Video Mosaicing for Curved Documents by Structure from Motion

Tomokazu Sato[*] NAIST & NEC
Akihiko Iketani[†] NEC
Sei Ikeda[‡] NAIST
Masayuki Kanbara[§] NAIST & NEC
Noboru Nakajima[¶] NEC
Naokazu Yokoya[‖] NAIST & NEC

1 Introduction

This sketch describes a novel method for digitizing documents on curved surfaces by mosaing using a hand-held video camera. Generally, the video mosaicing techniques give a cheap and low-resolution web-cam device the capability of high resolution and wide angle capturing of a target with improved imaging quality. However, most of existing video mosaicing methods[Szeliski, R. 1994; Hsu, C.T. et al. 2000] are only for flat or very far targets. Although there are some video mosaicing methods which can deal with curved surfaces[Grattono, P. and Spertino, M. 2003], they require an active camera and a slit light projection device. In this sketch, a novel video mosaicing method based on structure from motion is proposed. The proposed method can generate a distortion-free image of curved document from a video captured by a single camera, and thus requires no additional devices nor a priori knowledge on the target. The assumptions here are that intrinsic camera parameters are fixed and calibrated in advance to correct lens distortion. For curved documents, the curve of the target must lie along one direction and its curvature changes smoothly along this direction.

2 Approach

In our method, a user should just take a video image of a curved target by moving a video camera. The system automatically estimates 3-D information from the motion of image features, and a still image without geometric distortion is then generated. Our video mosaicing method is constructed of three phases as follows.
Phase 1. 3-D reconstruction: 3-D positions of the feature points and 6-DOF extrinsic camera parameters in each frame are estimated by tracking image features in the video stream[Sato, T. et al. 2002].
Phase 2. Parameter refinement and surface fitting: The estimated camera parameters are globally optimized, and the shape of the target is approximated by fitting a polynomial of one variable to the point cloud. Reappeared features in video stream are also detected and utilized to increase the accuracy of 3-D reconstruction.
Phase 3. Mosaic image generation: An unwrapped mosaic image is generated using the approximated shape and the extrinsic camera parameters.

3 Experiments

We have developed a prototype video mosaicing system which mainly consists of a laptop PC (IBM ThinkPad X31 1.6GHz, Memory 1GB) and an IEEE1394 CCD camera (Aplux C104T, VGA). Experiments have been carried out for a thick bound book with curved pages (Figure 1(a)) and a label on a wine bottle (Figure 2(a)). Each target is captured with the web-cam as a 5 fps VGA sized video (250 frames for the book, 100 frames for the wine label). Figures 1(b) and 2(b) show sampled frames of each input video stream. Cross marks in the figure indicate positions of tracked image features detected by Harris operator. 3-D reconstruction results are shown in Figures 1(c) and 2(c), where the curved line shows the camera path, pyramids show camera postures in every 10 frames, and black points are 3-D positions of feature points. The set of grid points in each figure represents a surface fitted to the 3-D point cloud. The unwrapped mosaic images are shown in Figures 1(d) and 2(d). As can be seen, the distortion on the target has successfully been removed in the resultant image.

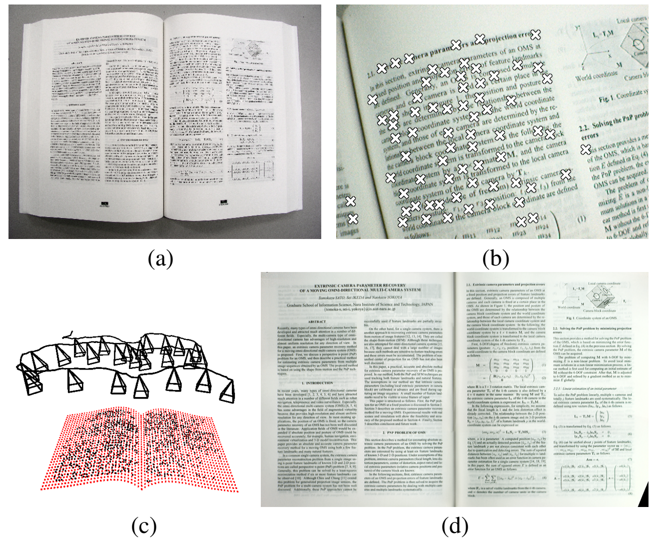

Figure 1: Experimental results for thick bound book.

Figure 2: Experimental results for wine label.

[*] e-mail: tomoka-s@is.naist.jp
[†] e-mail: iketani@cp.jp.nec.com
[‡] e-mail: sei-i@is.naist.jp
[§] e-mail: kanbara@is.naist.jp
[¶] e-mail: n-nakajima@ay.jp.nec.com
[‖] e-mail: yokoya@is.naist.jp

References

Grattono, P., and Spertino, M. 2003. A Mosaicing Approach for the Acquisition and Representation of 3D Painted Surfaces for Conservation and Restoration purpose. *Machine Vision and Applications*, vol.15, no.1, 1–10.

Hsu, C.T., Cheng, T.H., Beuker, R.A., and Hong, J.K. 2000. Feature-based Video Mosaicing. *Proc. ICIP*, vol.II, 887–890.

Sato, T., Kanbara, M., Yokoya, N., and Takemura, H. 2002. Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera. *IJCV*, vol.47, no.1-3, 119–129.

Szeliski, R. 1994. Image Mosaicing for Tele-Reality Applications. *Proc. IEEE Workshop on Applications of Computer Vision*, 230–236.