Proc. ISMAR 2006 (IEEE/ACM Int'l Symp. on Mixed and Augmented Reality), Santa Barbara, CA, Oct. 22–25, 2006, pp. 241–242

Viewpoint Stabilization for Live Collaborative Video Augmentations

Taehee Lee, Tobias Höllerer
Four Eyes Laboratory
Department of Computer Science
University of California
Santa Barbara, California 93106 USA

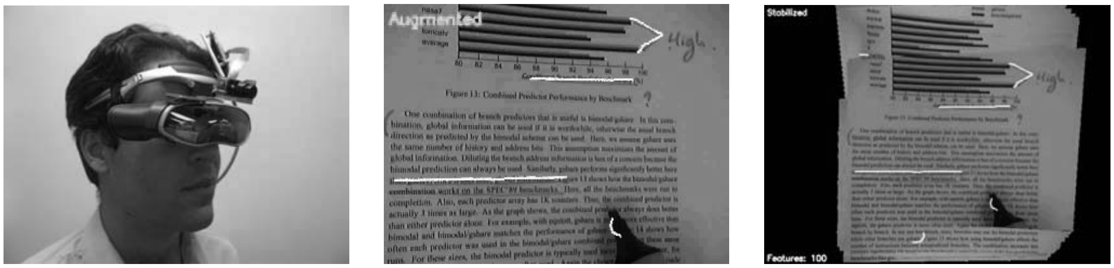

Figure 1: Our application scenario: a) AR user w. head-worn camera, looking at shared working plane, b) AR user's view including remote annotations, c) Remote user's stabilized view

Abstract

We present a method for stabilizing live video from a moving camera for the purpose of a tele-meeting, in which a participant with an AR view onto a shared canvas collaborates with a remote user. The AR view is established without markers and using no other tracking equipment than a head-worn camera. The remote user is allowed to directly annotate the local user's view in real time on a desktop or tablet PC. The planar homographies between the reference frame and the other following frames are maintained. In effect, both the local and remote participants can annotate the physical meeting space, the local AR user through physical interaction, the remote user through our stabilized video. When tracking is lost, the remote user can still continue annotating on a frozen video frame. We tested several small demo applications with this new form of transient AR collaboration that can be established easily, on a per need basis, and without complicated equipment or calibration requirements.

1 Introduction

Augmented Reality is a very attractive paradigm for collaboration, in which local and remote participants can share a physical space and annotate it [2][3][10]. One practical hindrance is the high setup cost resulting from the need to either instrument the user or the environment with tracking systems and/or fiducials for marker-based camera tracking. Markerless vision-based tracking in unprepared environments [4][7][9][12] has taken great strides towards becoming an important complement and increasingly viable alternative to sensor-based tracking approaches, but real-time and robustness demands still cause problems for general scenes.

We are interested in specific AR application scenarios, for which current state-of-the-art real time computer vision approaches are useful, practical, and sufficient. This poster describes a simple scenario for collaboration between an AR user and a remote user interacting with the AR user's video feed (see Figure 1). The head-worn camera is the only tracking equipment used. If we make the additional assumption that the local user be looking at a planar collaboration space (e.g., a table top or whiteboard), we do not need to solve the complete localization and mapping problem [4] in order to get useful results. The most important consideration for this application is to create a stable view onto the collaboration surface for the remote user to annotate. Dynamic scenes, affected by movements of the worn camera or objects in a view, make it very hard in general for a remote user to annotate live videos. Using planar homographies, we build up a large stabilized and rectified [8][6][11] view of the work environment from the AR user's camera stream (Figure 1c), in which all scene parts outside of the current (smaller) camera view are retained from the most recent camera frame that included them. The quality of the homography is monitored so that we know when tracking is lost (reasons include the user looking away from the working surface or temporary occlusion). In that case, we freeze the last known good complete picture of the work environment for the remote user until we can re-establish the homography in the local AR user's view, either automatically, or, in the worst case, by a very simple matching action prompted and guided by screen messages and overlays.

We present two simple illustrations of our AR collaboration setup in use: a Tic-Tac-Toe game between the local and remote participant and the collaborative editing of a paper document.

2 System Description

A local AR user, wearing a head mounted display and camera, collaborates with other participant(s) in the local physical workspace, which is augmented by everybody's annotations. Other participants without AR equipment use desktop or tablet PCs remotely with our viewpoint stabilization software.

2.1. Viewpoint Stabilization

The remote users' view is coupled with the AR user's camera. To instantiate collaboration, the AR user looks at the target workspace. The remote users can initiate "stabilization mode", which relies on tracking optical flow features in the AR user's video stream; the initial view is the reference frame. Using successfully tracked features, we calculate planar homographies between consecutive frames. Then the video frames from the AR user are projected to the reference frame by the corresponding homographies, stabilizing the remote user's view. As an inverse, the homography also enables the remote user's drawings to correctly annotate the AR user's view, as shown in Figure 1.

1-4244-0651-X/06/$20.00 ©2006 IEEE

2.1.1. Homography Calculation

By considering the reliability of each homography, we can estimate the homography between the initial reference frame and the current frame as below:

$$H_{i,0} = H_{i,j} H_{j,0}$$

where *i* is the current frame index and *j* is the most recent frame index with a valid homography $H_{j,j-1}$. This sequential computation of homography ignores invalid homographies in between frame *j* and *i*. Also, this makes it possible for the AR user to reset the lost homography. Being guided by frame *j* added to the view in a ghosted screen-stabilized visual overlay, the AR user can move the camera near to the last known good view so that the system may automatically pick up the homography or choose another viewpoint manually.

2.1.2. Feature Management and Error Recovery

The point correspondences are derived from tracking features throughout the video. For each frame, 100 Harris corners [13] are detected, including previously reliable features. Newly generated features are matched to the nearest feature of the previous frame with 3 pixel distance threshold. Additional new features are added in order to maintain up to 100 features. Among these features, only the inliers from the RANSAC algorithm [5] are used to compute the homography. The number of inliers also determines the reliability of the homography, with an empirically chosen threshold of 40.

3 Results

We have tested our prototype system with some simple collaborative AR applications, including remote Tic-Tac-Toe and collaborative paper editing.

Remote Tic-Tac-Toe is played as follows: 1) AR user draws grid lines on whiteboard or sheet of paper, 2) remote user starts stabilizing the view, 3) either of them starts the game, and 4) AR user places pieces by physically drawing them and views the opponent's responses as AR overlays through the HMD. The remote user annotates the stabilized video using pen or mouse input. A screenshot of the remote user's view is shown in Figure 2.

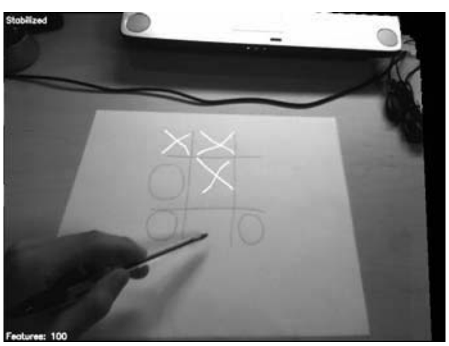

Figure 2: Tic-Tac-Toe on our stabilized video

Collaborative Paper Editing. In a second example, the local AR user and the remote user collaborate on editing a physical paper copy in the AR user's workspace. Again, both participants can add annotations to the document, physically in the AR user's case and virtually on the stabilized video by the remote user (cf. Figure 1). Physical annotations could be captured, e.g., using Anoto technology [1] so that the local and remote interactions could be fused into a final digital document.

4 Future Work

We are currently evaluating our technique in the overall framework of a transpacific teleconferencing project, including tests of user acceptance for different methods of reinstating homographies. We are working towards further improving the tracking robustness, considering frame-to-frame constraints on feature patterns in the scene. We also aim to bring up the frame rate by both increased computational efficiency and use of hardware acceleration, e.g. a GPU implementation of the warping operation. Finally, we are interested in tackling applications involving more general 3D geometries, which leads us into the direction of SLAM algorithms [4].

Acknowledgement

This work was supported by the Korea Science and Engineering Foundation Grant (No.2005-215-D00316).

References

[1] Anoto AB, http://www.anotofunctionality.com/, accessed 08/04/2006.
[2] M. Billinghurst, S. Weghorst, and T.A. Furness III. 1998. Shared Space: An Augmented Reality Approach for Computer Supported Collaborative Work, Virtual Reality, Springer, Vol. 3, No. 1, pp. 25–36.
[3] A. Butz, T. Höllerer, S. Feiner, B. MacIntyre, and C. Beshers. 1999. Enveloping users and computers in a collaborative 3D augmented reality. In *Proc. IWAR '99*, San Francisco, CA, Oct. 20-21 1999, pp. 35–44.
[4] A.J. Davison, W.W. Mayol, and D.W. Murray. 2003. Real-Time Localisation and Mapping with Wearable Active Vision. In *Proc. ISMAR 2003*, pp. 18–27.
[5] M.A. Fischler and R.C. Bolles, 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (Jun. 1981), 381–395.
[6] J.B. Hayet, J. Piater, and J. Verly. 2004. Robust incremental rectification of sports video sequences. British Machine Vision Conference (BMVC'04). pp. 687–696.
[7] M.I. Lourakis and A.A. Argyros, 2005. Efficient, causal camera tracking in unprepared environments. *Computer Vision and Image Understanding* 99, 2 (Aug. '05), 259–290.
[8] Y. Matsushita, E. Ofek, X. Tang and H-Y. Shum. Full-frame Video Stabilization. IEEE CVPR 2005, vol. 1, pp. 50–57.
[9] G. Simon, A. Fitzgibbon, and A. Zisserman, 2000. Markerless Tracking using Planar Structures in the Scene. *Proc. ISAR '00*, pp. 120–128.
[10] N. Takao, J. Shi, and S. Baker. Tele-Graffiti: A Camera-Projector Based Remote Sketching System with Hand-Based User Interface and Automatic Session Summarization. *Int. Journal of Computer Vision*, 53(2), July 2003, pp. 115–133.
[11] D. Koppel, Y.-F. Wang, H. Lee, 2005. Robust and Real-Time Image Stabilization and Rectification. *IEEE WACV/MOTION '05*, Breckenridge, CO, pp. 350–355.
[12] K. Xu, S. J. D. Prince, A. D. Cheok, Y. Qiu, and K. G. Kumar, 2003. Visual Registration for Unprepared Augmented Reality Environments, *Journal of Personal and Ubiquitous Computing*, Vol. 7, pp. 287–298.
[13] C. J. Harris and M. Stephens, 1988. A combined corner and edge detector. In *Alvey Vision Conf.*, pp. 147–151.