Improvement of panorama-based annotation overlay using omnidirectional vision and inertial sensors

Masakatsu Kourogi† Takeshi Kurata‡ Katsuhiko Sakaue‡ Yoichi Muraoka†

†Graduate School of Science and Engineering, Waseda University,
3-4-1 Okubo, Shinjuku, Tokyo, 169-8555, JAPAN

Phone: +81-3-3209-5198, E-mail: {kourogi,muraoka}@muraoka.info.waseda.ac.jp

‡Eletrotechnical Laboratory, 1-1-4 Umezono, Tsukuba, Ibaraki, 305-8568, JAPAN
Phone: +81-298-61-5789, E-mail: {kurata,sakaue}@etl.go.jp

Abstract

Annotation overlay on live video frames is an essential feature of augmented reality (AR), and is a well-suited application for wearable computers. A novel method of annotation overlay and its real-time implementation is presented. This method uses a set of panoramic images captured by omnidirectional vision at various points of environment and annotations attached on the images. The method overlays the annotations according to the image alignment between the input frames and the panoramic images. It uses inertial sensors not only to produce robust results of image registration but also to improve processing throughput and delay.

1 Introduction

Displaying input video frames overlaid with annotation about a scene is an essential feature of augmented reality (AR) and one of the most potential applications for wearable computers[1].

We describe a fast and robust method for annotation overlay without using fiducials.

In our previous work[4], we proposed a method for annotation overlay which used a set of panoramic images captured by panning a camera at various points of environment and annotations manually attached to them, as a prior knowledge. The method overlaid the annotations according to the image alignment between the input video frame and the panoramic images.

The method, however, has two limitations. (1) It is time-consuming to capture panoramic images and manually place annotations on every image. (The method requires roughly one panoramic image per 1 m^2 of the usual office environment.) (2) The method will not work if the input video frame is mostly covered by objects that are not present in the panoramic images, or if the frame contains mostly featureless scene. Image alignment will fail in both cases.

To overcome these limitations, we improved the method in two ways. (1) We developed a method for finding correspondences between panoramic images. By using this method, when annotations are placed by hand on one panoramic image, they can be automatically put in the appropriate positions on other corresponding panoramic images. We also use omnidirectional vision[6] and create 360 degree panoramic images by mapping to the cylindrical coordinate system, so that panoramic images can be easily captured. (2) Previous works[2][3] have used inertial sensors combined with vision-methods that tracked the natural features on a frame-to-frame basis, or that estimated camera posture by tracking fiducials carefully placed in the environment. Our method uses three degree of freedom (3-DOF) inertial sensors to estimate image alignment parameters between successive frames along with the panorama-based method that estimates frame-to-panorama image alignment parameters to cover a large-scale environment without using fiducials. The estimated parameters can be used as an alternative if the panorama-based image alignment fails or is in progress. Since image alignment takes more processing time to estimate the parameters than the inertial sensors does, throughput and delays of annotation overlay **should** be greatly improved.

2 Creating a prior knowledge

Our proposed method requires three types of a prior knowledge about the environment.

1. A set of panoramic images captured at various points of the environment.
2. Annotations with their positions on the panoramic images.
3. The correspondence between the panoramic images.

It is time-consuming to manually create this knowledge. We reduce this burden in two ways.

2.1 Omnidirectional vision

Panoramic images can be created by one of two ways: by stitching multiple frames into one panoramic image (image mosaicing), or by projecting an omnidirectional image captured by HyperOmni Vision[6] to a cylindrical plane. We took the latter approach because it creates geometrically precise panoramic images more easily than the former.

2.2 Automatic placement of annotations on panoramic images

Panoramic images captured close to each other share a part of the same scene taken from slightly different viewpoints. Annotations manually placed on the shared scene in one panoramic image should be automatically placed at the corresponding positions on the other panoramic images.

Using alignment parameters between panoramic images obtained by a similar method to frame-to-panorama alignment, manually placed annotations can be automatically propagated to other panoramic images.

3 Improvement of image registration with inertial sensors

This method estimates the affine transform parameters, $A_{t \to p}$, for image alignment between an input

0-7695-0795-6/00 $10.00 © 2000 IEEE

frame, I_t, at time t and a panoramic image, I_P. However, the image alignment will fail if the frame contains featureless scenes, or if objects that are not present in the panoramic images account for the majority of the scene. To overcome such situations, we use inertial sensors.

The inertial sensors, fixed on a camera, can measure the camera's rotational angles around three axes. Let the yaw, pitch, and roll of the rotational angles between time $t+1$ and t be (ϕ, θ, ψ) , the rotation matrix be $\boldsymbol{R}(\phi, \theta, \psi)$, r_{ij} be an element of the matrix and f be the focal length of the camera. By setting $r_{31} = 0, r_{32} = 0$, and $r_{33} = 1$ to approximate by affine transform, the following transform matrix for image alignment between I_{t+1} and I_t is obtained.

$$\boldsymbol{A}_{t+1 \to t} = \begin{bmatrix} r_{11} & r_{21} & r_{31} f \\ r_{12} & r_{22} & r_{32} f \\ 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

Affine transform matrix $\boldsymbol{A}_{t+1 \to P}$ for image alignment between input frame I_{t+1} at time $t+1$ and the panoramic image can be computed by using

$$\boldsymbol{A}_{t+1 \to P} = \boldsymbol{A}_{t \to P} \boldsymbol{A}_{t+1 \to t}. \quad (2)$$

The method uses the affine parameters obtained by equation (2) as an alternative if the image-based alignment fails. If the mean squares error of the image brightness between the panorama and the frame is above a threshold, the image-based alignment is regarded as failed.

The affine parameters will also be used while the alignment is still in progress to improve the processing delays and throughput.

4 Experimental results

We implemented the proposed method on our developing *wearable vision system*[4][5], consisting of a wearable computer (OS: Windows 98, CPU: Mobile PentiumIII-500MHz) equipped with a head-worn display (MicroOptical, Clip-on display), a CCD camera (Toshiba, IK-SM43H), a 3-DOF motion sensor with gyro-sensors, accelerometers and compasses (Tokin, MS3D-U7) which claims to have resolution of ± 1.0 degree at 125Hz and a wireless LAN card complied with 11-Mbps IEEE 802.11b, and a remote PC cluster consisting of 5 PCs (OS: Linux-2.2.14 SMP supported, CPU: Dual PentiumIII-500MHz) as shown in Figure 1.

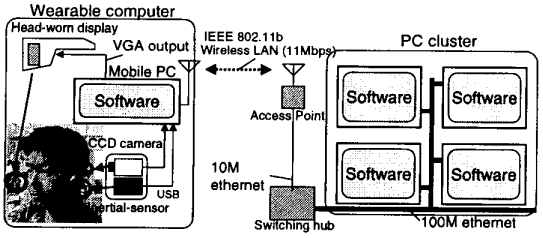

Figure 1: Diagram of our wearable vision system

The results of testing running under various conditions showed that by automatically switching the reference from image alignment to the inertial sensors output, this method can overlay the video frames with annotations even though the image-based alignment failed. Examples of output frames with annotations are shown in Figures 2 and 3. They showed that the method worked properly for a scene mostly covered by a held book and that it worked robustly for a featureless scene, respectively. Error of the annotation overlay was kept below 15 pixels by both the vision-only method and the hybrid method. The processing throughput and delay were improved from 6–8 frame/sec and 800–1000 msec, respectively, by the vision-only method to 12–15 frame/sec and 100–150 msec, respectively, by the hybrid method.

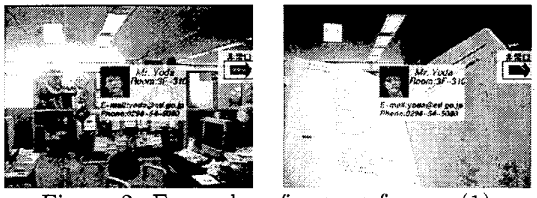

Figure 2: Examples of output frames (1)

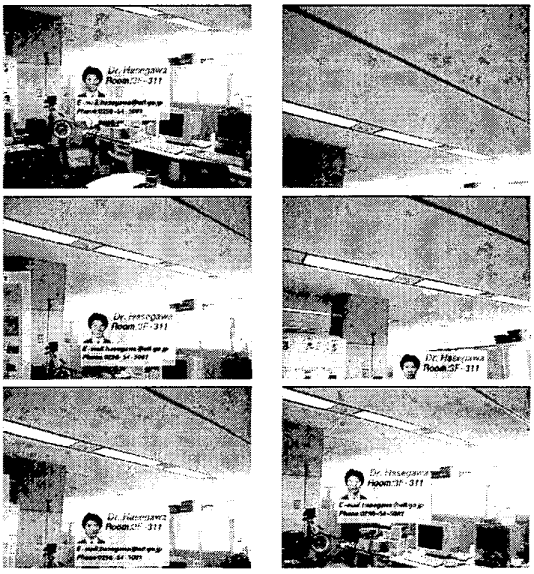

Figure 3: Examples of output frames (2)

References

[1] T. Starner et al., "Augmented reality through wearable computing," in *Presence*: Teleoperators and Virtual Environments, vol. 6, no. 4, pp. 386–398, 1997.

[2] S. You, U. Neumann and R. Azuma, "Hybrid Inertial and Vision Tracking for Augmented Reality Registration," *in Proc. of IEEE VR'99*, pp. 260–267, 1999.

[3] Y. Yokokoji et al., "Accurate Image Overlay on Video See-Through HMDs Using Vision and Accelerometers," in *Proc. of IEEE VR'2000*, pp. 247–254, 2000.

[4] M. Kourogi et al., "A real-time panorama-based technique for annotation overlay on video frames," in *Proc. of ICME'2000*, TA. 2.05, 2000.

[5] http://www.etl.go.jp/~kurata/demo/.

[6] K. Yamazawa et al., "Omnidirectional imaging with hyperboloidal projection," in IEEE Int. Conf. Intelligent Robots and Systems, no. 2, pp. 1029–1034, 1993.