International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.11 (2016), pp.281-290
http://dx.doi.org/10.14257/ijmue.2016.11.11.24

Non-Environment-Sensitive Facial Recognition System Using Two CCTV Cameras

Euisun Kim[1], Green Bang[1], Daniel Chung[2] and Ilju Ko[1]

[1]*Department of Media, Soongsil University*
[2]*Department of Sports Information Technology, Soongsil University*
{kes33, banglgreen, danielc, andy}@ssu.ac.kr

Abstract

With the rapid growth of the high image quality CCTV monitoring market, research on facial recognition from long as well as short distances is needed. We proposed a non-environment-sensitive hybrid facial recognition system using fixed-format and pan-tilt-zoom cameras. The proposed system enables efficient system load distribution and facial recognition in the transferred images from multiple sites and channels. Furthermore, this system enables the multiple face detection and image processing of multiple CCTV channels. This system can be used for security monitoring in the public areas, and can create a new intellectual video monitoring market.

Keywords*: object detection, face detection, non-environment-sensitive facial recognition, distributed processing*

1. Introduction

Recently, closed-circuit televisions (CCTVs) have become common, and the existing ones that have low image quality are being replaced with new ones that have higher image quality. Given this trend, the facial recognition technology using high image quality CCTVs with user-friendly interfaces and continuous tracking is becoming more and more important [1]. In addition to this, non-environment-sensitive facial recognition is required as the facial recognition technology advances and the impact of the surrounding environment is reduced. With the rapid growth of the high image quality CCTV monitoring market, research on facial recognition from long as well as short distances is needed.

Several Japanese companies (*e.g*. NEC, OMRON, *etc.*) are highly interested in facial recognition technology. In particular, NEC developed an application with the following functionalities: face detection and recognition, and gender detection [2]. In the United States, besides the defense companies, companies such as Microsoft, Apple, and Google are interested in and investing in facial recognition technology. After the terrorist attack of Sep. 11, 2001, many companies invested in the reinforcement of video surveillance technology related to national security [8-9].

Now, the face detection technology for short distances is advanced enough, but that for long distances needs to be improved for general use [3]. For efficient system load distribution and facial recognition in the transferred images from multiple sites and channels, we divided one camera into two cameras, the fixed format (FF) camera and the pan-tilt-zoom (PTZ) camera, for object detection and face detection, respectively [15]. Therefore, face detection performance from long distance can be improved, and the facial recognition can be executed in the simultaneous input images from the multiple channels [4-5]. Figure 1 shows the overall architecture of the proposed system.

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2016 SERSC

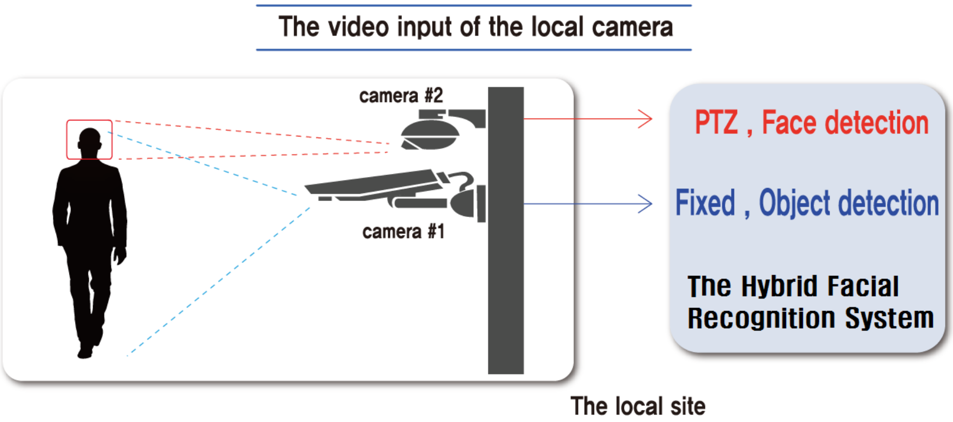


Figure 1. Overall Architecture of Hybrid Facial Recognition System

2. Face Detection Implementation for the FF Camera and the PTZ Camera

2.1. Object Separation and Tracking in the Background Image of the FF Camera

The background image is saved for object detection using the input image from long distances. Then, an object is tracked and separated using the difference image from the input image containing the object to the background image. At the same time, the PTZ camera is controlled to locate the separated object.

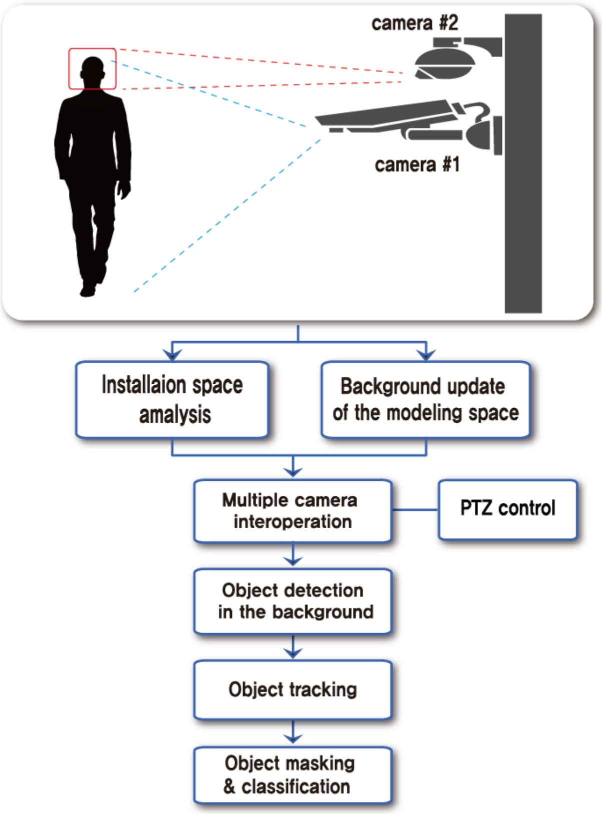


Figure 2. Flow of Applying Background

Copyright ⓒ 2016 SERSC

Figure 2 explains the image processing procedure of the FF camera. First, the background image without any object is saved and used as the background model. Second, an object in the background is separated using the background model. Third, the separated object is tracked considering the object's motion speed and direction. Fourth, the separated object is masked and the location of the masked object is sent to the PTZ camera control module. Last the PTZ camera is located to the quadrant that contains the location of the masked object for the facial recognition functionality. Figure 3 is a screenshot of the range of effect (ROE) configuration in the quadrant that contains the masked object. The computation amount of the face detection module is drastically reduced, since the quadrants that contain no object are excluded from the computation area. Therefore, this method efficiently reduces the system load of the face detection module.

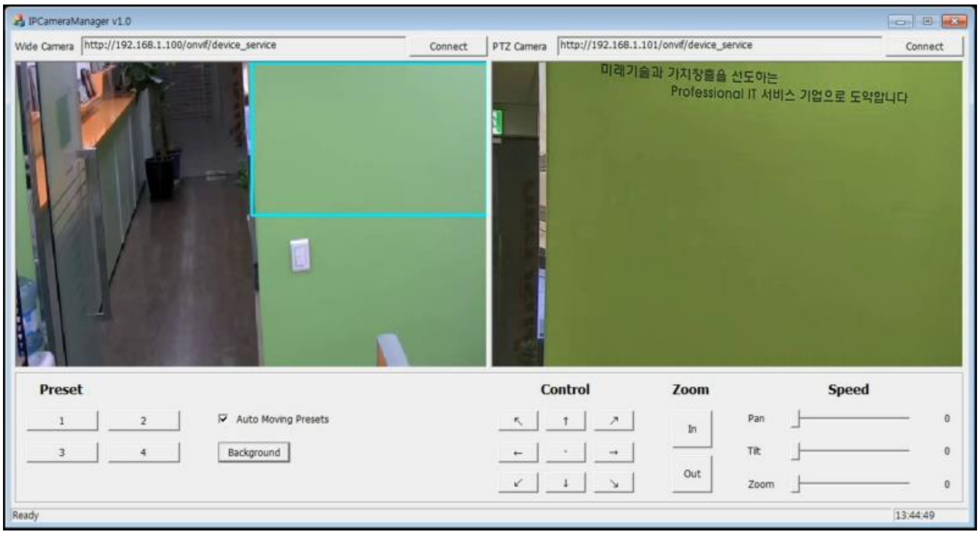


Figure 3. Screenshot of Applying Background Model

2.2. Image Separation and Object Detection in High-Resolution Input Image

To ensure efficient input image processing, we assigned the ROE and divided the input image into four quadrants. With this division, the moving time and distance of the PTZ camera are reduced, and we can take advantage of the high image quality. The input image size of the PTZ camera is 720p, which is equivalent to that of the FF camera. Thus, the face detection can be done without zooming into a specific object, which is required for face detection with low image quality.

We implemented the preset function for the PTZ camera. This function is used for focusing on objects from short and long distances. Moreover, this function enables automatic and active monitoring: automatic movement to a specific area at a specific moment, unmanned monitoring, and efficient camera management. Figure 4 shows the screenshots: the object detection of the input image with four quadrants, and the masking and tracking in the detected face area by moving the PTZ camera. Figure 5 shows the configuration of the face detection monitoring area for the environment of the installed cameras.

Copyright ⓒ 2016 SERSC

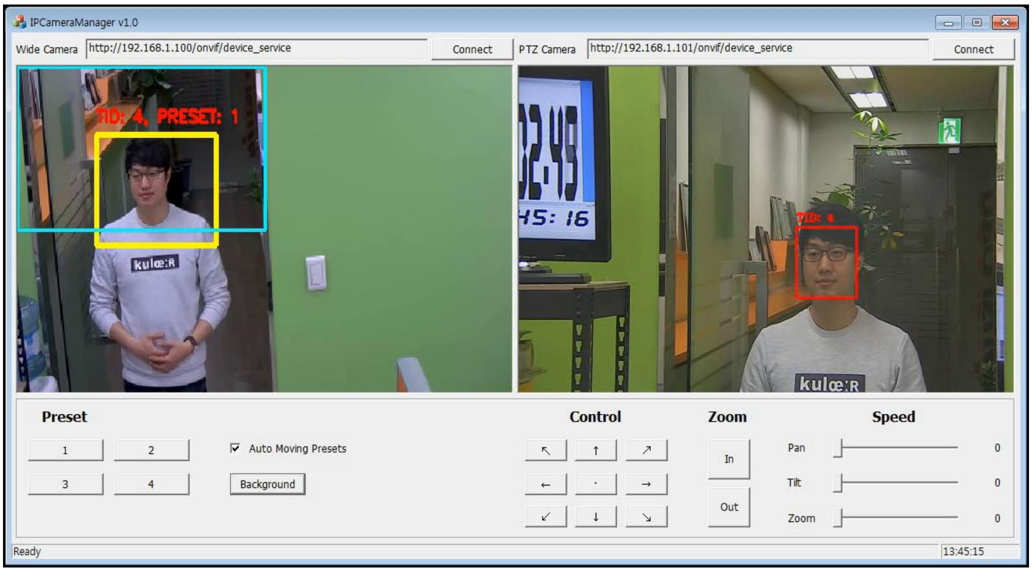
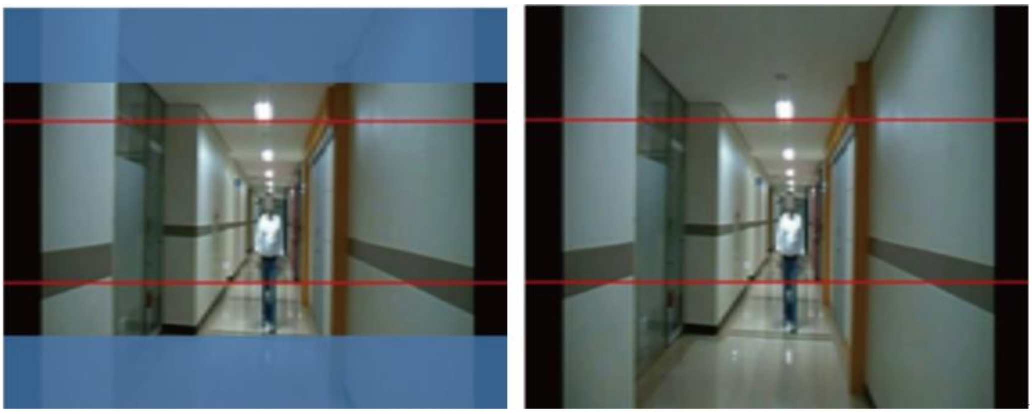


Figure 4. Screenshot of Applying Search Area of an Input Image



(a) (b)

Figure 5. Example of Effective Area Selection for Face Detection: (a) Effective Detection Area, (b) Distance from the Target

2.3. Multiple Face Detection and Tracking in the PTZ Camera

In modern CCTV cameras, high-resolution images are supported. Multiple face detection can be done automatically using high-resolution images, which do not require zooming into the face area of the detected object [6]. In this environment, we applied the automatic detection and tracking method for multiple faces. Figure 6 explains the method of ROE selection in an image from a PTZ camera. The face ID tracking engine recognizes the face area of a person and tracks it in real time from the two-dimensional color input image. This engine is composed of four modules: face detection, eye detection, noise reduction, and face tracking.

 Copyright ⓒ 2016 SERSC

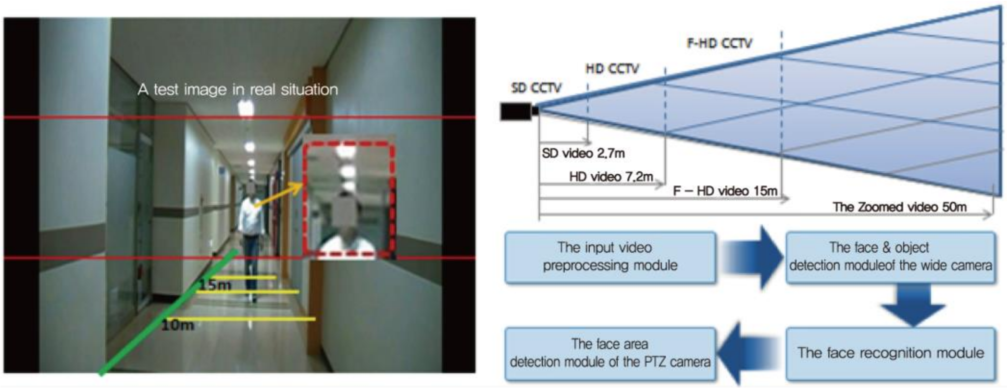


Figure 6. Effective Area of PTZ Camera

The face detection module is constructed by the learning process using the Adaptive Boosting (AdaBoost) algorithm [12]. The detailed learning process is as follows. First, the face learning database (DB) is created. The DB is composed of the 20x20-pixel face images of faces of a variety of races, genders, and ages. The face detection module is usually used for front face images, but we included additional face images with out-of-plane rotations from -30 degrees to +30 degrees and in-plane rotations from -20 degree to +20 degree to add variety to the learning DB. A difference in the eye location was allowed up to 10% of the width and the height of the image to add variety to the learning DB. We also constructed a non-face learning DB with millions of 20x20-pixel background images without faces. The binary classifier that classifies the face DB and the non-face DB is constructed by the AdaBoost algorithm. In the AdaBoost algorithm, a strong classifier is constructed with the weighted sum of hundreds of simple classifiers. The simple classifiers contain simple features composed of one Haar feature and one threshold. The AdaBoost algorithm constructs the classifiers of the serialized connection of the strong classifiers. The AdaBoost algorithm is a simple binary classifier, but it has an update rule. For wrong classification error, we assign more weight to the wrong classification to ensure more accurate classification. In other words, the weight for the classification of the face DB and the non-face DB is the same in the first stage of the learning process. In the next stage, the weight is increased for the wrong classification cases of the previous stage: the face images classified to non-face and vice versa. The classifier becomes more and more powerful via the repetition of the learning process. This process is called a “coarse-to-fine search.” The simple classifiers evaluated in each process are evaluated as the weighted sum of the strong classifier in the next stage. The eye detection module also uses the AdaBoost algorithm. To apply the algorithm, we constructed an eye learning DB and a non-eye learning DB. An element of the DBs has a 15x15-pixel image. The learning procedure of the eye detection module is similar to that of the face detection module.

The noise reduction module uses skin colors to decide if an area is a face area [10]. To make the decision, we constructed the skin color look up table (LUT) using the pixels in the skin area from the hundreds of face images. We used the RGB color space, and each channel of the color space had 16 intensity levels. For each color value from the color space, the color has the probability of skin color. In other words, if a pixel was given, the color value of the pixel was mapped to the color space, and the probability of being a skin color was evaluated by extracting the probability value saved to the skin color LUT. To decide if a specific area was a skin area, the average value of the skin color probability was computed for each pixel of the area. Using this method, we could remove non-skin areas. This procedure was applied only to the face areas marked as “unknown.” Otherwise, we knew the face ID by face and eye detection.

Copyright ⓒ 2016 SERSC

The face tracking module used the mean shift algorithm [11]. The areas identified as part of the face were registered to the object module. For the registration process, we used a 16x16x16 (12-bit) RGB color histogram. Then, we found the same color from the candidate image area with an area 1.5 times the width and 1.5 times the height of the tracked face area. Using the color from the candidate image, we found the center location of the candidate image corresponding to the center location of the tracked face image. To apply the above process iteratively, we could find the center location of the model. The face tracking works for 60 consecutive frames once the face is registered as the object model by the configuration. In this period, the object ID is renewed if a new face is detected. Moreover, the face tracking with the new object ID also works for the 60 consecutive frames. For this renewal, the merge stage, which merges the existing tracked face area and the newly detected face area, is needed. In our module, the face location priority is higher for the newly detected face, and the face ID is selected to the face ID that has the higher similarity score. The face detection process is continued to the following frames as long as a face is classified as "unknown." Figure 7 shows the visualization of the process progression.

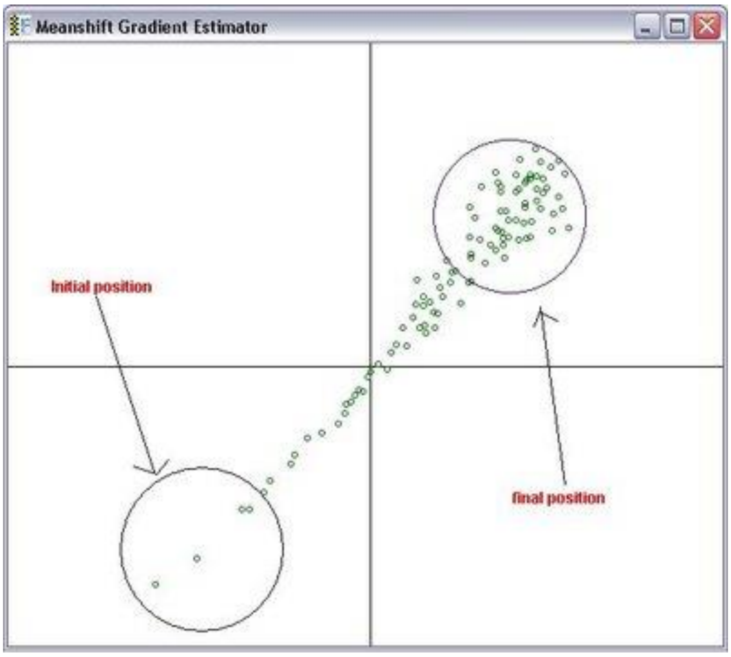


Figure 7. Object Detection Process by Mean Shift Algorithm

We also developed a hybrid face detection method fits for the environment using FF and PTZ cameras from long distances [13]. In this method, we used face template [14] for short, middle, and long distances, and defined ROI corresponding to each camera's image resolution. We changed the size of the face template for efficient face detection from various distances. Thus, we could detect the small-size faces from long distances which could not detect using the existing global face detection method. Figure 8 shows usage of face templates. We also set the face area considering the input image resolution and the size of objects for various distances. Image resolutions of CCTV cameras are as follows: standard-definition (SD), high-definition (HD), and full HD. Thus, we could define the face areas from short, middle, and long distances corresponding to CCTV camera resolution.

 Copyright ⓒ 2016 SERSC

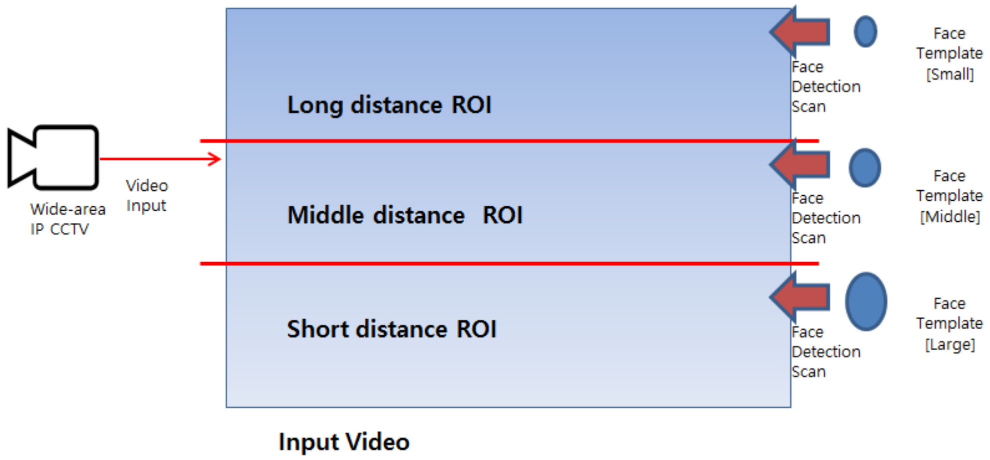


Figure 8. Face Detection using Face Template

3. Conclusion

We implemented the facial recognition system described above. The system enables the multiple face detection and image processing of multiple CCTV channels. Moreover, the system can recognize up to 15 faces and can process up to 200 CCTV channels if we use distributed threads in the facial recognition module.

Figure 9 shows the flowchart of the recognition process of the detected face image that is transferred to the facial recognition module. The detected face image is clipped and transferred. The facial recognition process is tried only in the case of successful detection of the face and the eye. During the facial recognition process, the similarity score, whose value ranges from 0 to 100, is evaluated. The facial recognition process is considered successful if the similarity score is 16 or higher and considered a failure otherwise. The facial recognition module classifies the failed case as “unknown.” The skin color level, whose value ranges from 0 to 255, is used for the noise reduction. The noise reduction is applied if the skin color level is less than or equal to 50. The facial recognition process can be executed in real time and can be designed considering the computing power of each component composing the facial recognition system. Figure 10 shows the process of displaying multiple faces simultaneously. The faces are masked and displayed by the face detection module. Figure 11 shows the facial recognition result. The facial recognition process is executed by the order of the transferred face image that is clipped by the distributed threads of the face detection module. This shows the possibility of distributed processing as follows: the local processing of object and face detection and the extensible remote processing of facial recognition.

The proposed system can be developed as a distributed architecture: face detection in the local site of the PTZ camera and the facial recognition servers in the remote site for the efficient facial recognition processing in the multiple CCTV images. The system also has the benefit of being able to increase the number of facial recognition servers and their computation power. The monitoring area can be extended with the extension of the facial recognition server. This system can be used for security monitoring in the public areas, such as subway boarding gates, and can create a new intellectual video monitoring market [6-7].

Copyright ⓒ 2016 SERSC

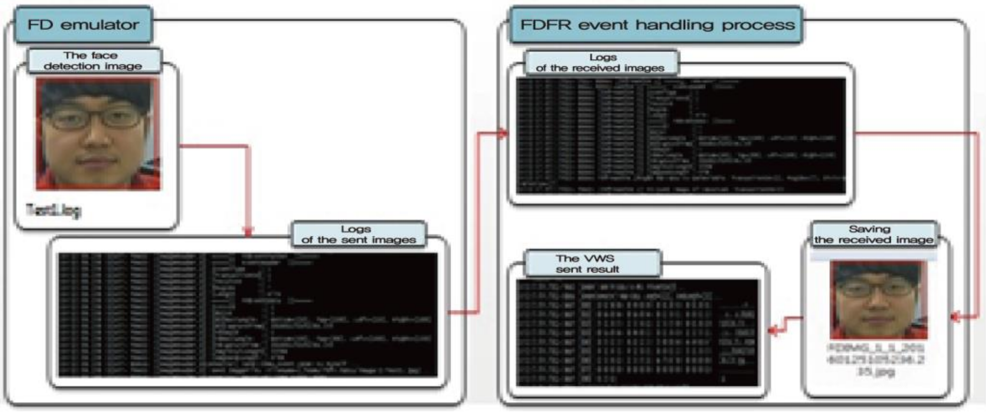


Figure 9. Recognition Process of Detected Face Image [15]

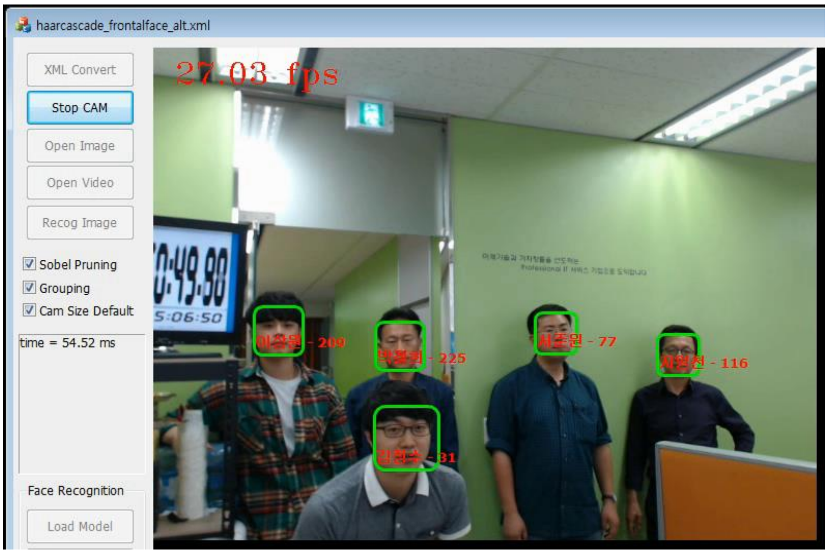


Figure 10. Multiple Face Recognition Result

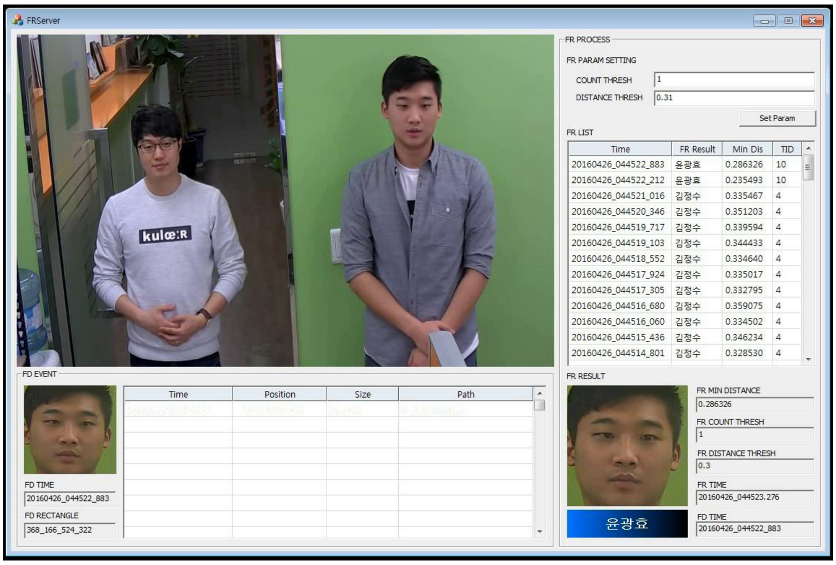


Figure 11. Facial Recognition Results of Transferred Face Image

 Copyright ⓒ 2016 SERSC

References

[1] C. Y. Jeong and J. W. Han, “Technical Trends of Abnormal Event Detection in Video Analytics”, Electronics and Telecommunications Trends, vol. 27, no. 4, **(2012)**, pp. 114-122.
[2] J. P. Lee and H. K. Kim, “Strategy of Big Data Utilization in Japan, Japan Insight”, LGERI, **(2013)**.
[3] Y. Deng, B.S. Manjunath and H. Shin, “Color image segmentation”, IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, **(1999)**.
[4] R. Adans and L. Bischof, “Seed Region Growing”, IEEE Pattern Analysis and Machine Intelligence, vol. 16, no. 6, **(1994)**.
[5] N. Ikonomakis, K. N. Ikonomakis and A. N. Venetsanopoulos, “Unsupervised seed determination for a region-based color image segmentation scheme”, Proceedings of International Conference on Image Processing, **(2000)**.
[6] J. Fan, D. K. Y. Yau, A. K. Elmagarmid and W. G. Aref, “Automatic image segmentation by integrating color-edge extraction and seeded region growing”, IEEE Transactions on Image Processing, vol. 10, no. 10, **(2001)**.
[7] M. Szummer and R. W. Picard, “Indoor-outdoor image classification”, Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Database, **(1998)**, pp. 42-51.
[8] Marcus Nieto, “Public Video Surveillance: is it an Effective Crime Prevention Tool?”, California Research Bureau, California State Library, **(1997)**.
[9] “Chicago’s Video Surveillance Cameras: A Pervasive and Unregulated Threat to Our Privacy”, ACLU of Illinois, **(2011)**.
[10] Y. Deng, B. S. Manjunath and H. Shin, “Color image segmentation”, Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, vol. 2, **(1999)**.
[11] R. Li, B. Aeng and M. L. Liou, “A New Three-Step Search Algorithm for Block Motion Estimation”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 4, no. 4, **(1994)**, pp. 438-442.
[12] Z. Yang and J. Kuo, “A Semantic Classification and Composite Indexing Approach to Robust Image Retrieval”, Proceedings of 1999 International Conference on Image Processing, vol. 1, **(1999)**, pp. 134-138.
[13] A. L. Ratan and W. E. L. Grimson, “Training Templates for Scene Classification Using a Few Example”, Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries, **(1997)**, pp. 90-97.
[14] P. Lipson, E. Grimson and P. Sinha, “Configuration Based Scene Classification and Image Indexing: Training Templates for Scene Classification Using a Few Example”, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(1997)**, pp. 1007-1013.
[15] E. Kim, G. Bang, D. Chung, and I. Ko, “Environment Independent Hybrid Face Recognition System Using a Fixed Camera and a PTZ Camera”, ASTL vol. 129, **(2016)**, pp. 196-202.

Copyright ⓒ 2016 SERSC

 Copyright ⓒ 2016 SERSC