Face Registration Using Wearable Active Vision Systems for Augmented Memory

Takekazu Kato Takeshi Kurata Katsuhiko Sakaue

Intelligent Systems Institute,
National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 JAPAN
t.kato@aist.go.jp

Abstract

This paper describes a wearable active vision system called VizWear-Active in which an active camera is used to obtain more information about the wearer and his or her environment for a wearable vision system. We have constructed the prototype system based on VizWear-Active and implemented two reflex actions: gaze direction stabilization and active tracking. For the gaze direction stabilization, the direction of the camera-head is controlled using an inertial sensor, which reduces the influence of the wearer's motion on the input images. For the active tracking, the system tracks a person by controlling the direction of the camera-head to observe the person even if the attention of the wearer is focused else where. Our prototype system performs for these reflex actions in real time.

Autonomous face registration was implemented on our prototype system for visual augmented memory applications. Facial images can be used to retrieve visual memory cues related to a human subject if various facial expressions are registered in face dictionary. Our system automatically registers the facial images in the its dictionary and uses them to retrieve visual memory cues when the system finds a particular person. We confirmed the basic functions of autonomous face registration in experiments.

1. Introduction

Wearable systems are attracting more attention as wearable devices become smaller and more efficient. The advantages of wearable systems are that they can experience the environment of the wearers and can directly assist the wearer by understanding the context of the wearer and his or her environment. In this regard, visual information is important for understanding contexts. We are researching wearable systems, interfaces and applications that use computer vision techniques. We call them collectively *VizWear*[1, 5].

Visual augmented memory is a promising application of wearable systems. It assists the wearer to recall previously experienced episodes. A visual augmented memory can be realized by storing and retrieving visual memory cues in an episode database. Visual memory cues may include information related to previous encounters with persons, such as their location, time and situations. Farringdon and Oni [3] used face recognition techniques to retrieve visual memory cues. A facial image can be used to retrieve visual memory cues if a variety of facial expressions for each person are registered in the face dictionary. It is, however, difficult to prepare a face dictionary in real-world environments. We propose autonomous face registration in which the face dictionary is automatically constructed when the system finds a particular person.

Wearable systems often use body-mounted cameras to obtain visual contexts. Usually, the cameras are fixed on the wearer's body or head and have the same field of view as the wearer. Many applications, however, require a versatility in excess of that possible with a fixed body-mounted camera, because both the wearer and objects likely move about independently. Mayol et al. [6] proposed wearable visual robots that use wearable active cameras, and evaluated some basic vision tasks. Their robots can observe objects even if the attention of the wearer is not kept on the objects. We also use a wearable active camera on the *VizWear* to extend to cognitive abilities associated with wearer's eyesight. We call this concept *VizWear-Active*. Our first prototype system based on *VizWear-Active* implemented face registration for the visual augmented memory application.

The rest of the paper is organized as follows. In the next section, we describe prototype system based on *VizWear-Active* and its basic actions. In Section 3, we discuss visual augmented memory in terms of the face registration task and show results of an experimental implementation our prototype system of *VizWear-Active*. The Section 4 is a brief conclusions.

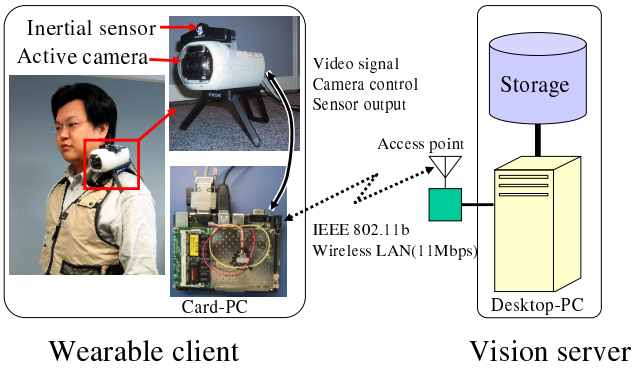


Figure 1. Prototype system for *VizWear-Active*.

2. VizWear-Active

Many current wearable systems have wearable cameras fixed on the wearer's body or head, which means they have, at best, the same field of view as the wearer. This makes visual observations dependent on the wearer's posture. Furthermore, since input images often become unstable when the wearers moves, certain vision algorithms may not work correctly. The concept of *VizWear-Active* is intended to cope with these problems. The camera can change the direction of the camera-head according to situations and purpose of the application and provides the wearable systems with a field of view independent of the wearer's motion.

2.1. Prototype System for VizWear-Active

We have constructed a prototype system based on *VizWear-Active* which consists of a wearable client and a vision server as shown in Figure 1. The wearable client includes a Card-PC (Intel mobile PentiumIII 500MHz), an active camera and an inertial sensor. The Card-PC is enough small to wear (140 mm$\times$105 mm$\times$40 mm), and it is connected to the active camera and the inertial sensor. The active camera is mounted on the wearer's shoulder. The direction of the camera-head is controlled about elevation and panning by the Card-PC. The inertial sensor is attached to the active camera, and it measures the posture of the active camera as the wearer moves.

The vision server is a high-performance desktop PC (Intel Xeon 1.7GHz dual) equipped with large storage devices. Many vision algorithms are too computationally heavy for existing stand-alone wearable computer. In our system, such tasks are implemented on the vision server, which supplements the wearable client through an on-line connection. The wearable client and the vision server communicate via a wireless LAN network (IEEE802.11b 11Mbps).

Two types of actions are required for *VizWear-Active*. The first is a reflex action for controlling the wearable active camera according to conditions of the wearer and his or her environment. The reflex action should respond in real-time to cope with various situation changes. The second is a cognitive action that understands and archives visual contexts occurring in the real-world environment. The cognitive action requires large computational resources and a large amount of storage. In our system, the reflex action is implemented on the wearable client in order to directly control the active camera in real-time, whereas the cognitive action is implemented on the vision server to use the rich resources.

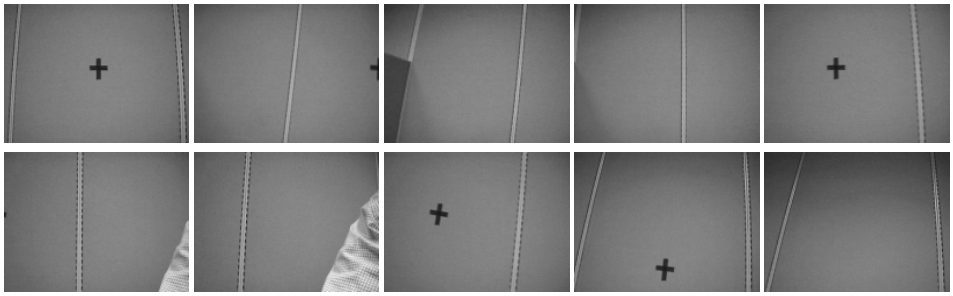
The reflex actions play important roles in *VizWear-Active* to obtain more information about the wearer and his or her environment. The basic reflex actions, image stabilization and active person tracking, are described in the rest of this section.

2.2. Gaze direction stabilization

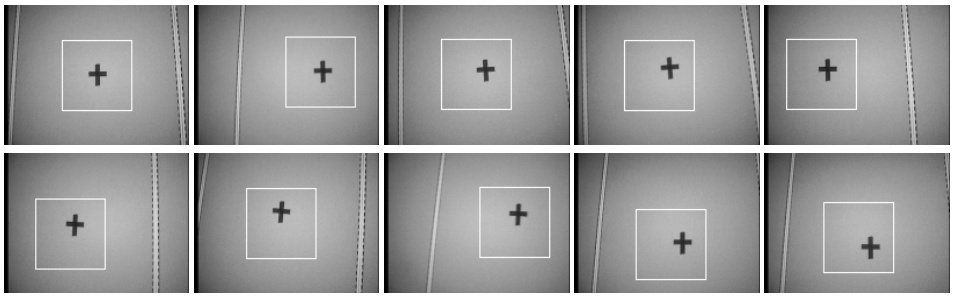
In wearable systems, the input image sequence is affected by the motion of the wearer. Gaze direction stabilization aims to keep the wearable active camera pointing independent of the wearer's body motion. This is done using the inertial sensor. The posture of the camera is measured by the inertial sensor. The camera-head is controlled to keep the gaze direction on a previously determined reference direction according to the measured posture of the camera.

Figure 2 shows input image sequences captured by our prototype system: without stabilization (a) and with stabilization (b). The sequences were captured with two cameras that were joined to each other. The "+" marker indicates the attention point of the system. The marker is not visible in some input images in (a). On the other hand, the marker is appears in all input images in (b).

The system, however, often failed to keep the marker position in the center of the input images. This problem was caused by a delay in the camera control. To reduce its influence, a virtual fovea region was added to the input images according to the error estimated by the time lag on between the camera motion and the direction control of the camera-head. The white rectangles in Figure 2 (b) indicate the virtual fovea region. The virtual fovea region keeps the marker in the center of the input image. Figure 3 shows the gaze direction stabilization comparison. The blue line indicates the error between the marker position and the center of input images without stabilization, the green line indicates the error with stabilization, and the red line is the error between the marker position and the center of virtual fovea region. We can see that the error can be reduced by the gaze direction stabilization.



(a) Without stabilization



(b) With stabilization

Figure 2. Results of gaze direction stabilization.

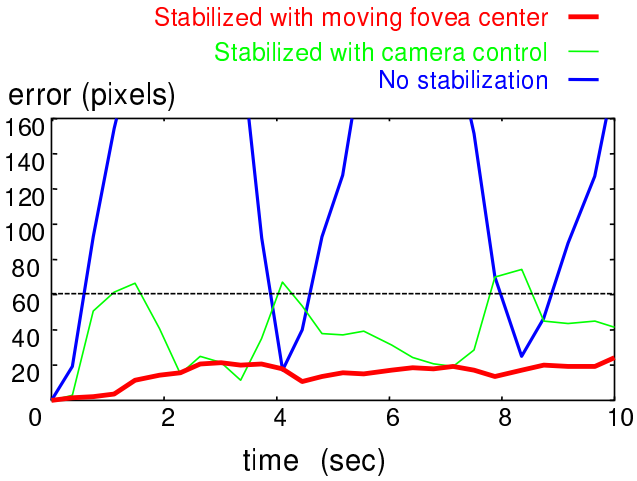


Figure 3. Gaze direction stabilization comparison.

2.3. Active Tracking

In many applications, it is important to observe not only subjects being watched by the wearer but also other subjects. When the wearer encounters a person, our system tracks the person by using the direction control of the camera-head to observe the person even if the attention of the wearer is not kept on the person.

Initially, the camera has to focus on the front of the wearer with gaze direction stabilization, and the person's head region is detected in the virtual fovea region by fitting it to an elliptic head model [2]. After detecting the head region, it is tracked by continuously fitting the elliptic model around the head region in the previous frame. Then, the direction of the camera-head is controlled to keep the head region in the center of input image.

Figure 4 shows results of active tracking. The ellipses indicate the tracked region. We can see that the person can be continuously tracked in the wide area by keeping the head region in the center of the input images.

3. Visual Augmented Memory on VizWear-Active

The visual augmented memory assists the wearer to recall episodes in his or her life. It is realized by storing and retrieving the visual memory cues in the episode database. This section describes the visual augmented memory for face registration and recognition and shows the results of an automatic face registration experiment using *VizWear-Active*.

Figure 4. Results of active tracking of a subject.

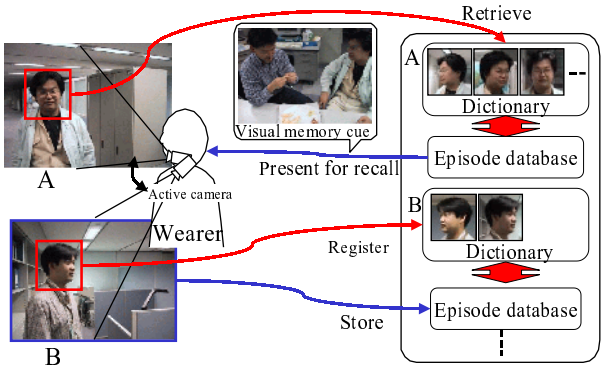


Figure 5. Visual augmented memory.

3.1. Face Registration and Recognition for Visual Augmented Memory

The episode database consists of visual memory cues which are video logs displaying previously encountered people and their environment. Face recognition techniques are used to retrieve the visual memory cues. Then, the face dictionary indexes the episode database to retrieve the visual memory cues of each person.

Whenever the wearer encounters a person, the visual memory cues of the encounter are stored in the episode database, and these are related to the person by using the face dictionary. If the wearer encounters the same person later, the visual memory cues are retrieved from the episode database using face dictionary and presented to the wearer.

To robustly recognize faces in a real-world environment, the face dictionary should contain various facial expressions for each person encountered. It is, however, difficult to prepare a sufficient number of face images in advance, because who the wearer will meet cannot be specified in advance. In [4], the authors propose a cooperative distributed face registration that can automatically and efficiently construct the face dictionary using many active cameras which are distributed and fixed in the room. We will apply this concept to the visual augmented memory applications of systems base on *VizWear-Active*.

3.2. Autonomous Face Registration

Figure 5 shows the overview of our visual augmented memory application in which the face dictionary is automatically constructed on the spot using stored visual memory cues.

Facial images are extracted from the input images, and the facing direction is estimated using the eigenface method [10, 8]. The facial image is recognized using the face dictionary. If the facial image matches an entry in the face dictionary, visual memory cues of the person are presented to the wearer from the episode database, and the input images are additionally stored in the episode database for matched person. If the facial image cannot be matched to any entry in the face dictionary, the input images are stored in the episode database for a new person. Then, the facial images are registered in the face dictionary and indexes the person in the episode database. Facial images are continuously registered in the face dictionary until there are a sufficient number of them. Since required facial images are dependent on a face recognition method, the facial images should be evaluated adjusting the face recognition method.

Our system uses the subspace method [9, 7] to recognize faces. A subspace is created from the facial images for each facing. If the DFFS (distance from feature space)[7] is small between the facial image and a dictionary entry subspace, the facial image is matched to that entry. Appearances of facial images are variously changed by influence of the environment changes. Especially, the facing and lighting conditions seriously affect the appearances. To cope with facing changes, the facial images are categorized according to the facing. Therefore, subspaces for each facing should include the images reflecting a large number of lighting conditions.

Lighting condition is evaluated using averaged face templates for typical lighting conditions [4]. The averaged face template for each lighting condition is created by averaging facial images of many people. If a person is sufficiently registered in the face dictionary, the DFFS from the subspace of the person becomes small to any averaged face template. On the other hand, the DFFS often becomes large if the

Target-A

Target-B

Figure 6. Input image sequences.

person is insufficiently registered. Therefore, the lighting condition of the registered facial images is evaluated using the maximum DFFS between the subspace and all averaged face templates for typical lighting conditions.

3.3. Experimental results

Figure 6 shows the input images obtained while tracking two persons: Target-A and Target-B. Facial images were extracted with facing estimation. These were registered in the dictionary as shown in Figure 7. Target-A was sufficiently registered for each facing. On the other hand, Target-B was not registered for the left facing because he was observed for only a short time. Figure 8 shows average images and eigenvectors of the subspace.

Figure 9 shows test images and extracted facial images of Target-A (test 1 and test 2) and Target-B (test 3 and test 4), which were observed in environments different from Figure 6. Table 1 shows the distances from the subspaces shown in Figure 8. Tests 1 and 2 were correctly matched to Target-A, and test 3 was correctly matched to Target-B. On the other hand, test 4 was not matched to any registered person, because the left profile of Target-B had not been registered in the dictionary. In this case, the facial image was registered for a new person by mistake. To solve this problem, the face dictionary could merge two or more entries that are judged to be similar.

4. Conclusion and Future Work

This paper described the concept of *VizWear-Active* in which a wearable system uses a wearable active camera to obtain more information about the wearer and his or her environment. A face registration task was implemented for a visual augmented memory application on the prototype system. Facial images were automatically registered in a face dictionary indexing the episode database when the system observed the person.

However, we were able to confirm only the most basic functions with the prototype system. To realize complete visual augmented memory applications, the episode database should be analyzed and pigeonholed so that it presents the visual memory cues desired by the wearer. Furthermore, the face dictionary should be more efficiently constructed by finding ways to eliminate failed registration and merge dictionaries in which same person are separately registered. In addition, the current prototype system uses a large camera, so we are constructing the new system that uses much smaller wearable active camera.

Acknowledgments

This work is supported by Special Coordination Funds for Promoting Science and Technology of MEXT of the Japanese Government.

References

[1] http://unit.aist.go.jp/is/hcv/vizwear/.

[2] S. Birchfield. Elliptical head tracking using intensity gradients and color histogram. In *CVPR'98*, pages 232–237, Santa Barbara, California, June 1998.

[3] J. Farringdon and V. Oni. Visual augmented memory. In *4th International Symposium on Wearable Computers(ISWC2000)*, pages 167–168, 2000.

[4] T. Kato, Y. Mukaigawa, and T. Shakunaga. Cooperative distributed tracking for effective face registration. In *2000 IAPR Workshop on Machine Vision Applications(MVA2000)*, pages 353–358, 2000.

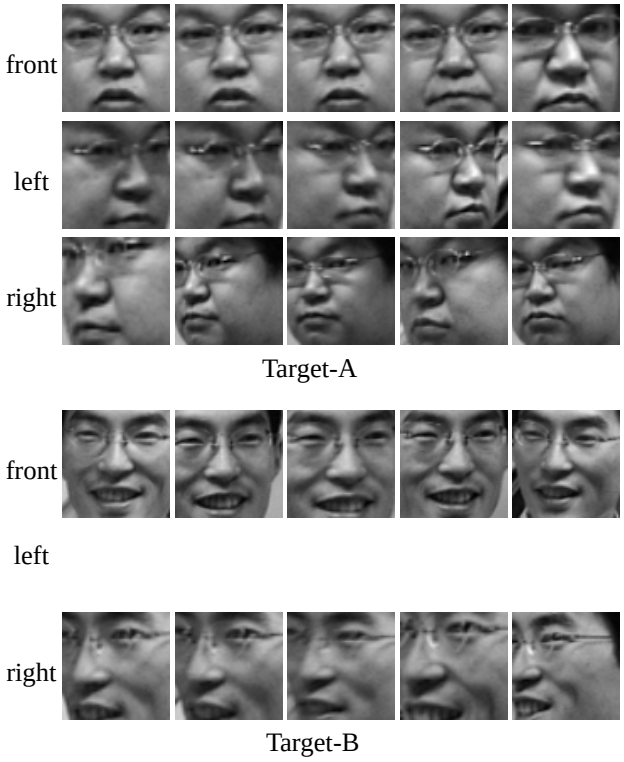


Figure 7. Examples of registered facial images.

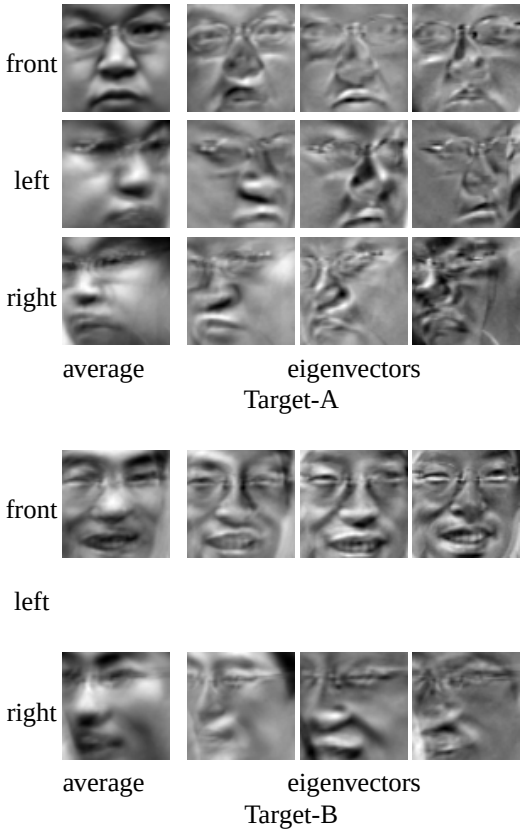


Figure 8. Averaged images and subspaces created from registered facial images.

[5] T. Kurata, T. Okuma, M. Kourogi, T. Kato, and K. Sakaue. Vizwear: Toward human-centered interaction through wearable vision and visualization. In *PCM2001*, pages –, 2001. (to appear).

[6] W. Mayol, B. Tordoff, and D. Murray. Wearable visual robots. In *4th International Symposium on Wearable Computers(ISWC2000)*, pages 95–102, 2000.

[7] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Anal. & Mach. Intell.*, vol.19(no.7):696–710, July 1997.

[8] T. Shakunaga, K. Ogawa, and S. Oki. Integration of eigentemplate and structure matching for automatic facial feature detection. In *The Third International Conference on Automatic Face and Gesture Recognition(FG'98)*, pages 94–99, Nara, Japan, Apr. 1998.

[9] Y. Sugiyama and Y. Ariki. Facial region tracking and recognition by subspace method. In *VSSM'96*, pages 225–230, Sept. 1996.

[10] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, vol.3(no.1):71–86, Jan. 1991.

Figure 9. Input images for recognition test.

Table 1. Results of recognition test.

	direction	Target-A	Target-B
test 1 (Target-A)	front	0.67	1.02
test 2 (Target-A)	left	0.65	-
test 3 (Target-B)	front	1.00	0.77
test 4 (Target-B)	left	1.09	-