameters. This transformation from $\hat{\mathbf{x}}_{fp}$ to $\mathbf{x}_{fp}$ is applied only to coordinate calculation in the real-time stage regarding computational cost, whereas it will be considered precisely as a part of image dewarping in off-line stage, and omitted for simplicity in the rest of this paper.

Next, the error function used for 3D reconstruction is described. In general, the projected position $\mathbf{x}_{fp}$ of $\mathbf{S}_p$ to the *f*-th subimage frame does not coincide with the actually detected position $\mathbf{x}'_{fp} = (u'_{fp}, v'_{fp})$, due to errors in feature detection, extrinsic camera parameters and 3D feature positions. In this paper, the squared error E_{fp} is defined as an error function for the feature p in the *f*-th frame as follows:

$$E_{fp} = \left| \mathbf{x}_{fp} - \mathbf{x}'_{fp} \right|^2. \quad (2)$$

2.2. Real-time Stage for Image Acquisition

In the real-time stage, extrinsic camera parameters are estimated in real-time to show some information that helps users in image capturing. First, a user may very roughly set the image plane of the camera parallel to the target paper. Note that this setting is not laborious for users because it is done only for the first frame. This setting is used to calculate initial values for the real-time camera parameter estimation. The user then starts image capturing with free camera motion.

As shown in Figure 2, the real-time stage is constructed of two iterative processes for each frame. First, the extrinsic camera parameter is estimated by tracking features (a). A preview image of generating a mosaic image and instruction for controlling camera motion speed are then shown in the display on the mobile computer and updated every frame (b). The following describes each process of the real-time stage.

Step (a): Feature tracking and camera parameter estimation.

An iterative process to estimate the extrinsic camera parameters and 3D position $\mathbf{S}_p$ of each feature point is described. This process is an extension of the structure-from-motion method in [13].

In the first frame, assuming that the image plane in the first frame is approximately parallel to the target, rotation and translation components in $\mathbf{M}_f$ are set to an identity matrix and 0, respectively. For each feature point detected in the first frame, its 3D position $\mathbf{S}_p$ is set to $(u_{1p}, v_{1p}, 1)$, based on the same assumption. Note that these are only initial values, which will be corrected in the refinement process (Figure 2 (d)).

In the succeeding frames ($f > 1$), $\mathbf{M}_f$ is determined by iterating the following steps until the last frame.

Feature point tracking: All the image features are tracked from the previous frame to the current frame by using a standard template matching with Harris corner detector [14]. The RANSAC approach [15] is employed for eliminating outliers.

Extrinsic camera parameter estimation: Extrinsic camera parameter $\mathbf{M}_f$ is estimated using the feature position (u'_{fp}, v'_{fp}) and its corresponding 3D position $\mathbf{S}_p = (x_p, y_p, z_p)$. Here, extrinsic camera parameters are obtained minimizing the error summation $\sum_p E_{fp}$ w.r.t. 6-DOF elements in $\mathbf{M}_f$ by the Levenberg-Marquadt method. For 3D position $\mathbf{S}_p$ of the feature p, the estimated value in the previous iteration is used.

3D feature position estimation: For every feature point p in the current frame, its 3D position $\mathbf{S}_p = (x_p, y_p, z_p)$ is refined by minimizing the error function $\sum_{i=1}^{f} E_{ip}$. In the case of the flat mode, z value of the $\mathbf{S}_p$ is fixed to a constant to improve the accuracy of estimated parameters.

Addition and deletion of feature points: In order to obtain accurate estimates of camera parameters, stable features should be selected. The set of features to be tracked is updated based on the feature reliability [13].

Iterating the above steps, extrinsic camera parameters $\mathbf{M}_f$ and 3D feature positions $\mathbf{S}_p$ are estimated.

Step (b): Preview image generation and instruction notification for user.

Figure 4 shows displayed information for a user in our system. Usually, the user watches the right side window in the real-time stage. If necessary, the user can

check the input subimages and estimated camera path on the left side windows. The preview image helps the user to confirm a lacking part of the target to be scanned.

For each frame, the captured subimage is resized to a lower resolution and stored to texture memory. Every stored texture is dewarped onto the mosaic image plane by texture mapping using Eq. (1). The boundary of the current frame image is colored in the preview so that the user can grasp what part of the target has currently been captured.

On the other hand, the instruction shown in the bottom of the right window in Figure 4 helps the user to control the speed of camera motion. In our system, there is appropriate speed for camera motion that is shown by an arrow mark in the speed gauge, and it is graded as three ranges; “too slow", “appropriate," and “too fast." Too fast camera motion causes a worse mosaic image quality because the error in camera parameter estimation increases due to tracking errors and shortage of tracking span of each feature. In contrast, too slow motion consumes computational resources redundantly because the number of images increases to capture the whole document. To obtain a better result, the user controls camera speed so as to keep the arrow mark being inside “appropriate" range. In turn, too much slanted camera posture, which is still tolerable for 3D reconstruction, can cause degradations such as irregular resolution and partial defocus blur in a mosaic image. The camera pose can also be checked in the left-bottom window.

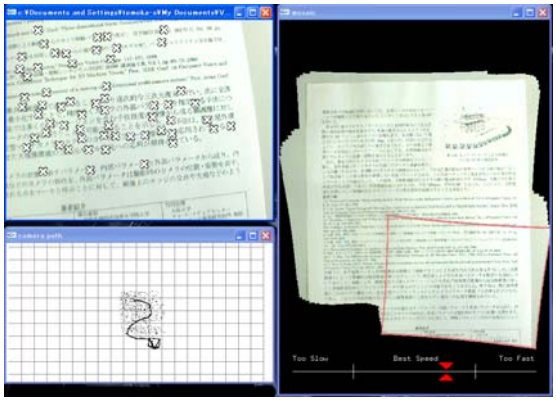


Figure 4. User Interface for real-time stage. Left-top: input image and tracking feature point. Left-bottom: temporarily estimated camera path and posture. Right: preview of ongoing mosaicing preview and speed indicator of camera motion.

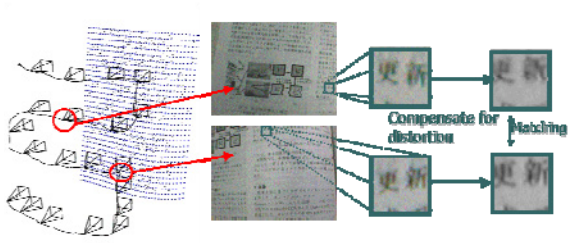


Figure 5. Registrations of reappearing features. (a) Camera path and 3D feature position. (b) Two temporally distant frames sharing part of scope. (c) Templates of the corresponding features. (d) Templates in which perspective distortion is corrected.

2.3. Off-line Stage for 3D Reconstruction Refinement and Mosaic Image Generation

This section describes the 3D reconstruction parameter refinement and shape regression, so as to minimize the summation of the error function all over the input sequence and feature misalignment from a parameterized columnar polynomial surface. First, as shown in Figure 2, feature points that reappear in temporally distant frames are homologized (c). Using these reappearing features, the 3D reconstruction is refined (d). Then, by fitting a curved surface to 3D feature points, the target shape is estimated (e). After a few iterations of steps (c) to (e), a dewarped mosaic image of the target is generated according to the target shape and the extrinsic camera parameters (f). Details of steps (c) to (f) are described below.

Step (c): Reappearing feature registration.
Due to the camera motion, the features relatively pass though the scope of the frames. Some features reappear in the sight after the previous flameout, as shown in Figure 5. This step detects these reappearing features, and distinct tracks belonging to the same reappearing feature are linked to form a single long track. This will give tighter constraints among the sequence of camera parameters in temporally distant frames, and thus makes it possible to reduce the cumulative errors caused by sequential camera scanning.

Reappearing features are detected by examining the similarity of the patterns among features belonging to distinct tracks. The problem here is that even if two patterns belong to the same feature on the target, they can have different appearance due to perspective distortion. To remove this effect, first, templates of all the features are projected to the fitted surface (described later). Next, feature pairs whose distance in 3D space is less than a given threshold are selected and tested with the normalized cross correlation function. If the corre-

lation is higher than a certain threshold, the feature pair is regarded as reappearing features. Note that in the flat mode, the shape of the target is simply assumed as a plane. When the system works in the curved mode, this step is skipped at an initial iteration of steps (c) to (e) because surface parameters have not been estimated at the first iteration.

Step (d): Refinement of 3D reconstruction.
Since the 3D reconstruction process described in Section 2.2 is an iterative process, its result is subject to cumulative errors. In this method, by introducing a bundle-adjustment framework [16], the extrinsic camera parameters and 3D feature positions are globally optimized so as to minimize the sum of re-projection errors, which is given as follows:

$$E_{all} = \sum_{f}\sum_{p} E_{fp} \ . \tag{3}$$

As for reappearing features, all the tracks belonging to the same reappearing feature are linked, and treated as a single track. This enables the extrinsic camera parameters and 3D feature positions to be optimized, maintaining consistency among temporally distinct frames.

Step (e): Target shape estimation by surface fitting.
In this step, assuming the curvature of the target lies along one direction, the target shape is estimated using 3D point clouds optimized in the previous step (d). Note that this stage is skipped in the case of the flat mode. First, the principal curvature direction is computed from the 3D point clouds as shown in Figure 6. Next, the 3D position of each feature point is projected to a plane perpendicular to the direction of minimum principal curvatures. Finally, a polynomial equation of variable order is fitted to the projected 2D coordinates, and the target shape is estimated.

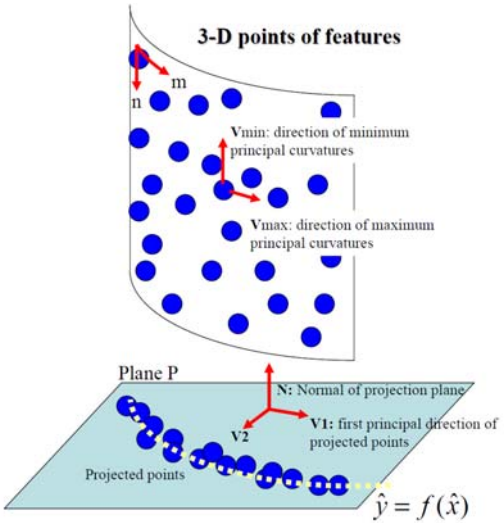


Figure 6. Polynomial surface regression for target shape approximation

Let us consider for each 3D point $\mathbf{S}_p$ a point cloud $\mathbf{R}_p$ that consists of feature points lying within a certain distance from $\mathbf{S}_p$. First, the directions of maximum and minimum curvatures are computed for each $\mathbf{R}_p$ using local quadratic surface fit. For a target whose curve lies along one direction, as assumed in this paper, the minimum principal curvature must be 0, and its direction must be the same for all the feature points. In practice, however, there exists some fluctuation in the directions of minimum curvature, due to the estimation errors. Thus, a voting method is applied to eliminate outliers and the dominant direction $\mathbf{V}_{\min} = (v_{mx}, v_{my}, v_{mz})^T$ of minimum principal curvatures for the whole target is determined.

Next, 3D position $\mathbf{S}_p$ for each feature point is projected to a plane whose normal vector $\mathbf{N}$ coincides with $\mathbf{V}_{\min}$ i.e. $P(x,y,z) = v_{mx}x + v_{my}y + v_{mz}z = 0$. The projected 2D coordinates $(\bar{x}_p, \bar{y}_p)$ of $\mathbf{S}_p$ is given as follows:

$$\begin{pmatrix} \bar{x}_p \\ \bar{y}_p \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{pmatrix} \mathbf{S}_p \ , \tag{4}$$

where $\mathbf{V}_1$ is a unit vector parallel to the principle axis of inertia of the projected 2D coordinates $(\bar{x}, \bar{y})$, and $\mathbf{V}_2$ is a unit vector that is perpendicular to $\mathbf{V}_1$ and $\mathbf{V}_{\min}$, i.e. $\mathbf{V}_2 = \mathbf{V}_1 \times \mathbf{N}$.

Finally, the target shape parameter $(a_1, \ldots, a_m)$ is obtained by fitting the following variable-order polynomial equation to the projected 2D coordinates $(\bar{x}, \bar{y})$.

$$\bar{y} = f(\bar{x}) = \sum_{i=1}^{m} a_i \bar{x}^i \ . \tag{5}$$

Here, the optimal order *m* in the above equation is automatically determined by using geometric AIC [17].

In the case where the target is composed of multiple curved surfaces, e.g. the thick bound book shown in Figure 1(a), the target is first divided with a line where the normal vector of the locally fitted quadratic surface varies discontinuously, and then the shape parameter is computed for each part of the target. The estimated shape is used for generating a dewarped mosaic image in the next process, as well as for removing the perspective distortion in the reappearing feature detection process (Figure 2(c)).

Step (f): Mosaic Image Generation.
Finally, a mosaic image is generated using extrinsic camera parameters and surface shape parameters. The super-resolution and shading correction are implemented here regarding coordinate transformation be-

low. Let us consider a 2D coordinate (m,n) on the dewarped mosaic image, as shown in Figure 6. The high-definite pixel value at (m,n) on the dewarped mosaic image is estimated from the intrinsic pixel values at all the corresponding coordinates (u_f, v_f) in the input subimage [11,12].

The relation between (m,n) and its corresponding 3D coordinate $(\bar{x}, f(\bar{x}), \bar{z})$ on the fitting surface is given as follows:

$$(m,n) = \left(\int_0^{\bar{x}} \sqrt{1 + \left(\frac{d}{dx} f(x) \right)^2}\, dx, \bar{z} \right). \quad (6)$$

The coordinate $(\bar{x}_p, f(\bar{x}_p), \bar{z})$ is transferred to the corresponding 2D coordinate (u_f, v_f) on the *f*-th image plane by the following equation.

$$\begin{pmatrix} au_f \\ av_f \\ a \end{pmatrix} = \mathbf{M}_f \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \\ \mathbf{N}^T \end{pmatrix} \begin{pmatrix} \bar{x} \\ f(\bar{x}) \\ \bar{z} \end{pmatrix}. \quad (7)$$

Assumption of the flat target shape (flat mode) simplifies these as planar perspective transformations.

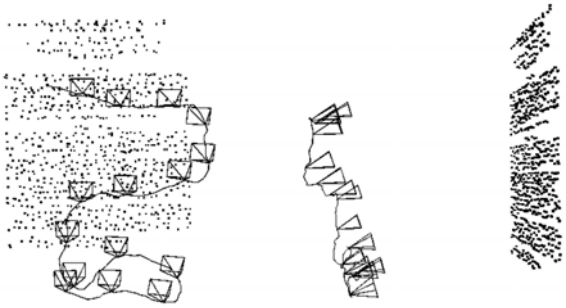
3. Experiments

We have developed a video mosaicing prototype system that consists of a mobile PC (Pentium-M 1. 2GHz, Memory 1GB) and a USB web-cam. The appearance of the prototype system has been shown in Figure 3. In our system, a user starts image capturing, referring the monitor of mobile computer. The real-time stage for image capturing is automatically finished when a video buffer becomes full. Experiments have been carried out for flat and curved documents. In these experiments, the intrinsic camera parameters are calibrated in advance by Tsai's method [18], and are fixed throughout image capturing. Note that in the current version of the prototype system, the curved mode is not implemented on the mobile system. Thus, the latter experiment for the curved surface is carried out by using a desktop PC (Xeon 3.2GHz, Memory 2GB), and the initial camera parameter estimation is processed offline after image capturing using an IEEE 1394 web-cam. Experimental results chiefly focused on 3D reconstruction are shown below.

3.1. Flat target

As shown in Figure 7, a flat target document is captured as 150 frame images of 640x480 pixels at 6 fps with initial camera parameter estimation. Image features tracked in the real-time stage are depicted with cross marks. Note that none of the input image planes are parallel to the target document. Figure 8 illustrates estimated extrinsic camera parameters and feature positions on the mosaic image plane. The curved line shows the estimated camera path and pyramids show the camera postures every 10 frames. The generated mosaic image is shown in Figure 9. We can confirm that the slanted effect is correctly removed in the final mosaic image by comparing the result with the homography-based result shown in Figure 1(b) where the same input images and the same feature trajectories with this experiment were used. Although the target paper is not perfectly flat, there are no apparent distortions in the final mosaic image. The performance of the

Figure 7. Input subimage samples and tracked features for flat target.



(a) Top view (b) Side view

Figure 8. Estimated extrinsic camera parameter and 3D feature distribution for flat document.

Figure 9. Reconstructed mosaic image.

prototype system for this sequence is as follows: 6 fps for image acquisition and initial parameter estimation with preview of mosaicing, 17 seconds for camera parameter refinement, and 39 seconds for generating the final mosaic image.

3.2. Curved target

An example of curved documents used in experiments is shown in Figure 10. The target is composed of 2 pages of curved surfaces: one page with texts and the other with pictures and figures. The target is captured with a web-cam as a video of 200 frames at 7.5 fps, and is used as an input to our system. Sampled frames of the captured video are shown in Figure 11. Tracked feature points are depicted with cross marks. The 3D reconstruction result obtained by the proposed method is shown in Figure 12. The curved line shows the camera path, pyramids show the camera postures every 10 frames, and the point cloud shows 3D positions of feature points. As can be seen, the point cloud coincides with the shape of the thick bound book. The shape of the target is estimated after 3 time iterations of reappearing feature detection and the surface fitting process. The estimated shape of the target is shown in Figure 13. In the proposed method, the optimal order of the polynomial surface fitted to the target is automatically determined by geometric AIC. In this experiment, the order is 5 and 4 for the left and right pages, respectively. The dewarped mosaic images before and after shadow removal are shown in Figure 14(a) and (b), respectively. As can be seen, the distortion on the target has been removed in the resultant image. The performance of the system on the desktop PC for this sequence is as follows: 27 seconds for initial 3D reconstruction, 71 seconds for camera parameter refinement, and 113 seconds for generating the final mosaic image. Some other mosaicing results for curved targets are shown in Figure 15.

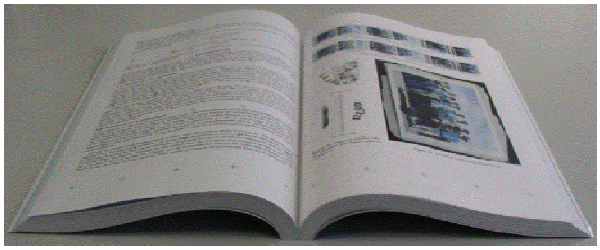


Figure 10. Curved target.

1st frame 66th frame 123rd frame 200th frame

Figure 11. Input subimage samples for curved target.

(a) Front view (b) Side view

Figure 12. Estimated extrinsic camera parameter and 3D feature distribution for curved document.

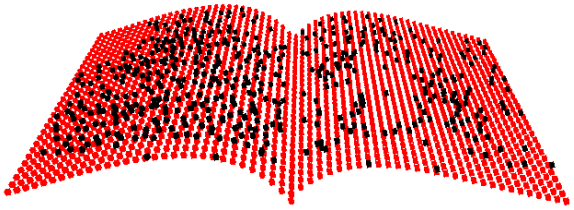
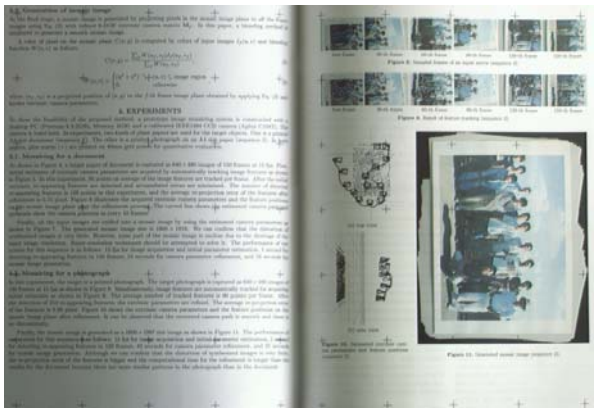
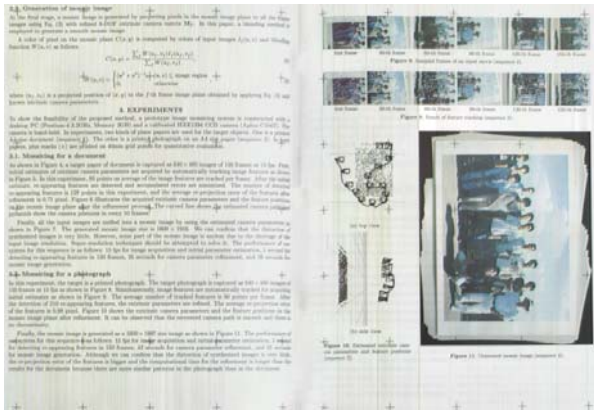


Figure 13. Estimated shape.

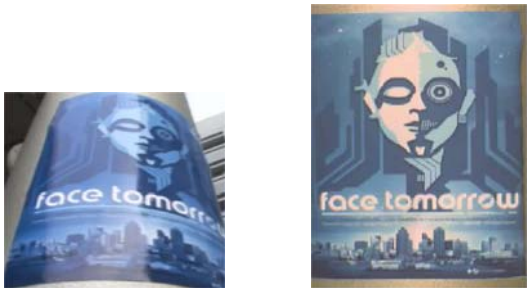


(a) Before shading removal



(b) After shading removal

Figure 14. Reconstructed mosaic image.



(a) Poster on a round pillar

(b) Bottle label

Figure 15. Mosaicing result examples. Left: target appearances. Right: reconstructed images.

4. Conclusion

A novel video mosaicing method for generating a high-resolution and distortion-free mosaic image for flat and curved documents has been proposed. With this method based on 3D reconstruction, the 6-DOF camera motion and the shape of the target document are estimated without any extra devices. In experiments, a prototype system of mobile video mosaicing has been developed and has been successfully demonstrated. For a flat target, a mosaic image without the slanted effect is successfully generated. Although the curved mode has not yet been implemented on the current version of mobile system, we have shown the off-line result for curved documents generated by desktop PC. In the curved mode, assuming the curve of the target lies along one direction, the shape model is fitted to the feature point cloud and a dewarped image without shadow is automatically generated. Our future work involves reducing the computational cost in the curved mode and improving the accuracy of 3D reconstruction around the inner margin of a book.

References

[1] K. Junga, K. I. Kim, and A. K. Jain, “Text information extraction in images and video: a survey,” *Pattern Recognition*, Vol. 37, No.5, 2004, pp. 977 – 997.

[2] A. Rosenfeld, D. Doermann, and D. DeMenthon, *Video Mining*, Kluwer, Massachusetts, 2003.

[3] D. Doermann, J. Liang, and H. Li, “Progress in camera-based document image analysis,” *Proc. Int. Conf. Document Analysis and Recognition*, 2003, pp. 606–616.

[4] R. Szeliski, “Image Mosaicing for Tele-Reality Applications," *Proc. IEEE Workshop Applications of Computer Vision*, pp. 230-236, 1994.

[5] S. Takeuchi, D. Shibuichi, N. Terashima, and H. Tominage, “Adaptive Resolution Image Acquisition Using Image Mosaicing Technique from Video Sequence," *Proc. Int. Conf. Image Processing*, vol. I, pp. 220-223, 2000.

[6] C. T. Hsu, T. H. Cheng, R. A. Beuker, and J. K. Hong, “Feature-based Video Mosaicing," *Proc. Int. Conf. Image Processing*, Vol. 11, pp. 887-890, 2000.

[7] M. Lhuillier, L. Quan, H. Shum, and H. T. Tsui, “Relief Mosaicing by Joint View Triangulation," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, vol. 19, pp. 785-790, 2001.

[8] P. T. McLauchlan, and A. Jaenicke, “Image Mosaicing Using Bundle Adjustment," *Image and Vision Computing*, 20, pp. 751-759, 2002.

[9] D. W. Kim and K. S. Hong, “Fast Global Registration for Image Mosaicing," *Proc. Int. Conf. Image Processing*, 11, pp. 295-298, 2003.

[10] P. Grattono, and M. Spertino, “A Mosaicing Approach for the Acquisition and Representation of 3D Painted Surfaces for Conservation and Restoration Purposes," *Machine Vision and Applications*, Vol. 15, No. 1, pp. 1-10, 2003.

[11] D. Capel, *Image Mosaicing and Super-resolution*, Springer-Verlag, London, 2004.

[12] Y. Weiss, “Deriving intrinsic images from image sequences,” *Proc. Int. Conf. Computer Vision*, pp. 68-75, 2001.

[13] T. Sato, M. Kanbara, N. Yokoya, and H. Takemura, “Dense 3D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera," *Int. J. Computer Vision*, Vol. 47, No. 1-3, pp. 119-129, 2002.

[14] C. Harris, and M. Stephens, “A Combined Corner and Edge Detector," *Proc. Alvey Vision Conf.*, pp. 147-151, 1988.

[15] MA Fischler, and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, Vol. 24, No. 6, pp. 381-395, 1981.

[16] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, “Bundle Adjustment - a Modern Synthesis," *Proc. Int. Workshop Vision Algorithms*, pp. 298-372, 1999.

[17] K. Kanatani, “Geometric Information Criterion for Model Selection,", *Int. J. Computer Vision*, Vol. 26, No. 3, pp. 171-189, 1998.

[18] R. Y. Tsai, “An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 364-374, 1986.