

## VizWear-3D: A Wearable 3-D Annotation System Based on 3-D Object Tracking using a Condensation Algorithm

Takashi OKUMA

Takeshi KURATA

Katsuhiko SAKAUE

*Intelligent Systems Institute,*

*National Institute of Advanced Industrial Science and Technology (AIST)*

[takashi-okuma@aist.go.jp](mailto:takashi-okuma@aist.go.jp)

[t.kurata@aist.go.jp](mailto:t.kurata@aist.go.jp)

[k.sakaue@aist.go.jp](mailto:k.sakaue@aist.go.jp)

### Abstract

*We have developed a method for recognizing and tracking known three-dimensional objects in scene images captured from a user's viewpoint. The method uses multiple color histogram matching and condensation tracking. We implemented a prototype 3-D online manual system based on our method.*

### 1. Introduction

Our research group is developing wearable computing systems, interfaces, and applications based on computer vision techniques. We call them collectively *VizWear* [1][2]. These systems have been developed so that we can evaluate their advantages, disadvantages, and applications in situations when the user always wears the system with a camera. We have also investigated an augmented reality (AR) system for wearable computing, because AR is a good way to show information intuitively. Therefore AR on a wearable system is very useful for some applications, such as instruction systems that show handling of real objects in three dimensions. To display 3-D information effectively, the system should support 3-D object recognition, tracking the position and orientation (6-DOF parameters) of a 3-D object, and annotation with 3-D virtual objects.

The use of 3-D position sensors, such as magnetic and ultrasonic sensors, that are used in virtual reality systems are also widely used in augmented reality systems [4]. Such devices have the limited range because they need an emitter and a receiver to acquire 6-DOF parameters. Therefore, these devices are not suitable for wearable design. Vision-based tracking systems have been investigated with the aim of overcoming the range, accuracy, and practical problems associated with tracking systems [5][6]. They estimate 6-DOF parameters by analyzing images captured from the user's viewpoint (viewed-scene images). They use methods for

determining the position and orientation of a camera according to the 2-D positions of  $N$  known points in an image. This determination problem is called a Perspective  $n$ -Points (PnP) problem. To use methods for solving PnP problems, a vision-based tracking system faces difficulty in knowing exact 2-D positions of known points in a viewed-scene image. Therefore, some good fiducial markers have been developed [6][7]. Such markers enable positions to be detected and tracked easily. However, AR systems using a vision-based tracking with fiducial markers require that the markers are attached to all annotated objects.

### 2. Tracking of a known 3-D object by using a CONDENSATION algorithm

We have previously developed a wearable AR system which uses methods for solving PnP problems without fiducial markers in [3]. This system uses multiple color histogram matching\*, template matching, and the Lucas-Kanade feature-tracking method to acquire 2-D positions of known points. However, this system only uses a minimum number of feature points to estimate 6-DOF parameters, so estimated parameters are easily affected by detection error and tracking error in the 2-D positions of known points.

The goal of the present study is to develop a more robust method that enables 3-D objects to be tracked without the need for fiducial markers. Therefore, we devised a stochastic algorithm to track 6-DOF parameters. We modified the CONDENSATION algorithm, which is used for tracking the contours of objects [8]. This algorithm is based on factored sampling but can be extended to apply iteratively to successive images in a sequence.

In our method, a sample is a set of 6-DOF parameters

---

\*A "multiple color histogram" is referred to a "multiplied color histogram" in reference [3].

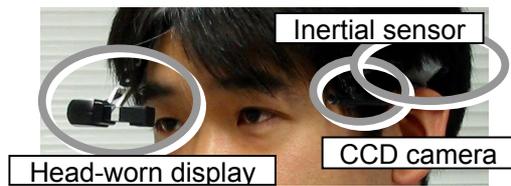


Figure 1: Headset of the VizWear-3D system

of a tracked object. We use  $N$  samples to estimate 6-DOF parameters, so  $n$  samples are generated by following two steps; 1) a center of a distribution of samples are decided by modifying 6-DOF parameters in the previous frame with difference of 3-DOF orientation parameters from the previous frame that is acquired by an inertial sensor, then 2)  $n$  samples are generated randomly around the decided center of the distribution. And  $N - n$  samples are calculated by a method for solving P3P. Three feature points, which are used by the method for solving P3P, are randomly selected from feature points that are tracked by the Lucas-Kanade's method [9]. The observation step calculates likelihood by using the distance between detected corners in the viewed-scene image and the 2-D positions, onto which the sample projects known feature points. Then, 6-DOF parameters are calculated by taking a weighted average of samples that have higher likelihood than a  $T$  % maximum likelihood. In this method, estimated parameters are rarely affected by some tracking error of 2D positions and miss-tracking.

### 3. VizWear-3D: 3-D annotation on wearable computing system

To evaluate the above-described tracking algorithm, the developed tracking method was implemented on a VizWear system, consisting of wearable equipments, remote host computers, and wireless LAN devices. Figure 1 shows one of our headsets. It was designed so as not to obscure the eyes and face of the user.

Remote hosts have an annotation database that contains sets of annotated scene information. A set of annotated object information contains object images, color information, sets of feature point information, and initial 6-DOF parameters of tracking process. A captured image is sent to remote hosts. When the hosts receive images from the client, they recognize annotated objects by multiple color histogram matching [3]. If they decide that the annotated object is present in the scene, they start to track the 6-DOF parameters of the annotated object. Then, the annotated virtual object is rendered with determined 6-DOF parameters. The client receives the synthesized scene image from remote hosts through wireless LAN.



Figure 2: Output stills of a 3-D online manual

Figure 2 shows some output stills from a prototype system of a 3-D online manual for building a PC. The proposed algorithm was implemented on a PC (CPU: Pentium Xeon 1.7 GHz). Each digitized image consists of 320x240 pixels. Sample number  $N$  was set to 1024, and  $n$  was set to 512. The system outputs synthesized images at 10 frames per second. In this prototype system, threshold rate of likelihood  $T$  was set to 80%.

### References

- [1] *VizWear*, <http://unit.aist.go.jp/is/hcv/vizwear/>.
- [2] T. Kurata, T. Okuma, M. Kourogi, T. Kato, and K. Sakaue, "VizWear: Toward Human-Centered Interaction through Wearable Vision and Visualization," Proc. PCM2001, pp. 40-47, 2001.
- [3] T. Okuma, T. Kurata, and K. Sakaue, "Real-Time Camera Parameter Estimation for 3-D Annotation on a Wearable Vision System," IEICE Trans. Inf. Syst., VolE84, 2001 (to appear).
- [4] R.T. Azuma, "A survey of augmented reality," Presence, vol. 6, No. 4, pp. 355-385, 1997.
- [5] U. Neumann and Y. Cho, "A self-tracking augmented reality system," Proc. VRST 96, pp. 109-115, 1996.
- [6] U. Neumann, S. You, Y. Cho, J. Lee, and J. Park, "Augmented Reality Tracking in Natural Environments," Mixed Reality - Merging Real and Virtual Worlds, Ohmsha & Springer-Verlag, pp. 101-130, 1999.
- [7] J. Rekimoto, "Matrix: A Realtime Object Identification and Registration Method for Augmented Reality," APCHI'98, 1998.
- [8] M.A. Isard, "Visual Motion Analysis by Probabilistic Propagation of Conditional Density," Ph.D. thesis, Department of Engineering Science, University of Oxford, 1998.
- [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.