Human Visual Augmentation Using Wearable Glasses with Multiple Cameras and Information Fusion of Human Eye Tracking and Scene Understanding

Seung-Ho Yang
IT College, Kyungpook National University
1370 Sankyuk-dong, Buk-gu, Daegu, Korea, 702-701
+82-53-950-7233
yangsh@ee.knu.ac.kr

Hyun-Woo Kim
IT College, Kyungpook National University
1370 Sankyuk-dong, Buk-gu, Daegu, Korea, 702-701
+82-53-950-7233
kimhw2109@ee.knu.ac.kr

Min Young Kim
IT College, Kyungpook National University
1370 Sankyuk-dong, Buk-gu, Daegu, Korea, 702-701
+82-53-950-7233
mykim@ee.knu.ac.kr

ABSTRACT
A smart wearable robot glasses system is proposed to assist human visual augmentation in daily life, providing a refined visual recognition result to users from multiple input images of the proposed system. It consists of a glasses-type wearable device with a front-view camera, eye-view camera, mounted display, earphone, and computing unit for signal processing. The scene-understanding process on the input image from the front-view camera can be computationally accelerated with the support of the eye-view camera that monitors the eye position of the user. For efficient information processing, the eye view camera catches the user's visual intention and attention in a given situation. It is correlated to the eye viewing direction estimated from the eye-position monitoring results of eye-view camera. The proposed device can be used to augment the human visual capability in various daily life applications.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Representation**]: Multimedia Information Systems –*Artificial, augmented, and virtual realities.*

General Terms
Design, Experimentation, Human Factors, Verification.

Keywords
wearable robot; smart glasses; human visual augmentation.

1. WEARABLE ROBOT DEVICE
Human Augmentation Robots, which augment confined human ability to a higher level, are attracting robotics researcher and consumer interest boosted up with the technological innovation of information technology industries. To augment human sensing/ perception/ recognition abilities, several new visual wearable robotic devices have been developed, such as First-Person Vision or Inside-Out Vision [1-2]. Two recent keynote speeches reviewing the history and recent advances in egocentric vision revealed that most current research projects on First-Person Inside-Out Vision are related to the storage of daily human activity and its behavioral recognition for daily life analysis. However, trials to augment human sensing/ perception/ recognition abilities have not yet been extensively researched, and especially information fusion for robust and efficient recognition results from human visual augmentation devices with multiple sensors has rarely observed. The main purpose of the proposed robotic device is to assist and augment the human recognition ability by analyzing the input images, recognizing the scene contents, and serving the recognition results to the wearer in forms of visual and auditory information. Finally, the recognized results can be transferred to an audio output device in the form of an audio signal, and to a image display device in an image signal. Figure 1 shows a conceptual diagram of the proposed system, where a glasses -type design is chosen to allow a higher DOF in terms of easily changing the camera view based on monitoring the human intention. To view the scene in front and check the human intention, two cameras, one for scene monitoring and the other for human eye detection, are combined in the proposed augmentation device. All the information gathering and serving devices are integrated in the proposed wearable robot, as the embedded system for the recognition process is connected to the input camera system and the recognized results are transferred to the earphone via Bluetooth or to the video display device in front of the eye for visual information.

a) normal mode b) display mode

Figure 1. Design and implementation of wearable robot glasses with multi-cameras and display for human visual augmentation

2. VISUAL AUGMENTATION SYSTEM
To demonstrate the feasibility of the proposed system, human face recognition application was implemented for the proposed augmentation device, which is useful for meetings, conferences, and party situations. Using input images from the eye-view and front-view cameras, the purpose is to find human faces and identify the face the wearer is interested in. The identified face recognition result from embedded signal processing units is then translated into audio information or image information for the

HRI'11, March 6–9, 2011, Lausanne, Switzerland.
ACM 978-1-4503-0561-7/11/03.

wearer, as shown in Fig. 2. For efficient processing, the human intention is checked and this information utilized to endow certain regions of the input image with specialized visual attention. Here, to estimate the user's intention in the input scene, the eye position is extracted from the eye-view image, and the real eye-pointing direction of the wearer is then estimated from its image position.

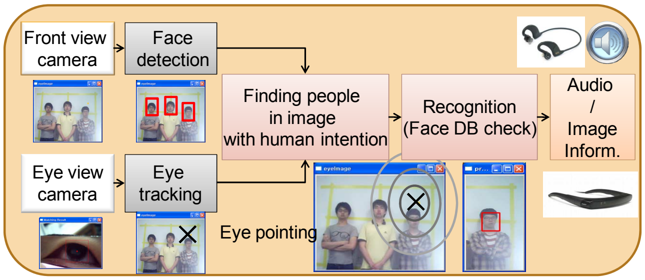

Figure 2. Software framework for human face recognition, prompted by monitoring wearer's intention

Figure 3 shows the face recognition method using eigen-face space. First, to create an eigen-face set as eigen-vectors based on the well-known principle component analysis, a covariance matrix is calculated from difference images, and the resulting covariance matrix used to calculate the eigen-values and related eigen-vectors. If the input image for online recognition is given, its face part is projected onto eigen-face space and compared with DB component values based on the learned eigen-vectors. In the case of the proposed system, image signal processing is only performed on the face areas extracted from the input images. After identifying a candidate set of saliency areas on a saliency map, AdaBoost algorithms try to filter and find the exact positions of the face detection areas based on initial information acquired from the saliency map. For efficient face recognition, the information acquired from the two input devices needs to be fused into the recognition procedure naturally. Here, the information fusion is performed in an identical space called the saliency map.

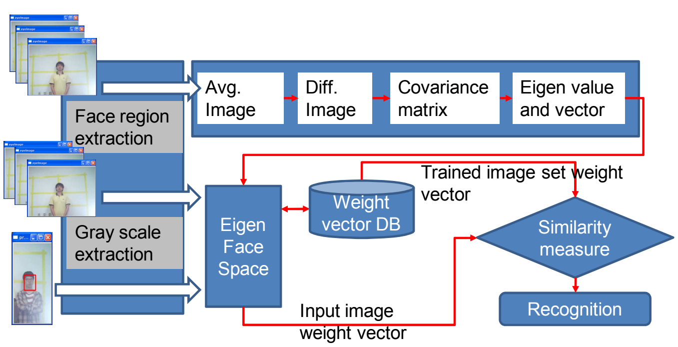

Figure 3. Face recognition procedure

3. FACE RECOGNITION EXPERIMENTS

For the face recognition, the eye detection program was first applied to the eye-view image. Next, a saliency map of the human intention in the input scene image is created. This saliency map is then combined with the face saliency map for the same input image. In this experiment, the use and non-use of the eye-tracking function implemented by the eye-view camera were compared in terms of the processing time and recognition accuracy. When comparing with general egocentric visions composed of a single camera device, the experimental results showed that the proposed vision system not only shortened the processing time for face recognition, but also improved the recognition accuracy with the help of human intention monitoring. In particular, as shown in Figure 4, when there was no human intention given, the face candidate area detection frequently failed due to illumination noise or image non-uniformity. However, when using the human intention information from the eye-view camera, the face of the person of interest was included in the face candidate area, thereby helping in successful recognition results. In summary, as shown in Table I, the processing time tended to decrease proportionally to the number of people in the input images, as the human intention defined the regions of interest for face recognition and focused the visual attention on those areas. More detailed description of the experiments can be founded in our previous publication [3].

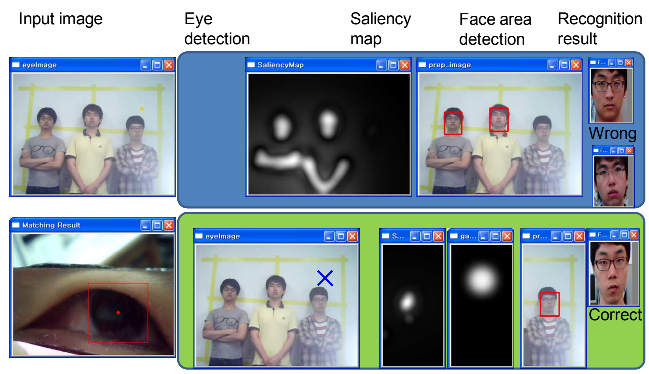

Figure 4. Recognition results with light condition variations (comparison between conventional and proposed devices)

TABLE I. SUMMARY OF EXPERIMENTAL RESULTS WITH AND WITHOUT HUMAN INTENTION MONITORING

Recognition time		**Recognition result (Success/Fail)**
Human intention check based on Eye-tracking	***Time(msec)***	
Without	534	F
With	328	S

4. ACKNOWLEDGMENTS

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (2010K001130).

5. REFERENCES

[1] T. Kanade, "First-person, Inside-Out Vision," Keynote speech, Proceedings of 2009 IEEE Computer Vision and Pattern Recognition Workshops : First Workshop on Egocentric Vision, 2009.

[2] B. Schiel, "Wearable and Egocentric Recognition and Scene Analysis," Keynote speech, Proceedings of 2009 IEEE Computer Vision and Pattern Recognition Workshops : First Workshop on Egocentric Vision, 2009.

[3] H. Kim, M.Y. Kim, S. Yang, K. Kim, H. Son, and Y. Lee, "Smart Wearable Robot Glasses for Human Visual Augmentation Based on Human Intention and Scene Understanding," Proceedings of 2010 Int. Symp. on Optomechatronic Technology, 2010. (published)