Geometrically and Temporally Consistent Visual Annotation for Smart Glasses

Kanade Sumino*, Naoya Wakita and Ikuhisa Mitsugami

Hiroshima City University, JAPAN
3-4-1, Ozuka-Higashi, Asaminami-ku, Hiroshima, 731-3194, JAPAN

Abstract
In this study, we propose a wearable face recognition system using commercially available smart glasses. For this system, there are two technical contributions. First, we propose a geometric calibration between the display area on the user's visual field and the camera mounted on the smart glasses for correctly overlaying the visual annotations on the physical world observed by the user's eyes. Secondly, we propose a method for reducing the delay in showing the visual annotation for maintaining geometric and temporal consistency. We developed the whole system and experimentally confirmed that the system could show geometrically and temporally correct annotations.

Keywords
Wearable system, face detection, face recognition, calibration, multi-processing

1. Introduction

Many of you should have experienced situations where you could not recall the name or affiliation of a person who you happened to meet though you knew his/her face. It would be useful if there were a system that superimposes the name and affiliation of the person at his/her face in your view of sight. Such a system is also useful in many situations such as helping elderly people with dementia to recall people around them. In this study, we thus propose a wearable face recognition system using commercially available smart glasses. The system performs face detection and recognition processing and shows visual annotations on a transparent display in the user's field of view, which enables the user to know the names and attributes of the people in front of you. There are two technical challenges to realizing this system. The first challenge is that the annotations must be shown appropriately at the face in the user's field of view, for which we propose a geometric calibration method for the display in the field of view and a camera. The second problem is that even the geometric calibration is done the delay of the face detection and visual annotation causes the misalignment of the annotation since the user's face and the person in front of him/her is always moving. To solve this problem, we propose a multi-process architecture where the face recognition and visual annotation run as separate processes. This architecture significantly reduces the delay and realizes the geometrically and temporally correct visual annotation.

2. System Configuration

Augmented Reality (AR) is a technology that superimposes digital information on the view of sight of a person observing the real world to enable a visually augmented representation of reality. In most existing studies, they often use special devices such as Microsoft HoloLens or video see-through VR goggles. Those devices are useful as they offer functions for maintaining the geometric and temporal consistency between the annotations and the real world. However, due to their weight and special appearance, they are not suitable for us to wear in our daily lives. For wearing in our daily lives, smart glasses can be good alternatives. Though they usually do not have functions for the geometric and temporal consistency, they are small and lightweight and look like usual glasses, which are important characteristics to use them in reality. In this study, therefore, we use optically transparent smart glasses EPSON Moverio BT-30E [1]. It has binocular transparent displays that are shown around the center of the user's field of view and a camera to capture his/her field of view. Figure 1 shows an overview of the system. The device is connected to a PC via USB, and the displays and camera are recognized as an external display and webcam, respectively.

3. Proposed Method

Figure 1 shows an overview of the proposed system. It detects and recognizes a face from captured images, and

APMAR'22: The 14th Asia-PacificWorkshop on Mixed and Augmented Reality, Dec. 02-03, 2022, Yokohama, Japan
*Corresponding author.
sumino@sys.info.hiroshima-cu.ac.jp (K. Sumino);
wakita@sys.info.hiroshima-cu.ac.jp (N. Wakita);
mitsugami@hiroshima-cu.ac.jp (I. Mitsugami)
http://www.sys.info.hiroshima-cu.ac.jp/ (I. Mitsugami)
0000-0002-4306-8684 (I. Mitsugami)

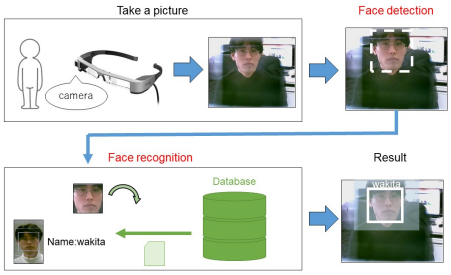

Figure 1: The proposed method

show the visual annotation at the face in the user's field of view. Since the positions of the camera and the eyes are not identical, it is necessary to geometrically calibrate them in advance to realize the positionally correct visual annotation. The following sections describe each step of the proposed system.

3.1. Face detection

For face detection, we apply a Haar-like feature-based face detector [2]. When a face is detected within the area corresponding with the display in the user's field of view, the face is cropped and saved in the system. When multiple people are detected at the same time. the system detects only the person closest to the center of the image.

OpenFace [3] , which is one of the popular face recognition libraries, is then applied to the cropped face images. OpenFace calculates similarities between the face image and the images of people stored in a database and returns a confidence level (the range of 0 to 1) for each person. In our system, we experimentally determined the threshold of the confidence level to 0.5.

3.2. Geometric Calibration for Visual Annotation

While the face detection is performed on the images captured by the camera, the visual annotation for the detected face should be shown on the display in the user's field of view. It is thus required to obtain the relation between the camera and display coordinates, as shown in Figure 2.

To obtain the relation, we propose a homography-based calibration method. As a homography matrix is a 3×3 matrix with a scale uncertainty, it has 8 unknowns, which means that four pairs of points are required to calculate. It is thus often done that the display showing (no fewer than) four points is captured by the camera to obtain the four pairs of points. In the case of this system, however, it is impossible to capture the display by the camera. To solve this problem, we propose to obtain the four pairs of points in an indirect way as follows. We first put a landmark point in the real world, and ask a user

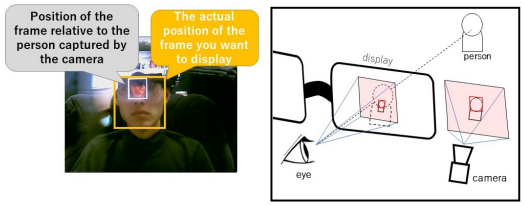

Figure 2: Necessity of geometric calibration between the camera and display.

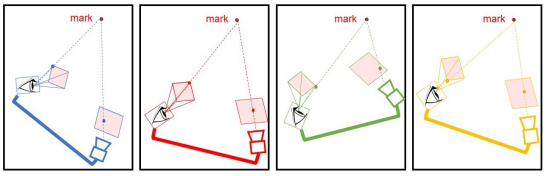

Figure 3: Gazing at the landmark while matching display corners to it in the user's sight.

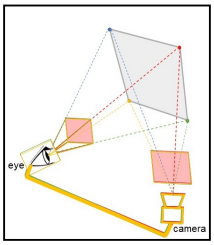

Figure 4: Four pairs of points on the image and display coordinates.

wearing the system to gaze at the landmark while matching each corner of the display to the landmark in his/her sight while the camera captures the landmark, as shown in Figure 3. We then integrate those four cases corresponding with the four corners as Figure 4. As shown in this figure, this process gives four pairs of points on the camera image and the display in the user's field of view, so that it is possible to calculate the homography matrix between those image coordinates. Once the homography is obtained, the position of the face in the camera coordinate can be transformed into the display coordinate, so that the visual annotation can be shown correctly at the face in the user's field of view as shown in Figure 5.

Note that this homography-based calibration assumes 1) the four landmarks and the face to be recognized should be located on the same plane, or 2) the centers of the eyes and camera should be colocated. Though they are not fulfilled, as the distance between the eyes and the camera is so small compared with the distance to the face, the second assumption is reasonable. Besides, considering the first assumption, it is desirable that the landmark should be at the similar depth to the face to be recognized in its actual use.

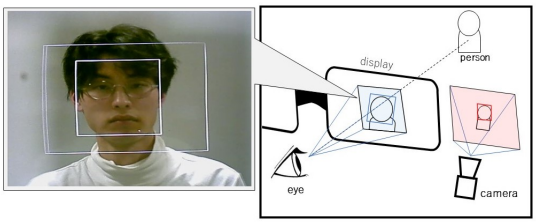

Figure 5: Visual annotation at correct position.

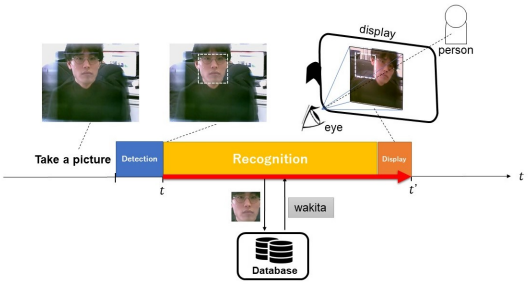

Figure 6: Displacement of visual annotation due to delays.

Table 1
Experimental comparison of average delays.

	Single-proc.	Multi-proc. (proposed)
Delay (ms)	720	71

3.3. Multi-processing for Reducing Delays

Even after the geometric calibration, it would still happen that the visual annotation is misaligned due to a delay by the face recognition process. For example, as shown in Fig. 6, even when the annotation is drawn at the position where the face was detected, if the person in the real world moves during the process, a gap occurs between the annotation and the person in the user's sight. In this study, we thus propose a method for reducing the delay by separating the whole process into that for the face recognition and that for calculating the position of the visual annotation, considering that the face recognition process takes much longer time than the others and the identity of the person in front of the user never changes so frequently while the face position changes frame by frame. Figure 7 shows the idea of the proposed method. By separating the face recognition process from the others, the facial annotation can be shown in the high frame rate and very little delay.

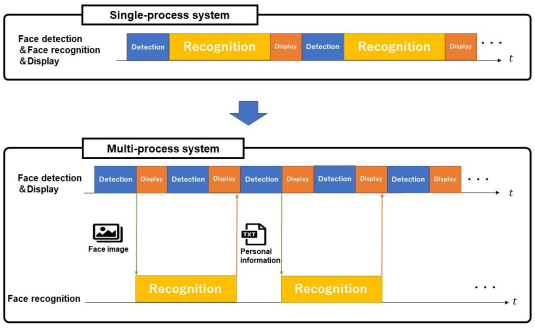

Figure 7: Multi-process processing for reducing delays.

4. Experiments

We experimentally evaluated the performance of the proposed method. We implemented the system and asked a participant (user) to keep looking at another person who was moving in front of him. It was confirmed that the visual annotation was always shown at the face in the user's sight even when the person is moving. Table 1 shows the effect of the proposed method. By applying the multi-process method, the delay is reduced by 90%.

5. Conclusion

In this paper, we propose a wearable face recognition system using commercially available smart glasses. The system performs face detection and recognition processing and shows visual annotations on a transparent display in the user's field of view, which enables the user to know the names and attributes of the people in front of you. The main contributions of this system are 1) the geometric calibration between the camera and the display in the user's sight, and 2) the multi-processing for reducing the delay in showing the visual annotation. We confirmed the effectiveness of those contributions by actually implementing the system and performing the experiments.

Future work includes making the system smaller and lighter for practical use. In the current system, the smart glasses are connected to a desktop PC, but in practical situations, they must be a wearable mobile PC or a smartphone. Another important issue for practical use is to consider the way to register people to be recognized.

References

[1] https://www.epson.jp/products/moverio/bt35e/
[2] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.
[3] B. Amos, B. Ludwiczuk, M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.