ISMR 2001 in Yokohama, Japan, pp.184-185 (2001)

3-D annotation of images captured from a wearer's camera based on object recognition

Takashi Okuma, Takeshi Kurata, Katsuhiko Sakaue
Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba, Ibaraki, 305-8568, JAPAN
Email: {okuma, kurata, sakaue}@etl.go.jp

Abstract

This paper describes a method for annotating the real scene image sequences with 3-D virtual objects based on image recognition. In order to annotate the real scene with 3-D virtual objects, we have to know not only where is user seeing from, but also what is user seeing. Therefore, we are trying to combine simple image recognition method with a vision based video see-through mixed reality system. In this presentation, it is proposed the method for starting estimation of camera position and orientation using image recognition result. Our demonstration system is a 3-D poster of this research that is a real paper poster annotated with 3-D virtual objects using proposed method.

Keywords: 3-D annotation, wearable system, real-time system

1 Introduction

Mixed Reality (MR) environment[1] has attracted much attention, because it enables us to add information on the real environment [2,3]. Some applications of MR require registration between virtual objects and the real environment. Most of these applications assume that the unique reference coordinate system is set in the real environment[4-7]. Virtual objects are described using the reference coordinate values. In this case, virtual objects are related to positions of the real objects, not to real objects. Therefore, when the annotated objects are moved, the position of the annotated object has to be tracked. It is very difficult to track all of annotated objects in the real environment. In order to annotate real objects, the reference coordinate system should be set on each object and the system has to acquire the reference coordinate values of the user's viewing position.

3-D position sensors, such as magnetic and ultra-sonic sensor, consist of two parts. One is the transmitter of magnetic field or ultra-sound, the other is the receiver of them, such as coils or microphones. These sensors acquire the relative positions of receivers to the position of the transmitter. Therefore, in order to annotate the real objects, all of the annotated real objects have to be set these receivers.

On the other hand, vision-based tracking method estimates the viewing position from the 2-D positions on the image plane (the screen coordinate values) of known feature points. This method doesn't need to set anything to the annotated real objects. There is potentially no limitation in its measuring area. However, this method needs to find annotated objects and to track the known feature points only from images. It is difficult to acquire the screen coordinate values of the known feature points accurately because of noise, change of lighting condition and so on. It is also difficult to track the feature points stably because they are often occluded or go out of the view. Therefore, traditional vision-based tracking method uses fiducial markers or matrix code to recognize the annotated object and to track known feature points [4-7]. However, these methods need to set fiducial markers to the annotated objects. Therefore, we try to use color information to recognize annotated objects that are not set fiducial markers on. Our method also uses a vision-based tracking. An inertial motion sensor is used to make a vision-based tracking robust. Inertial motion sensors can acquire only its orientation, but they don't have a pair of transmitter and receiver.

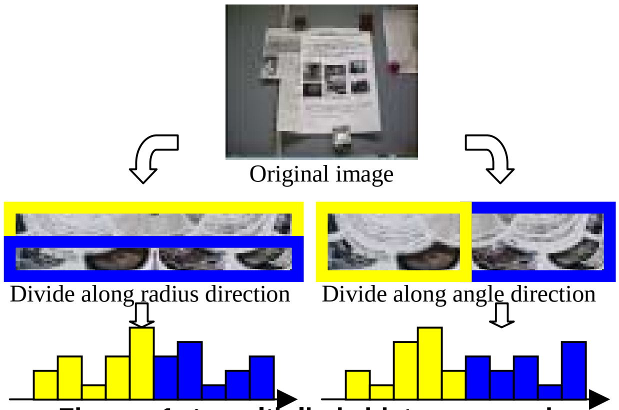


Figure 1 : multiplied histogram using polar coordinate transform

2 Multiplied color histogram

A lot of methods for image matching have been proposed. We focus on the color histogram matching[8]. Color histogram matching is stable and their calculation cost is low. However, its accuracy of matching is not so higher than that of the template matching, because color histogram matching is not to use the positional relation of the pixel. Therefore, our method divide scene image into plural parts and calculate histograms of each part. Then, these histograms are used for comparison of scene images.

In this paper, these histograms can be defined as "multiplied histogram". In order to create a multiplied histogram that isn't affected by the rotation of images, the polar coordinate transformed images are divided along the direction of radius (See upper side of Figure 1). Multiplied histogram that is created such as lower side of Figure 1 can be used to acquire the rotation angle between two images capture the same area.

3 Annotation with 3D Virtual Objects based on scene image recognition

Our method synthesis images by the following step:

(1) A list that has sets of annotated scene information is prepared. A set of annotated scene information contains the scene image, 3-D positions of some feature points and multiplied color histogram.

(2) Scene images captured from user's viewpoint (viewing scene image) are compared with annotated scene images using multiplied histogram matching. Similarity is calculated at each position and each rotation angle of viewing scene image. When the similarity is higher than a threshold, the system recognizes that users are seeing the annotated real object.

(3) Feature points registered in an annotated scene image are searched from viewing scene image using matched position and rotation angle. Screen coordinate values of known feature points are acquired. Found known feature points are tracked using Lucas-Kanade's feature points tracking method[9].

(4) Candidates of the model-view matrix are calculated with methods for a Perspective n-Points problem[10,11]. Then the model-view matrix is selected from these candidates using reprojection of the feature points and orientation data from inertial sensor.

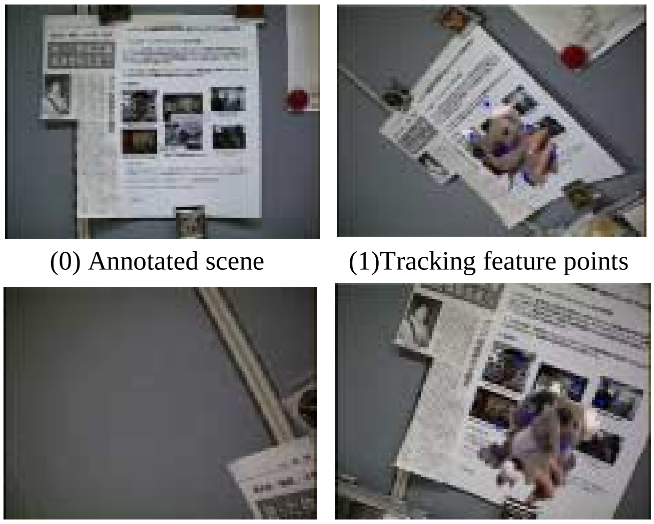
(5) The system render 3-D interactive virtual objects using estimated model-view matrix. Rendered virtual objects are overlaid on viewing scene image.

4 Experiment and Discussion

The proposed algorithms are implemented on a PC. It was found that an experimental system that is based on the proposed method can merge 3-D virtual objects into a 3-D real environment at right position. In Figure 2, the paper poster is annotated with virtual stuffed toy. When a user sees the annotated poster, the system recognizes this scene as annotated poster and virtual toy is displayed. This system can synthesize virtual and real images at 10 frames per second. In this system, 64 colors are used to make multiplied color histograms. The Bayer's dithering algorithm[12] is used to reduce the number of colors.

5 Conclusion

We have proposed a method for annotating the real object with 3-D virtual objects. The proposed method can recognize the annotated scene images, find and tracks natural feature points using a multiplied color histogram matching. The model-view matrix calculation method



(0) Annotated scene (1)Tracking feature points

(2)Going out of view (3)Recognize again

Figure 2: execution example (3-D poster)

consists of three algorithms for PnP problems. The experimental system using the proposed method works at 10 frames per second. This system shows the potential ability of interactive 3-D annotation with virtual objects.

In the future work, we should make error analyses of the proposed method and improve in accuracy and robustness. We should also investigate a method that can recognize the annotated scene image of different scale.

Reference

[1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual display," IEICE Trans. on Information and Systems, vol. E77-D, No.12, pp.1321-1329, 1994.

[2] Y. Ohta and H. Tamura Eds. , Mixed Reality – Merging Real and Virtual Worlds, Ohmsha & Springer-Verlag, 1999.

[3] R. T. Azuma, "A survey of augmented reality," *Presence*, vol.6, No.4, pp.355-385, 1997.

[4] A. State, G. Hirota, D. T. Chen, W. F. Garrett, M. A. Livingston, " Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking," SIGGRAPH96, pp.429-438, 1996.

[5] J. Rekimoto, "Matrix: A Realtime Object Identification and Registration Method for Augmented Reality", APCHI'98

[6] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto and K. Tachibana, "Virtual Object Manipulation on a Table-Top AR Environment," Proc. ISAR2000, pp.111-119, 2000.

[7] T. Okuma, T. Kurata and K. Sakaue, "Real-Time Camera Parameter Estimation from Images for a Wearable Vision System," Proc. MVA2000, pp.83-86, 2000.

[8] M. J. Swain, and D. H. Ballard, "Color Indexing," International Journal of Computer Vision, Vol.7, No.1, pp.11-32, 1991

[9] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," Proc. DARPA Image Understanding Workshop, pp. 121-130, 1981.

[10] K. Deguchi, "A Unified Approach to PnP Camera Calibration Problem by Projective Geometry," Proc. IPSJ Symposium, Vol.90, No. 20, pp.41-50, 1990. (in Japanese.)

[11] R. M. Haralick, C.-N. Lee and K. Ottenberg, "Analysis and Solutions of the Three Point Perspective Pose Estimation Problem," *Proc. CVPR `91*, pp.592-598, 1991.

[12] B. E. Bayer, "An Optimum Method for Two-Level Rendition of Continuous-Tone Pictures, Proc. Of IEEE 1973 Int. Conf. On Communications, Vol.1, pp.26-11—26-15, 1973.