

Article

Fuzzy System-Based Face Detection Robust to In-Plane Rotation Based on Symmetrical Characteristics of a Face

Hyung Gil Hong, Won Oh Lee, Yeong Gon Kim, Ki Wan Kim, Dat Tien Nguyen and Kang Ryoung Park *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 100-715, Korea; hell@dongguk.edu (H.G.H.); 215p8@hanmail.net (W.O.L.); csokyg@dongguk.edu (Y.G.K.); yawara18@hotmail.com (K.W.K.); nguyentientat@dongguk.edu (D.T.N.)

* Correspondence: parkgr@dongguk.edu; Tel.: +82-10-3111-7022; Fax: +82-2-2277-8735

Academic Editor: Angel Garrido

Received: 15 June 2016; Accepted: 29 July 2016; Published: 3 August 2016

Abstract: As face recognition technology has developed, it has become widely used in various applications such as door access control, intelligent surveillance, and mobile phone security. One of its applications is its adoption in TV environments to supply viewers with intelligent services and high convenience. In a TV environment, the in-plane rotation of a viewer's face frequently occurs because he or she may decide to watch the TV from a lying position, which degrades the accuracy of the face recognition. Nevertheless, there has been little previous research to deal with this problem. Therefore, we propose a new fuzzy system-based face detection algorithm that is robust to in-plane rotation based on the symmetrical characteristics of a face. Experimental results on two databases with one open database show that our method outperforms previous methods.

Keywords: TV environment; Face recognition; In-plane rotation of the face; Fuzzy systems; Symmetrical characteristics of a face

1. Introduction

With the rapid development of face recognition technology, it has been widely used in various applications such as authentication for financial transactions, access control, border control, and intelligent surveillance systems. Many studies on 2 dimensional (2D) face recognition have been performed [1–6] with 2D face detection [7,8], and there have been also previous studies on 3D face recognition [9,10]. They proposed fuzzy system-based facial feature fusion [1], convolutional neural network (CNN)-based face recognition [2,4,6], CNN-based pose-aware face recognition [3], and performance benchmarking of face recognition [5]. In addition, CNN-based face detection [7] with performance benchmarking of face detection [8] was also introduced. Three-dimensional face recognition based on geometrical descriptors and 17 soft-tissue landmarks [9] and the 3D data acquired with structured light [10] were performed as well. However, most of these previous studies were done with face images or data of high pixel resolution which are captured at a close distance from camera.

Along with the recent development of digital TV, studies have analyzed the viewers that use intelligent TV technologies such as smart TV and Internet protocol TV [11–15]. An intelligent TV provides a personalized service to the viewer. It includes a camera to obtain identity information in order to receive consumer feedback [11–15]. In order to obtain the information of the viewer using this camera, a face analysis system is used that includes the functionalities of face detection, recognition, and expression recognition [11–15]. However, different from previous research on face detection and recognition [1–10], because the camera is attached to the TV and the distance between the TV and viewer is far within the environment of watching TV, the input images are usually captured at a far

distance from the camera. Consequently, the image pixel resolution of a face area is low quality with blurring of the face image. In addition, it is often the case that people watch TV while lying on their sides. Therefore, the in-plane rotation of a face more frequently happens in images compared to the out-of-plane rotation (yaw and pitch) of a face because the face image is captured while people are watching TV, including the camera.

In previous research, An et al. adopted the methods of face detection and recognition in order to determine the identity of a TV viewer [11]. However, this method is only available for frontal face detection [11,13,16], and cannot be used for face recognition of in-plane or out-of-plane rotated faces [11]. In order to build a smart home environment, Zuo et al. proposed a method for face and facial expression recognition using a smart TV and home server, but this method did not deal with face rotation either [13]. In order to recognize a rotated face, previous methods for multi-view face detection have been based on the adaptive boosting (Adaboost) method [17–19]. However, an intensive training procedure is required to build the multi-view face detector, and these studies did not deal with the face recognition of rotated faces.

There are face detection and recognition studies that consider yaw, pitch, and in-plane face rotations [20–32]. Liu proposed a face recognition method that considers head rotation (yaw and pitch rotation) using Gabor-based kernels and principal component analysis (PCA), but this system does not deal with in-plane rotation [20] although the in-plane rotation of a face frequently occurs when a viewer watches TV while lying on his or her side. Mekuz et al. proposed face recognition that considers in-plane rotation using locally linear embedding (LLE) and PCA [26]. They also proposed face recognition methods that consider the in-plane rotation of a face using complex wavelet transforms [27] and Gabor wavelets [28]. However, they only considered in-plane rotations at small angles [26–28]. Anvar et al. proposed a method for estimating the in-plane rotation angle of a face based on scale invariant feature transforms (SIFTs), but they did not deal with face recognition [30]. In other research [31], Du et al. proposed a face recognition method based on speeded-up robust features (SURF). Their method can cope with in-plane rotated face images because of the characteristics of the scale and the in-plane rotation invariance of SURF. However, they did not show the specific experimental recognition results of in-plane rotated faces. In previous research [32], Lee et al. proposed a method of detecting the correct face box from in-plane rotated faces in a TV environment, but multiple face candidates are obtained by their method. Because all these candidates are used for face recognition, the processing time and recognition error are high.

Recently, there have been studies conducted on keypoint detection of a face image in References [33–35]. Using the results of the keypoint detection of a face image, the compensation of the in-plane rotation of a face can be possible. However, in most previous studies including References [33–35], keypoint detection has been done with face images of high pixel resolution which are captured at a close distance to the camera. In contrast, the input images captured at a far distance from the camera (maximum 2.5 m) are used in our research because our study aims at face recognition at far distances in the environment of watching TV. Consequently, the image pixel resolution of a face area is so low in addition to the blurring of a face image that the previous methods of keypoint detection are difficult to apply to the face images used in our research.

Therefore, in order to address the shortcomings of previous research, we propose a new face recognition algorithm that is robust to in-plane rotation based on symmetrical characteristics of a face in the TV environment. Compared to previous work, our research is novel in the following three ways, which are the main differences between our research and previous research [32].

- Multiple face region candidates for a face are detected by image rotation and an Adaboost face detector in order to cope with the in-plane rotation of a face.
- The credibility scores for each candidate are calculated using a fuzzy system. We use four input features. In general, the more symmetrical the left and right halves of the candidate face box are, the sharper the gray-level difference histogram (GLDH) (which is calculated by the pixel difference between the symmetrical positions based on the vertical axis that evenly bisects the

face box) is. Therefore, we define the degree of sharpness of the GLDH as the Y score in this research. Then, the differences in the Y score, pixels, average, and histogram between the left and right halves of the candidate face box are used as the four features based on the symmetrical characteristics of a face.

- The accuracy of face recognition is increased by selecting the face region whose credibility score is the highest for recognition.

The remainder of this paper is organized as follows. In Section 2, we explain the proposed fuzzy-based face recognition system. The experimental results with discussions and conclusions are described in Sections 3 and 4, respectively.

2. Proposed Face Recognition System

2.1. Overview of the Proposed Method

Figure 1 shows the overall procedure of our face recognition system. Using an image captured by the web camera connected to the set-top box (STB) for the smart TV camera (see the detail explanations in Section 3.1), the region of interest (ROI) of the face is determined by image differences between the captured and (pre-stored) background images, morphological operations, and color filtering [32]. The face region is detected within the face ROI by the Adaboost method and image rotation.

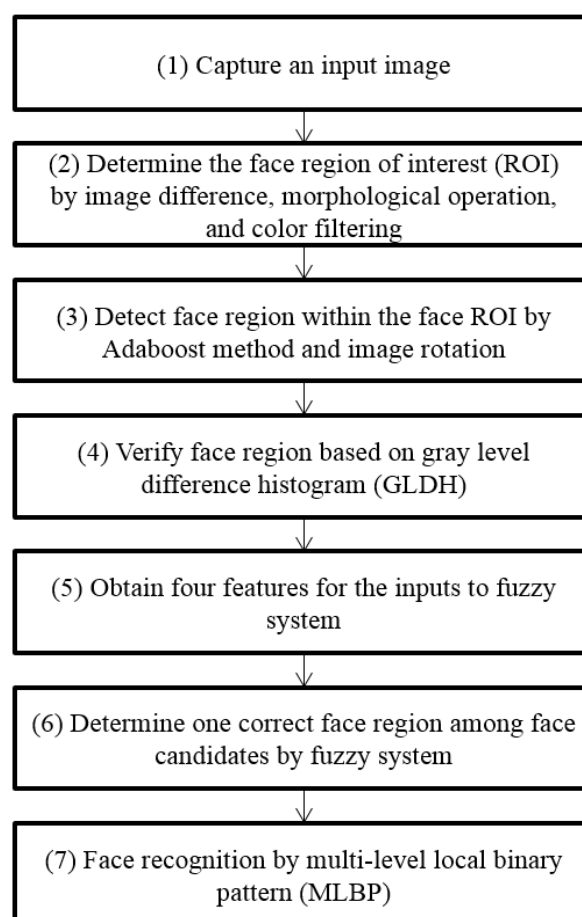


Figure 1. Flowchart of the proposed method.

Incorrect face regions can be removed using verification based on GLDH. With the face candidates, four features are extracted. Using these four features and the fuzzy system, one correct face region is selected from among the candidates. This selected face region is recognized using a multi-level

local binary pattern (MLBP). In previous research [32], steps (1)–(4) and (7) of Figure 1 are used, and steps (5) and (6) are newly proposed in our research. Through steps (5) and (6), one correct (upright) face candidate can be selected among multiple candidates, which can reduce the processing time and recognition error.

2.2. Detection and Verification of the Face Region

Using the image captured by the smart TV camera, the face ROIs are detected using image differencing (between the pre-stored background and current captured images), morphological operations, and color filtering [32]. The main goal of our research is face detection robust to in-plane rotation (not facial feature extraction or face recognition). Therefore, we use the simple method of using image differences in order to detect the rough ROI of the human body because this is not the core part of our research. Because the final goal of our research is to detect the correct face region (not the human body) from the roughly detected ROI of the human body, more accurate face ROI can be located by morphological operation, color filtering, and the Adaboost face detector with image rotation, which can reduce the error in the difference image caused by background change. That is, after the difference image is obtained by differencing (between the pre-stored background and current captured images), the area of the human body shows large difference values because the pixels within this area are different between the background and current captured image. Then, the rough area of the human body can be separated from other regions by image binarization. However, small holes inside the area of the human body in the binarized image still exist because some pixel values within this area can be similar between the background and current captured image. These holes give a bad effect on the correct detection of a face, and they can be removed by morphological operation. Because the area of the human body includes the hair, face, and body, the rough candidate region of a face can be separated by color filtering. Then, within the remaining area, more accurate face regions can be detected by the Adaboost face detector. To handle the in-plane rotation of a face, the multiple face regions are located by the face detector according to the in-plane rotation of the image.

The resulting image is shown in Figure 2a. Using the face ROIs, the face regions are detected by Adaboost and image rotation. The Adaboost algorithm is based on a strong classifier that is a combination of weak classifiers [17]. In a TV environment, the in-plane rotation of a viewer's face frequently occurs because he or she can watch the TV from a lying position, which degrades the accuracy of face detection. Therefore, we detected faces using Adaboost with the original image and six (in-plane rotated) images (at -45° , -30° , -15° , 15° , 30° and 45°). Because Adaboost detection is performed on the original image and six (in-plane rotated) images, multiple face boxes are detected even for areas that contain a single face, as shown in Figure 2b.

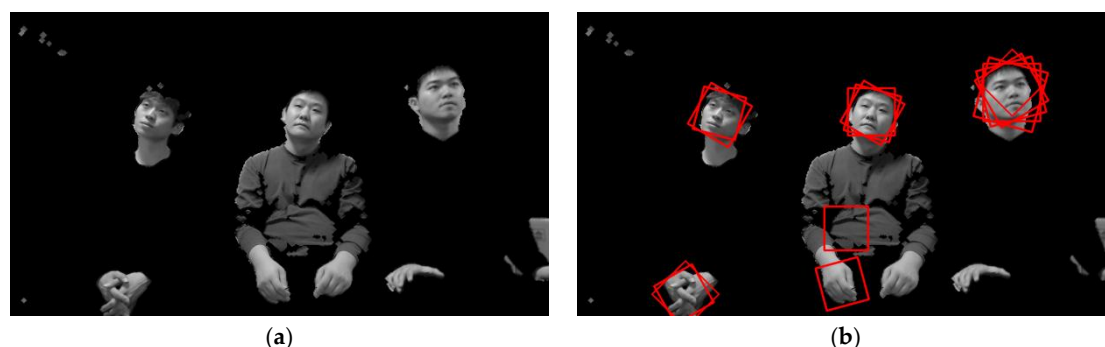


Figure 2. Cont.

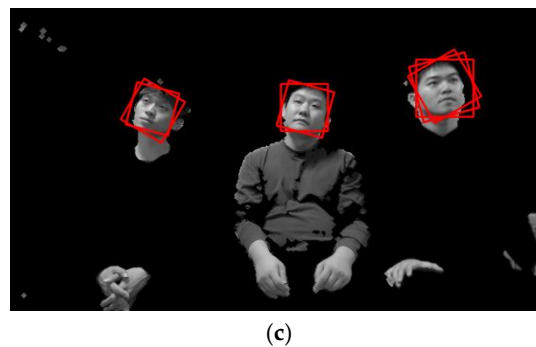


Figure 2. Detection of the face regions. (a) Detected face ROIs; (b) Multiple detected face boxes; (c) Results of the face detection using GLDH.

From the multiple detected face boxes, as shown in Figure 2b, we select candidates for correct face boxes using GLDH, as shown in Figure 2c. We use the GLDH method to select the correct box because it uses the characteristics of face symmetry to find a vertical axis that optimally bisects the face region [32]. The GLDH is calculated by the pixel difference between the symmetrical positions based on the vertical axis that evenly bisects the face box. Therefore, in general, the more symmetrical the left and right halves of the candidate face box are, the sharper the GLDH is. The GLDHs are shown at the bottom of Figure 3. The horizontal and vertical axes of the graphs respectively represent the gray-level difference (GLD) and number of occurrences [36].

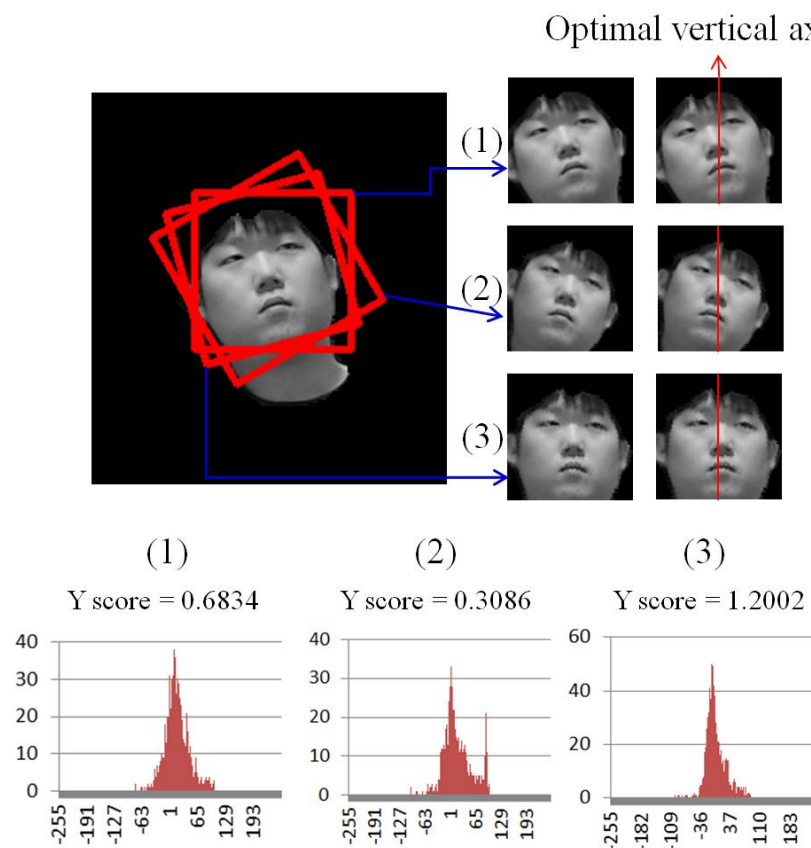


Figure 3. Y scores calculated using GLDH.

It is often the case that the face is originally rotated horizontally (yaw). Therefore, if we vertically bisect the detected face box into two equal areas, the left and right areas are not inevitably symmetrical.

Therefore, if the horizontal position of the vertical axis that evenly bisects the face box is defined as m , our system calculates the GLDHs at five (horizontal) positions ($m - 10$, $m - 5$, m , $m + 5$, and $m + 10$). If one of the five positions is the optimal vertical axis, the GLDH distribution at this position becomes a sharp shape with little variation. In an environment where a user is watching TV, severe rotation (yaw) of the user's head does not occur because he or she is looking at the TV. Therefore, the calculation of GLDH at these five positions can cope with all cases of head rotation (yaw). To measure the sharpness of the GLDH distribution, the Y score is calculated as follows [32,37]:

$$Y \text{ score} = \frac{MEAN}{\sigma^2} \quad (1)$$

where $MEAN$ is the number of pixel pairs whose GLD falls within a specified range (which we set at ± 5) based on the mean of the GLDH distribution. A high $MEAN$ represents a sharp GLDH distribution, which indicates that the corresponding bisected left and right face boxes are symmetrical. In addition, σ is the standard deviation of the distribution. Therefore, the higher the Y score, the more symmetrical the left and right halves of the face box are with respect to the vertical axis (the sharper the GLDH is).

The number of the face candidates is reduced using the Y score, as shown in Figure 2c. However, more than two face boxes still exist, even for a single face area, as shown in Figure 2c. Therefore, if multiple face candidates are used for face recognition, the processing time and face recognition error (false matches) will inevitably increase. In order to solve this problem, we propose a fuzzy-based method to select one correct face candidate. Details are given in Sections 2.3 and 2.4.

2.3. Obtaining Four Features Based on Symmetrical Characteristics of a Face for the Fuzzy System

In previous research [38,39], the characteristics of frontal face symmetry were used for face recognition. We also use four facial symmetry features as inputs for the fuzzy system. The four features (F_1 , F_2 , F_3 , and F_4) are shown below.

$$F_1 = 1 - Y \text{ score} \quad (2)$$

$$F_2 = \sum_{y=0}^{H-1} \sum_{x=0}^{W/2-1} \frac{|I(x, y) - I(W - x - 1, y)|}{W \times H} \quad (3)$$

$$F_3 = \left| \sum_{y=0}^{H-1} \sum_{x=0}^{W/2-1} \frac{I(x, y) - I(W - x - 1, y)}{W \times H} \right| \quad (4)$$

$$F_4 = \text{Chi-square distance between } HistoL \text{ and } HistoR \quad (5)$$

In Equation (2), F_1 is calculated from the Y score of Equation (1) after normalizing it to the range of 0–1. In Equations (3)–(5), $I(x, y)$ is the pixel value at position (x, y) , and W and H are the width and height of the detected face box, respectively. Equations (3)–(5) represent the differences between the left and right halves of the candidate face box based on the vertical axis that evenly bisects the face box. Equations (3) and (4) show the exemplary case where the vertical axis is positioned at the half of W .

In Equation (5), $HistoL$ and $HistoR$ respectively represent the histograms of the left-half and right-half regions of a face box.

Features F_2 – F_4 are normalized to the range of 0–1. As explained before, the higher the Y score, the more symmetrical the left and right halves of the face box are with respect to the vertical axis. In addition, F_2 – F_4 show the dissimilarity between the left and right halves of the face box. Therefore, the more symmetrical the left and right halves of the face box are with respect to the vertical axis, the smaller F_1 , F_2 , F_3 , and F_4 become. To prove this, we show the F_1 – F_4 values according to the in-plane rotation of a face as shown in Figure 4. As shown in Figure 4, the greater the amount of in-plane rotation of a face region is, the larger the F_1 – F_4 values. That is, the more symmetrical the left and right halves of the face box are with respect to the vertical axis (the smaller the amount of in-plane rotation of a face region is), the smaller F_1 , F_2 , F_3 , and F_4 become. From that, we can confirm that the F_1 – F_4

values can be used as inputs for the fuzzy system to select one correct (upright) face candidate among multiple candidates.

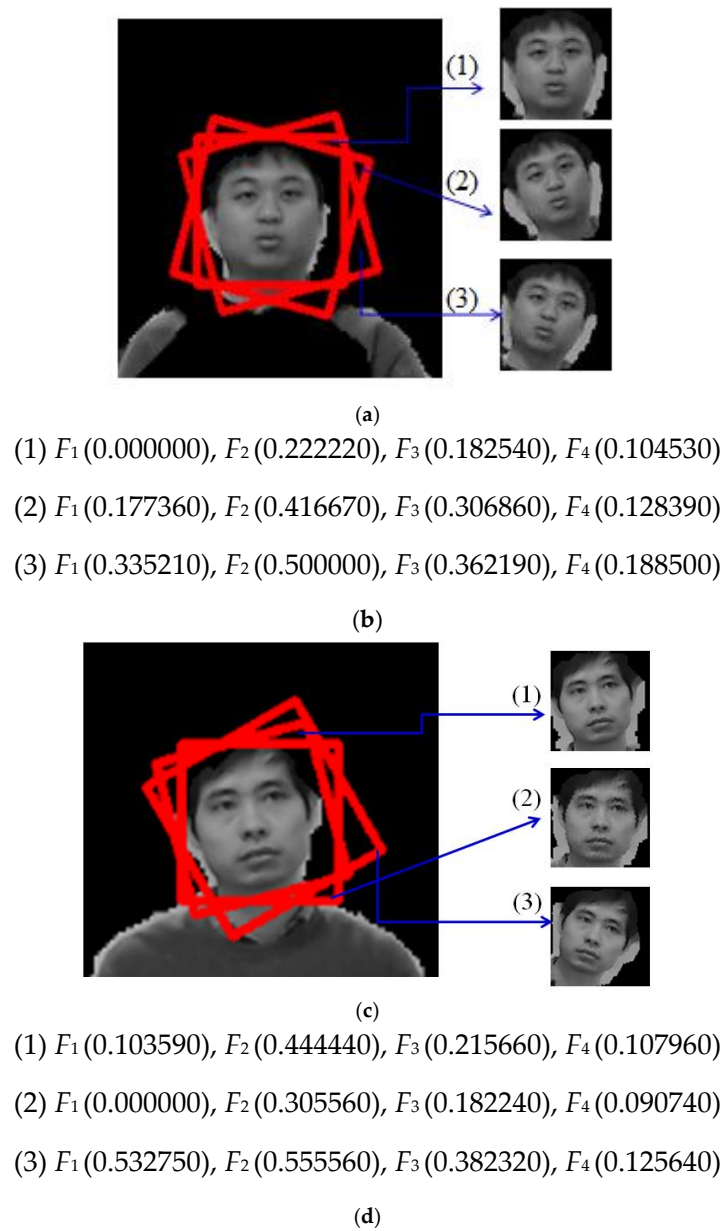


Figure 4. F_1 – F_4 values according to in-plane rotation of a face. (a,c) Examples of detected face regions including in-plane rotation; (b) F_1 – F_4 values of (1)–(3) face regions of (a); (d) F_1 – F_4 values of (1)–(3) face regions of (c).

2.4. Determining a Single Correct Face Region Using a Fuzzy System

2.4.1. Definition of Fuzzy Membership Functions and Fuzzy Rule Tables

The four features F_1 , F_2 , F_3 , and F_4 are used as inputs for the fuzzy system, and a single correct face box is its output. To achieve this, we define the input and output membership functions as shown in Figure 5a,b. Two linear functions respectively representing low (L) and high (H) are used as the input membership function. Three linear functions respectively representing low (L), medium (M), and high (H) are used as the output membership function. We acquire fuzzy output values using the input and output membership functions and the defuzzification method [40–44].

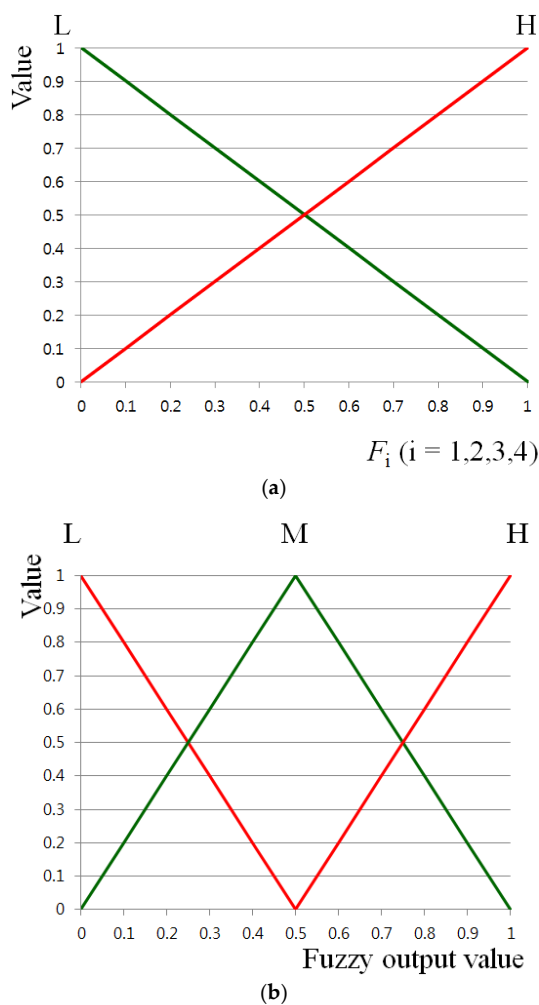


Figure 5. Input (a) and output (b) fuzzy membership functions.

As explained in Section 2.3, the more symmetrical the left and right halves of the face box are with respect to the vertical axis, the smaller F_1 , F_2 , F_3 , and F_4 become. Based on this fact, we designed the fuzzy rule table shown in Table 1. The fuzzy output values of L and H respectively represent smaller and larger amounts of symmetry of the left and right halves of the face box with respect to the vertical axis.

Table 1. Fuzzy rule table for obtaining the fuzzy system output.

Input 1 (F_1 of Equation (2))	Input 2 (F_2 of Equation (3))	Input 3 (F_3 of Equation (4))	Input 4 (F_4 of Equation (5))	Fuzzy Output Value
L	L	L	L	H
			H	H
		H	L	H
			H	M
	H	L	L	H
			H	M
		H	L	M
			H	L

Table 1. Cont.

Input 1 (F_1 of Equation (2))	Input 2 (F_2 of Equation (3))	Input 3 (F_3 of Equation (4))	Input 4 (F_4 of Equation (5))	Fuzzy Output Value
H	L	L	L	H
			H	M
		H	L	M
			H	L
	H	L	L	M
			H	L
		H	L	L
			H	L

2.4.2. Determining a Single Correct Face Region by Defuzzification

In this section, we explain the method for determining a single correct face region based on the output value of the fuzzy system. With one input feature from F_1 – F_4 of Equations (2)–(5), we can obtain two outputs using two input membership functions, as shown in Figure 6.

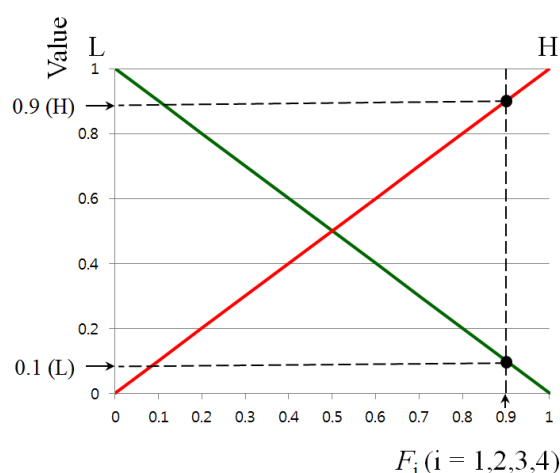


Figure 6. Obtaining two output values from a single input feature (F_i) using two input membership functions.

For example, if we assume that F_1 (of (1) face box of Figure 5a) is 0.9, 0.1 (L) and 0.9 (H) can be obtained from the L and H membership functions, respectively, as shown in Figure 6. Similarly, if F_2 , F_3 , and F_4 (of (1) face box of Figure 5a) are assumed to be 0.9, three pairs of 0.1 (L) and 0.9 (H) can be obtained. Consequently, the four pairs of 0.1 (L) and 0.9 (H) are obtained from F_1 , F_2 , F_3 , and F_4 using two input membership functions. Based on these four pairs of 0.1 (L) and 0.9 (H), we obtain the combined set as $\{(0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.1 \text{ (L)}), (0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.9 \text{ (H)}), (0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.9 \text{ (H)}, 0.1 \text{ (L)}), \dots, (0.9 \text{ (H)}, 0.9 \text{ (H)}, 0.9 \text{ (H)}, 0.1 \text{ (L)}), (0.9 \text{ (H)}, 0.9 \text{ (H)}, 0.9 \text{ (H)}, 0.9 \text{ (H)})\}$. With one subset, we can determine a single value (0.1 or 0.9) and a single symbol (L or H) based on the MIN or MAX methods [45,46] and the fuzzy rule table of Table 1.

For example, we can select 0.9 based on the MAX method with one subset (0.1 (L), 0.9 (H), 0.1 (L), 0.1 (L)). In addition, from the input of (L), (H), (L), and (L), we obtain (H) from Table 1. Consequently, we obtain 0.9 (H), which we call the inference value (IV) in this paper. Because the number of components in the combined set of $\{(0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.1 \text{ (L)}), (0.1 \text{ (L)}, 0.1 \text{ (L)}, 0.1 \text{ (L)},$

0.9 (H)), (0.1 (L), 0.1 (L), 0.9 (H), 0.1 (L)), ..., (0.9 (H), 0.9 (H), 0.9 (H), 0.1 (L), (0.9 (H), 0.9 (H), 0.9 (H), 0.9 (H))) is 16, we obtain 16 IVs.

We compared the performances of five defuzzification methods, the first of maxima (FOM), last of maxima (LOM), middle of maxima (MOM), mean of maxima (MeOM), and center of gravity (COG) [40–44]. FOM, LOM, MOM, and MeOM select one output value from the outputs determined by the maximum IV (0.9 (M) of Figure 7a). That is, FOM selects the first output value (S_2 of Figure 7a), and LOM selects the last output value (S_3 of Figure 7a). MOM selects the middle output ($(S_2 + S_3)/2$). MeOM selects the mean of all the outputs. In Figure 7a, MeOM also selects the $(S_2 + S_3)/2$.

Different from FOM, LOM, MOM, and MeOM which are based on the maximum IV, COG selects the center for the output based on the weighted average (S_5 of Figure 7c) of all the regions defined by all the IVs (the combined area of three regions R_1 , R_2 , and R_3 of Figure 7b). The method for calculating the weighted average by COG [42–44] is as follows:

$$S_5 = \frac{\int V \tilde{F}(S) \times S dS}{\int V \tilde{F}(S) dS} \quad (6)$$

Here, V and S respectively represent the variables for the vertical and horizontal axes of Figure 7b,c and \tilde{F} is the combined area of three regions, R_1 , R_2 , and R_3 , of Figure 7b.

Finally, we select one correct face box whose calculated output value by the defuzzification method is the largest. For example, if the output values of (1), (2), and (3) face boxes of Figure 5a are respectively 0.51, 0.38, and 0.79, the (3) face box is finally selected as the correct one which is used for face recognition.

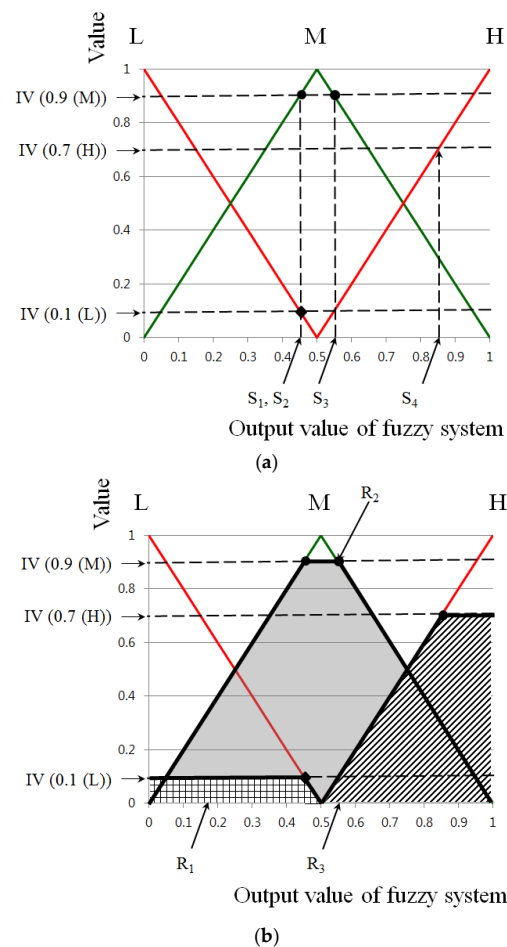


Figure 7. Cont.

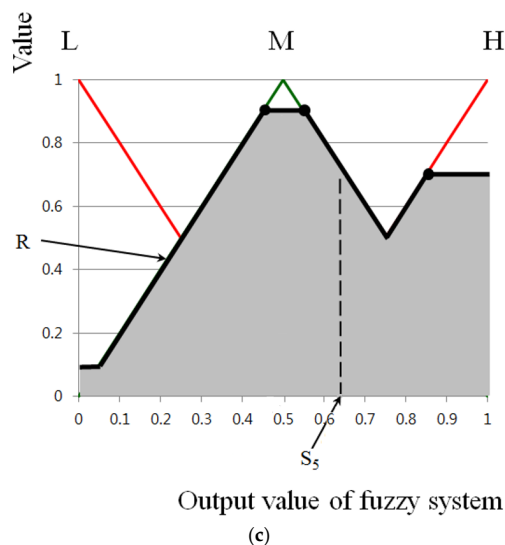


Figure 7. Obtaining the final fuzzy output value by various defuzzification methods: (a) by the first of maxima (FOM), last of maxima (LOM), middle of maxima (MOM), and mean of maxima (MeOM); (b) by the combined area of three regions of R_1 , R_2 , and R_3 ; and (c) by center of gravity (COG).

Figure 8 shows an example of the face boxes selected by the previous [32] and proposed methods. As shown in this figure, a more correct face box (where the left and right halves of the face box are more symmetrical) can be obtained using our method. Our system then recognizes faces using MLBP on the selected face box [32]. A more detailed explanation of the face recognition method can be found in [32].

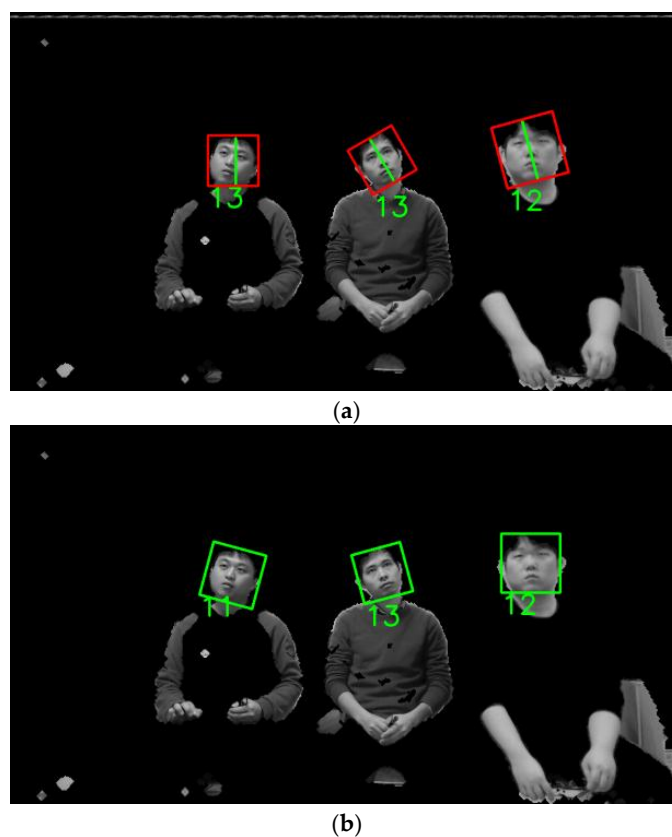


Figure 8. Examples of the final selected face boxes by (a) the previous method and (b) our method.

2.5. Face Recognition Using MLBP

The detected face regions are used for MLBP face recognition. MLBP is based on the local binary pattern (LBP) method, which assigns a binary code to each pixel based on a comparison between the center and its neighboring pixels [47]. MLBP is presented as a histogram-based LBP (concatenation of many histograms), and the LBP is a particular case of MLBP. If the center value is equal to (or greater than) the neighboring pixel, 1 is assigned; if it is less than the neighboring pixel, 0 is assigned. This basic LBP is extended to a multi-resolution method that considers various numbers P of neighboring pixels and distances R between the center and neighboring pixels as follows [32]:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \text{ where } s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (7)$$

where g_c is the gray value of the center pixel, g_p ($p = 1, \dots, P-1$) are the gray values of the p that has equally spaced pixels on a circle of radius R , and $s(x)$ is the threshold function for x . The obtained LBP codes are classified into uniform and non-uniform patterns. Uniform patterns include the number of transitions between 0 and 1 and are 0, 1, or 2. Other patterns are non-uniform patterns. The uniform patterns usually represent edges, corners, and spots, whereas the non-uniform patterns do not contain sufficient texture information. The histograms of uniform and non-uniform patterns are obtained and extracted from various sub-block levels, as shown in Figure 9 [32].

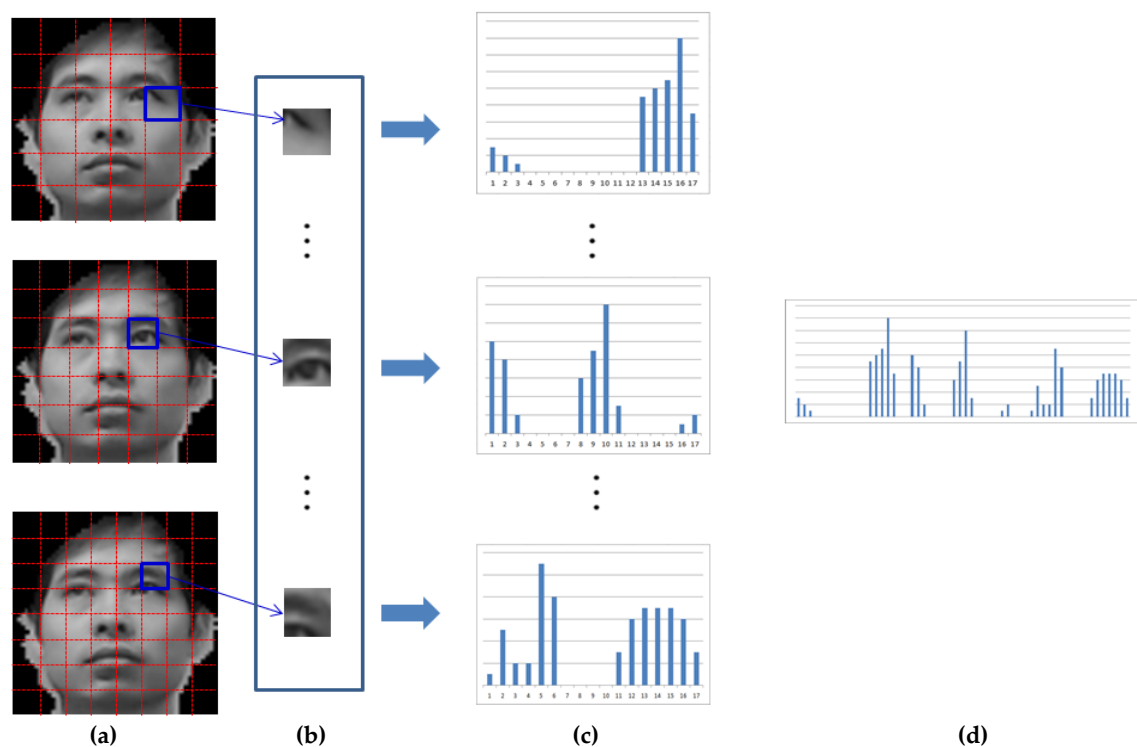


Figure 9. Example of histogram feature extraction using multi-level local binary pattern (MLBP) at three levels. (a) Face image divided into various sub-blocks; (b) Examples of sub-block regions; (c) Histograms for (b) obtained using local binary pattern (LBP); (d) Final facial feature histogram obtained by concatenating the histograms of (c).

In order to extract the histogram features globally and locally, sub-blocks of the faces are defined at three levels (the upper (6×6), middle (7×7), and lower (8×8) face of Figure 9). Because the larger-sized sub-blocks are used in the first level (upper face), the global (rough texture) features can be obtained from this sub-block. That is because the histogram information is extracted from the larger

area of a face. On the other hand, because the smaller-sized sub-blocks are used in the third level (lower face), the local (fine texture) features can be obtained from this sub-block. That is because the histogram information is extracted from the smaller area of a face.

As shown in Figure 9d, all of the histograms for each sub-block are concatenated in order to form the final feature vector for face recognition. The dissimilarity between the registered and input face histogram features is measured by the chi-square distance

$$\chi^2(E, I) = \sum_i \frac{(E_i - I_i)^2}{E_i + I_i} \quad (8)$$

where E_i is the histogram of the registered face, and I_i is the histogram of the input face. By using the histogram-based distance, a small amount of misalignment between two face images from the same person can be compensated for. In order to deal with faces in various poses (horizontal (yaw) and vertical (pitch) rotation), the histogram feature of the input face is compared with the five registered ones (which were obtained when each user looked at five positions (left-upper, right-upper, center, left-lower, and right-lower positions) on the TV during the initial registration stage) using Equation (8). If the distance calculated by Equation (8) is less than a predetermined threshold, the input face is determined to be a registered person.

3. Experimental Results and Discussions

3.1. Descriptions of Our Databases

Our algorithm is executed in the environment of a server-client-based intelligent TV. We aim to adopt our algorithm into an intelligent TV that can be used in underdeveloped countries where people cannot afford to buy smart TVs with high performance and cost. Therefore, most functionalities of the intelligent TV are provided by a low-cost STB. Additional functionalities requiring a high processing time are provided by a high-performance server, which is connected to the STB by a network. In this environment, our algorithm is executed on a STB (microprocessor without interlocked pipeline stages (MIPS)-based dual core 1.5 GHz, 1 GB double data rate 3 (DDR3) memory, 256/512 MB negative-and (NAND) memory) and server (3.5 GHz CPU and 8 GB of RAM). The STP is attached to a 60 in TV. Steps (1) and (2) of Figure 1 are performed on the STP, and steps (3) to (7) are performed on the server.

There are many face databases, e.g., FEI [48], PAL [49], AR [50], JAFFE [51], YouTube Faces [52], the Honda/UCSD video database [53], and the IIT-NRC facial video database [54]. However, most of them were not collected when a user was watching TV, and face images with in-plane rotation are not included. Therefore, we constructed our own database, which consists of images of users watching TV in natural poses, including face images with in-plane rotation. The database was collected using 15 people by separating them into five groups of three people for the experiments [32]. In order to capture images of users looking at the TV screen naturally, each participant was instructed to watch TV without any restrictions. As a result, we captured a total of 1350 frames (database I) (15 persons \times two quantities of participants (one person or three persons) \times three seating positions (left, middle, and right) \times three Z distances (1.5, 2, and 2.5 m) \times five trials (looking naturally)). In addition, a total of 300 images (database II) (five persons \times three Z distances (1.5 m, 2 m, and 2.5 m) \times two lying directions (left and right) \times 10 images) were collected for experiments when each person is lying on his or her side [32]. For face registration for recognition, a total of 75 frames (15 people \times five TV gaze points) were obtained at the Z distance of 2 m. Consequently, a total 1725 images were used for the experiments. We make our all databases (used in our research) [55] available for others to use in their own evaluations.

Figure 10 shows examples of the experimental images. For registration, five images were acquired, as shown in Figure 10a, when each user looked at five positions on the TV. Figure 10b shows examples of the images for recognition, which were obtained at various Z-distances, seating positions, and lying directions. Figure 10c shows examples of database II.



Figure 10. Cont.

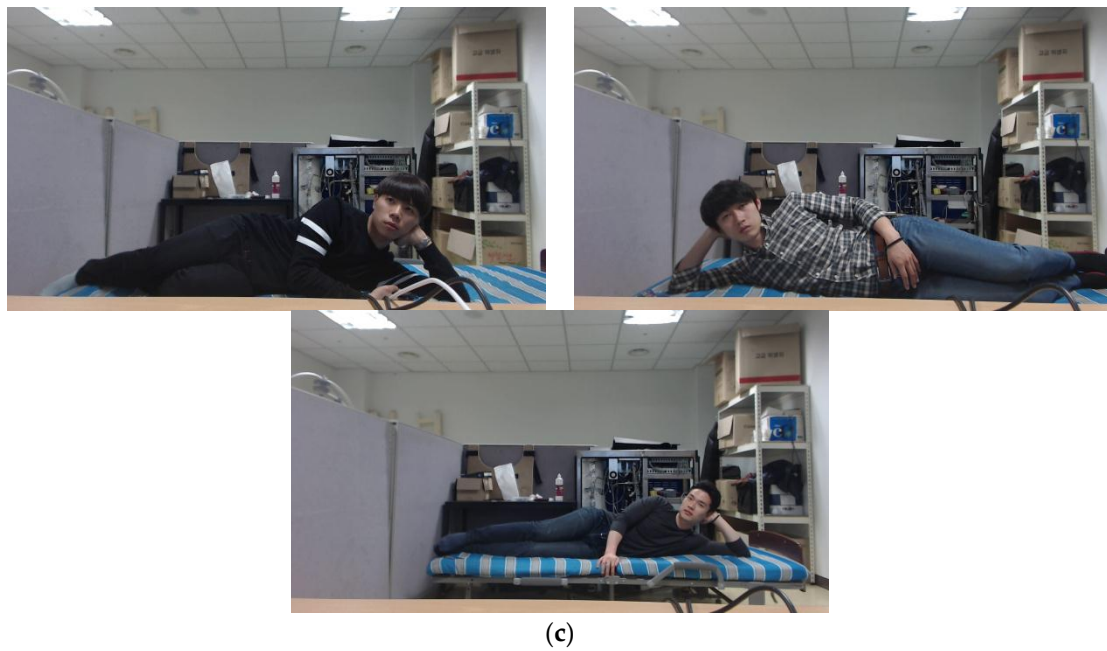


Figure 10. Examples of experimental images. (a) Images for face registration; (b) Images for recognition test (database I); (c) Images for recognition test (database II).

3.2. Experimental Results of the Face Detection and Recognition with Our Databases I and II

For the first experiment, we measured the accuracy of the face detection using database I. Accuracies were measured based on recall and precision, respectively, calculated as follows [32]:

$$\text{Recall} = \frac{\#TP}{m} \quad (9)$$

$$\text{Precision} = \frac{\#TP}{\#TP + \#FP} \quad (10)$$

where m is the total number of faces in the images; $\#FP$ and $\#TP$ are the number of false positives and true positives, respectively. False positives are cases where non-faces are incorrectly detected as faces. True positives are faces that are detected correctly. If the recall value is close to 1, the accuracy of the face detection is regarded as high. If the precision value is 1, all of the detected face regions are correct with an $\#FP$ of 0. As explained before, we measured the accuracies of the face detection according to the participant groups as shown in Table 2. In Table 2, recall and precision in the case of equal error rate (EER) are shown in bold type. EER means the error rate when the difference between the recall and precision is minimized in the trade-off relations between recall and precision. The reason why the recall at the EER point for Group 2 was lower than those for the other groups is that the face detection was not successful for the female who had hair occluding part her face and a small face. The reason why the precision at the EER point for Groups 2 and 3 is lower than those for other groups is that the colors of the subjects' clothes were similar to those of the facial skin, which caused false positives.

In Table 3, we measured the face detection accuracies according to the Z distances of the subjects in order to evaluate the effect of the change of image size (resolution). In Table 3, recall and precision in the case of equal error rate (EER) are shown in bold type as well. The recall at the EER point at a Z distance of 2.5 m is lower than for other cases because the face sizes are small, which caused the face detection to fail.

Table 2. Experimental results of the face detection according to participant groups (who have different gaze directions).

Group	Recall (%)	Precision (%)
1	94.94	99.91
	96.85	98.87
	97.35	97.13
	98.44	96.03
	99.87	94.83
2	80.38	99.28
	84.23	95.87
	89.44	91.15
	94.67	87.06
	99.45	82.11
3	90.38	99.38
	94.27	95.58
	97.78	93.92
	98.38	91.76
	99.89	90.89
4	93.09	99.94
	94.18	99.04
	95.74	98.86
	97.43	96.24
	99.91	94.03
5	96.87	99.95
	97.05	99.87
	98.7	99.26
	99.01	98.07
	99.87	97.16
Average	91.132	99.692
	93.316	97.846
	95.802	96.064
	97.586	93.832
	99.798	91.804

Table 3. Experimental results of the face detection according to Z distance.

Z Distance (m)	Recall (%)	Precision (%)
1.5	96.54	99.99
	97.42	99.02
	98.06	98.34
	99.13	97.24
	99.53	96.92
2	91.08	99.97
	93.12	97.67
	95.78	96.59
	97.59	93.33
	99.93	91.87
2.5	86.98	99.58
	90.98	96.71
	93.34	93.72
	96.26	90.17
	99.42	86.32
Average	91.53	99.85
	93.84	97.8
	95.73	96.22
	97.66	93.58
	99.63	91.70

The rows in each group (or Z distance) in Tables 2 and 3 show the changes of recall according to the decreases of precision. Because the recall and precision usually have a trade-off relationship (with a larger recall, a smaller precision is obtained, and vice versa), the changes of recall according to the decrease of precision are presented in our paper in order to show the accuracies of our face detection method more clearly through the various combinations of recall and precision.

In Tables 4 and 5, we respectively measured the accuracies of the face detection according to the seating positions and the number of participants in each image. As shown in Tables 4 and 5, the face detection accuracy is similar, irrespective of the seating position and number of people in each image.

Table 4. Experimental results of the face detection according to seating position.

Seating Position	Recall (%)	Precision (%)
Left	97.11	95.36
Middle	94.22	96.94
Right	95.78	96.15

Table 5. Experimental results of the face detection according to the number of people in each image.

Number of People	Recall (%)	Precision (%)
1	95.79	96.17
3	95.70	96.13

For the second experiment, we measured the accuracy of the face recognition with database I for various defuzzification methods. As explained in Section 2.5, the MLBP histogram of the incoming face is compared (using the chi-squared distance) to the five images of three individuals used to train it and the nearest is chosen as the identity, provided the calculated matching distance is less than the threshold. That is, it is a nearest neighbor classifier and only three identities are included in the tests. We measured the accuracy of the face recognition using the genuine acceptance rate (GAR). As shown in Table 6, the GAR by MOM with the MAX rule is higher than the GARs for other defuzzification methods. Using the MOM with the MAX rule, we compared the GAR of the proposed method to that of the previous one, as shown in Table 7, where it is clear that the GAR of our method is higher than that of the previous method for all cases.

Table 6. Experimental results (genuine acceptance rate (GAR)) of the face recognition using the proposed method and various defuzzification methods (%).

Method		GAR (%)
MIN rule	FOM	83.34
	MOM	90.65
	LOM	90.86
	MEOM	90.78
	COG	90.73
MAX rule	FOM	92.10
	MOM	92.93
	LOM	91.70
	MEOM	91.78
	COG	91.92

Table 7. Comparison of GARs of our method and the previous method according to participant group.

Group	GAR (%)	
	Previous Method [32]	Proposed Method
1	90.76	92.02
2	93.2	94.09
3	82.89	90.53
4	96.98	98.08
5	87.33	89.93
Average	90.23	92.93

In Tables 8–10, we compared the face recognition accuracy (GAR) of our method to that of the previous method with respect to the Z distance, sitting position, and number of people in each image, respectively. The GAR for various Z distances was measured in order to evaluate the effect of the change of the image size (resolution). The reason why the GAR at a Z distance of 2 m is higher than those at other Z distances is that the registration for face recognition was done with the face images captured at a Z distance of 2 m. The reason why the GAR at a Z distance of 2.5 m is lower than for other cases is that the face sizes in the images are smaller. As shown in Tables 8–10, we confirm that the GARs of our method are higher than those of the previous method in all cases, and the GARs of our method are not affected by the Z distance, sitting position, or the number of people in each image.

Table 8. Comparison of GARs for our method and the previous method for various Z distances.

Z Distance (m)	GAR (%)	
	Previous Method [32]	Proposed Method
1.5	89.11	92.22
2	92.97	96.35
2.5	88.61	90.49

Table 9. Comparison of GARs for our method and the previous method for various seating positions.

Seating Position	GAR (%)	
	Previous Method [32]	Proposed Method
Left	91.46	94.53
Middle	93.55	94.53
Right	85.64	89.42

Table 10. Comparison GARs for our method and the previous method for various number of people in each image.

Number of People	GAR (%)	
	Previous Method [32]	Proposed Method
1	90.12	92.19
3	90.57	93.67

For the next experiments, we compared the GARs of various face recognition methods [47,56–60] with our face detection method. In previous research [47], Ahonen et al. proposed LBP-based feature extraction for face recognition. PCA has been widely used to represent facial features based on eigenfaces [56,57]. Li et al. proposed a local non-negative matrix factorization (LNMf)-based method for the part-based representation of facial features [58]. In a previous study [59], they proposed support vector machine-discriminant analysis (SVM-DA)-based feature extraction for face recognition in order

to overcome the limitations of the linear discriminant analysis method that assumes that all classes have Gaussian density functions. Froba et al. proposed the modified census transform (MCT)-based facial feature extraction method which uses the average value of a 3×3 pixel mask, in contrast to the LBP method which uses the center value of a 3×3 pixel neighborhood [60]. As shown in Table 11, the GAR of our MLBP-based recognition method with our face detection method is higher than those of other methods. By using the MLBP histogram features of three levels, as shown in Figure 9, both local and global features can be efficiently used for face recognition, which improves the accuracy of the face recognition.

Table 11. Comparison of GARs of various face recognition methods with our face detection method according to groups in the database.

Group	GAR (%)						
	LBP [47]	PCA [56,57]	LNMF [58]	SVM-D [59]	MCT [60]	Previous Method [32]	MLBP
1	63.03	61.03	50.53	72.44	61.01	90.76	92.02
2	57.02	45.99	42.1	77.59	53.79	93.2	94.09
3	50.47	43.11	48.45	62.61	47.13	82.89	90.53
4	68.08	67.25	61.51	79.63	68.53	96.98	98.08
5	68.4	66.45	65.46	77.76	65.11	87.33	89.93
Average	61.4	56.77	53.61	74.01	59.11	90.23	92.93

As shown in Table 12, the GARs of our MLBP-based recognition method with our face detection method are higher than others irrespective of the change of image resolution which is caused by the change of Z distance. As explained before, because the MLBP-based method can use both local and global features for face recognition, the change of image resolution affects the facial features less using MLBP compared to other methods. In Tables 11 and 12, all the methods were applied to the same data of the face ROI detected by our face detection method for fair comparisons.

Table 12. Comparisons of GARs of various face recognition methods with our face detection method for various Z distances.

Z Distance (m)	GAR (%)						
	LBP [47]	PCA [56,57]	LNMF [58]	SVM-DA [59]	MCT [60]	Previous Method [32]	MLBP
1.5	63.06	53.51	52.71	76.55	58.59	89.11	92.22
2	64.96	57.16	56.02	79.29	63.78	92.97	96.35
2.5	56.18	59.4	52.1	66.17	54.98	88.61	90.49

Our research is mainly focused on selecting one correct (upright) face image among multiple (in-plane-rotated) face candidates (without the procedure of detecting eye positions or keypoints) based on a fuzzy system, and on enhancing the performance of face recognition by using only the selected face image. That is, the main goal of our research is face detection robust to in-plane rotation (not facial feature extraction or face recognition). In all the methods of Tables 11 and 12, our face detection method is also commonly used. That is, PCA means PCA-based face recognition with our face detection method. In the same manner, LBP means LBP-based face recognition with our face detection method. Therefore, Tables 11 and 12 just show the accuracies of various face recognition methods with our face detection method. PCA, LBP and MCT are not originally designed to be robust to in-plane rotation. Nevertheless, the reason why we selected PCA, LBP and MCT, etc. (instead of state-of-the-art methods such as deep learning-based face recognition, etc.), for comparisons in Tables 11 and 12 is to show that our face detection method can be used with any kind of traditional or even old-fashioned method whose accuracies are lower than the state-of-the-art methods for face recognition. If we use a recognition method showing high accuracies such as the deep learning-based method in Tables 11 and 12, it is difficult to analyze whether the high accuracies of recognition are

caused by our face detection method or the recognition method itself. Therefore, we include only the comparisons with traditional methods in Tables 11 and 12.

For the next test, we performed an additional experiment with database II, which includes extremely rotated faces, as shown in Figure 10c. The recall and precision of the face detection are, respectively, 96.67% and 99.39%, which are similar to those of database I in Tables 2–5. As shown in Table 13, the GAR of our method is 95.15%, which is higher than that of the previous method. In addition, the GAR of our method is similar to those of Tables 6–10. This result confirms that our method can be applied to highly rotated face images.

Table 13. Face recognition accuracy for images of highly rotated faces (database II).

Method	GAR (%)
Previous method [32]	93.10
Proposed method	95.15

Figure 11 shows the examples for which our face recognition method is successful. Figure 12 shows the examples where the face recognition failed. The failures (left person of the left figure of Figure 12 and right person of the right figure of Figure 12) are caused by false matching by the MLBP method, although the correct face boxes are selected by our method.

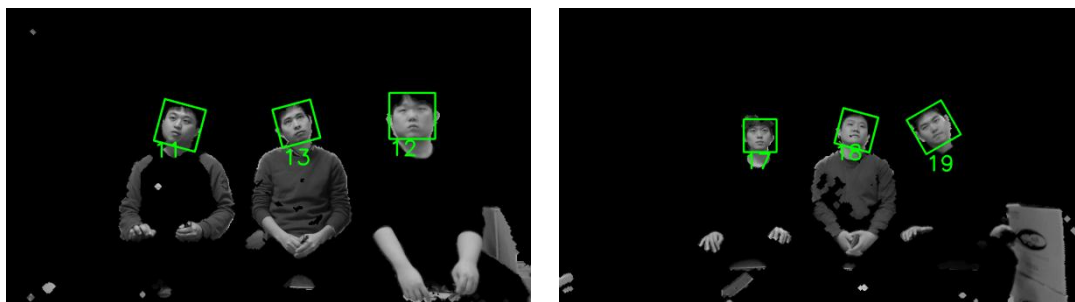


Figure 11. Examples of the success of the face recognition.

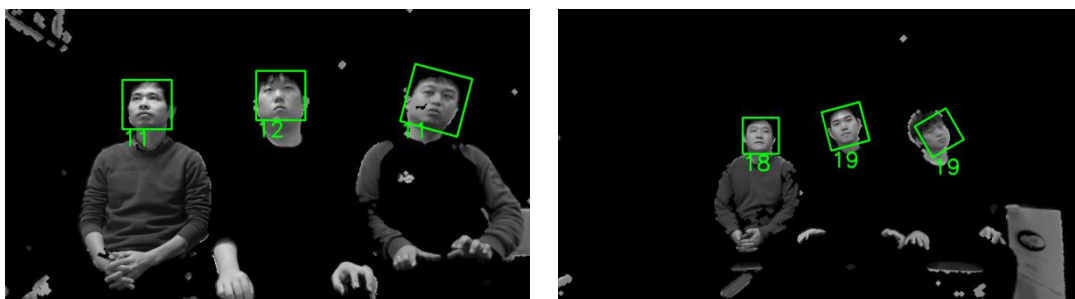


Figure 12. Examples of the failure of the face recognition.

Our method (including fuzzy system-based face detection and MLBP-based face recognition) does not require any training procedure. Even for face candidate detection, we used the original Adaboost face detector provided by the OpenCV library (version 2.4.9 [61]) without additional training. Therefore, all the experimental data were used for testing.

For the next experiment, we measured the processing time of our method. Experimental results show that the processing time per each image is approximately 152 ms. Therefore, our system can be operated at a speed of approximately six or seven frames per second. The processing time of our method is smaller than that of the previous method (185 ms) [32] because only a single face region is

selected per individual for recognition. The target applications for TV of our method are the systems for automatic audience rating surveys, program recommendation services, personalized advertising, and TV child locks. Face detection and recognition do not necessary need to be executed at every frame (real-time speed) in these applications. Therefore, our system at the current processing speed of approximately six or seven frames per second can be used for these applications.

Previous research on rotation-invariant face detection exists [62,63]. Their method can detect the correct face region from the face images including various rotations of a face based on the real Adaboost method [62]. However, the processing time of their method is so high (about 250 ms for a 320×240 image on a Pentium 4 2.4 GHz PC) that their method cannot be used in our system. In previous research [63], they show that their method can also locate the correct face region from face images including various rotations of a face by a neural network. However, the processing time of their method is so high (about six seconds to process a 160×120 pixel image on an SGI O2 workstation (Silicon Graphics Inc., Sunnyvale, CA, USA) with a 174 MHz R10000 processor (Silicon Graphics Inc., Sunnyvale, CA, USA)) that their method cannot be used in our system, either. In our system, the total processing time per one input image (1280×720 pixels) by our method is taken as 152 ms on a desktop computer (3.5 GHz CPU and 8 GB of RAM) including the processing time of steps (1) and (2) of Figure 1 on a set-top box (STB) (MIPS-based dual core 1.5 GHz, 1 GB DDR3 memory, 256/512 MB NAND memory). Although the processing time of the previous methods [62,63] includes only the procedure of face detection, our processing time of 152 ms includes both face detection and recognition. In addition, the face images in our research are considerably blurred as shown in Figure 13c,d compared to those in their research because our face images are acquired at far distance of a maximum of 2.5 m (from the camera to the user). Therefore, their methods for face detection based on the training of the real Adaboost or a neural network are difficult to apply to face images in our research.

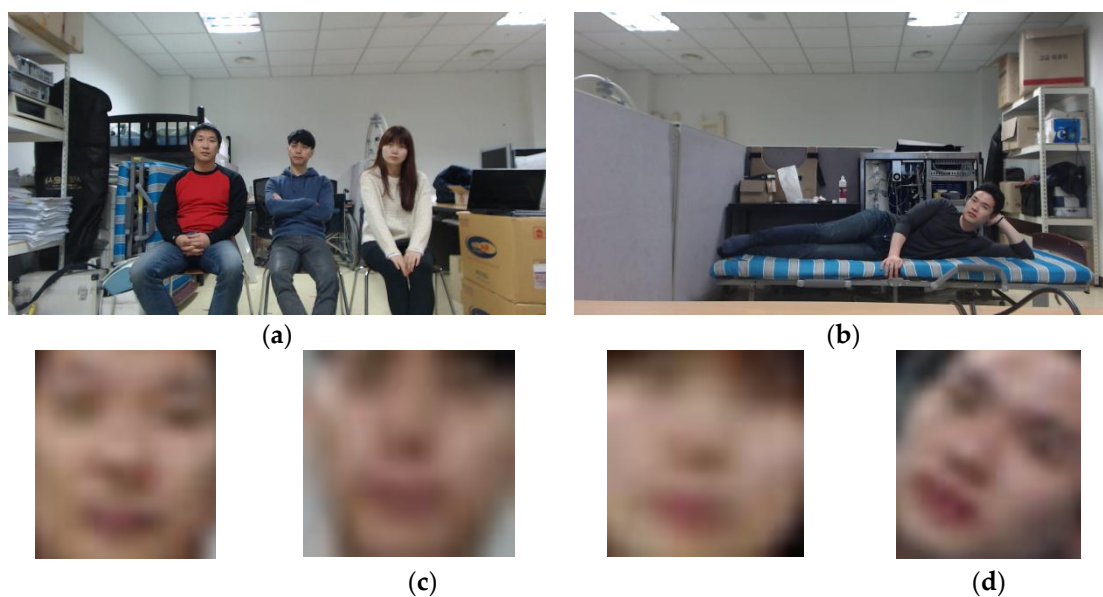


Figure 13. Comparisons of the face images in our research with those in previous studies. (a,b) Input images in our research; (c) Face images of (a); (d) Face images of (b).

In addition, we include the comparative experiments by our method with other rotation-invariant face detection methods [63]. Because our fuzzy-based method is applied to both databases I and II without any parameter tuning or training according to the type of database, the neural network of their method [63] is trained with all the images of databases I and II for fair comparison, and the testing performance are shown with databases I and II, separately.

As shown in Table 14, the accuracy of face detection by our method is higher than that by the previous method with database I. The reason why the accuracy of the previous method is lower than that of our method is that the face images in database I are blurred and the pixel resolution of the face images in database I is very low, as shown in Figure 13c. As shown in Table 15, the accuracy of face detection by our method is also higher than that of the previous method with database II. The reason why the accuracy of the previous method is lower than that of our method is that the pixel resolution of face images in database II is very low and there also exist many variations of in-plane rotation of the face images in addition to the blurring effect as shown in Figure 13d.

Table 14. Comparisons of the face detection accuracy of our method with previous method (database I).

Method	Recall (%)	Precision (%)
Previous method [63]	92.21	92.87
Proposed method	95.80	96.06

Table 15. Comparisons of the face detection accuracy of our method with previous method (database II).

Method	Recall (%)	Precision (%)
Previous method [63]	92.94	93.26
Proposed method	96.67	99.39

3.3. Experimental Results with Labeled Faces in the Wild (LFW) Open Database

As the next experiment, we measured the accuracies of the face detection with the LFW database [64]. Because our research is mainly focused on face detection robust to the in-plane rotation of a face, face images including other factors such as severe out-of-plane rotation and occlusion, etc., are excluded by manual selection for experiments among the images of the LFW database. This manual selection was performed by four people (two males and two females). Two people are in their twenties and the other two people are in their thirties. All four people are not the developers of our system and did not take part in our experiments for unbiased selection. We gave instructions (to the four people) to manually select the face images by comparing the images of the LFW database with those of our databases I and II. Then, only the images (selected by the consensus of all four people) are excluded in our experiments.

In addition, we included the comparative results of our method and the previous method [64]. As shown in Table 16, the accuracies of face detection by our method with the LFW database are similar to those with databases I and II of Tables 14 and 15. In addition, our method outperforms the previous method [63] with the LFW database.

Table 16. Comparisons of the face detection accuracy of our method with the previous method (LFW database).

Method	Recall (%)	Precision (%)
Previous method [63]	91.89	91.92
Proposed method	95.21	95.53

3.4. Discussions

There has been a great deal of previous researches on keypoint detection of a face image in References [33–35]. However, in most previous research including References [33–35], keypoint detection has been done with face images of high pixel resolution which are captured at close distance to the camera. In contrast, the input images captured at a far distance from the camera (maximum 2.5 m) are used in our research because our study aims at face recognition at far distances in the

environment of watching TV. Consequently, the image pixel resolution of a face area is so low (less than 40×50 pixels), in addition to the blurring of the face image as shown in Figure 13c,d, that the previous methods of keypoint detection or eye detection are difficult to apply to the face images used in our research.

As an experiment, we measured the accuracies of eye detection by the conventional Adaboost eye detector [17] and subblock-based template matching [65]. Experimental results showed that the recall and precision of eye detection by the Adaboost eye detector within the detected face region were about 10.2% and 12.3%, respectively. In addition, the recall and precision of eye detection by subblock-based template matching within the detected face region were about 12.4% and 13.7%, respectively. These results show that reliable eye positions or keypoints are difficult to detect in our blurred face images of low pixel resolution. Therefore, the procedures of detecting keypoints, alignment (removing in-plane rotation), and face recognition cannot be used in our research.

To overcome these problems, we propose the method of selecting one correct (upright) face image among multiple (in-plane-rotated) face candidates (without the procedure of detecting eye positions or keypoints) based on a fuzzy system, and enhancing the performance of the face recognition by using only the selected face image.

If we synthetically modify (manually rotate) the images of the open dataset, the discontinuous region (between the face and its surrounding areas) occurs in the image as shown in Figure 14b (from the YouTube dataset) and Figure 14e (from the Honda/UCSD dataset), which causes a problem in face detection and the correct accuracy of face detection is difficult to measure with these images. In order to prevent the discontinuous region, we can rotate the whole image. However, the background is also rotated as shown in Figure 14c,f, where an unrealistic background (which does not exist in the real world) is produced in the rotated image, which affects the correct measurement of the face detection accuracy.

As explained before, as shown in Figure 13c,d, the pixel resolution of images used in our research of face recognition is very low in addition to the blurring effect of a face image compared to images in open databases such as the LFPW [33], BioID [34], HELEN [35], YouTube Faces (Figure 14a), and Honda/UCSD (Figure 14d) datasets. These kinds of focused images of high pixel resolution cannot be acquired in our research environment of watching TV where the user's face is captured by a low-cost web camera at the Z distance of a maximum of 2.5 m between the camera and user (as shown in Figure 13c,d). Therefore, the experiments with these open databases cannot reflect the correct measurement of the face recognition accuracy in the environment of watching TV. There is no other open database (acquired at the Z distance of a maximum of 2.5 m) that includes large areas of background and face images of in-plane rotation like our dataset includes, as shown in Figure 13c,d.

Our method cannot deal with occluded or profiled faces. However, the cases of occluded or profiled faces do not occur in our research environment where the use is usually watching TV, as shown in Figure 10. That is because more than two people do not occlude their faces and a profiled face caused by the severe out-of-plane rotation of a face cannot happen when watching TV. Therefore, we do not consider the cases of occluded or profiled faces in our research.

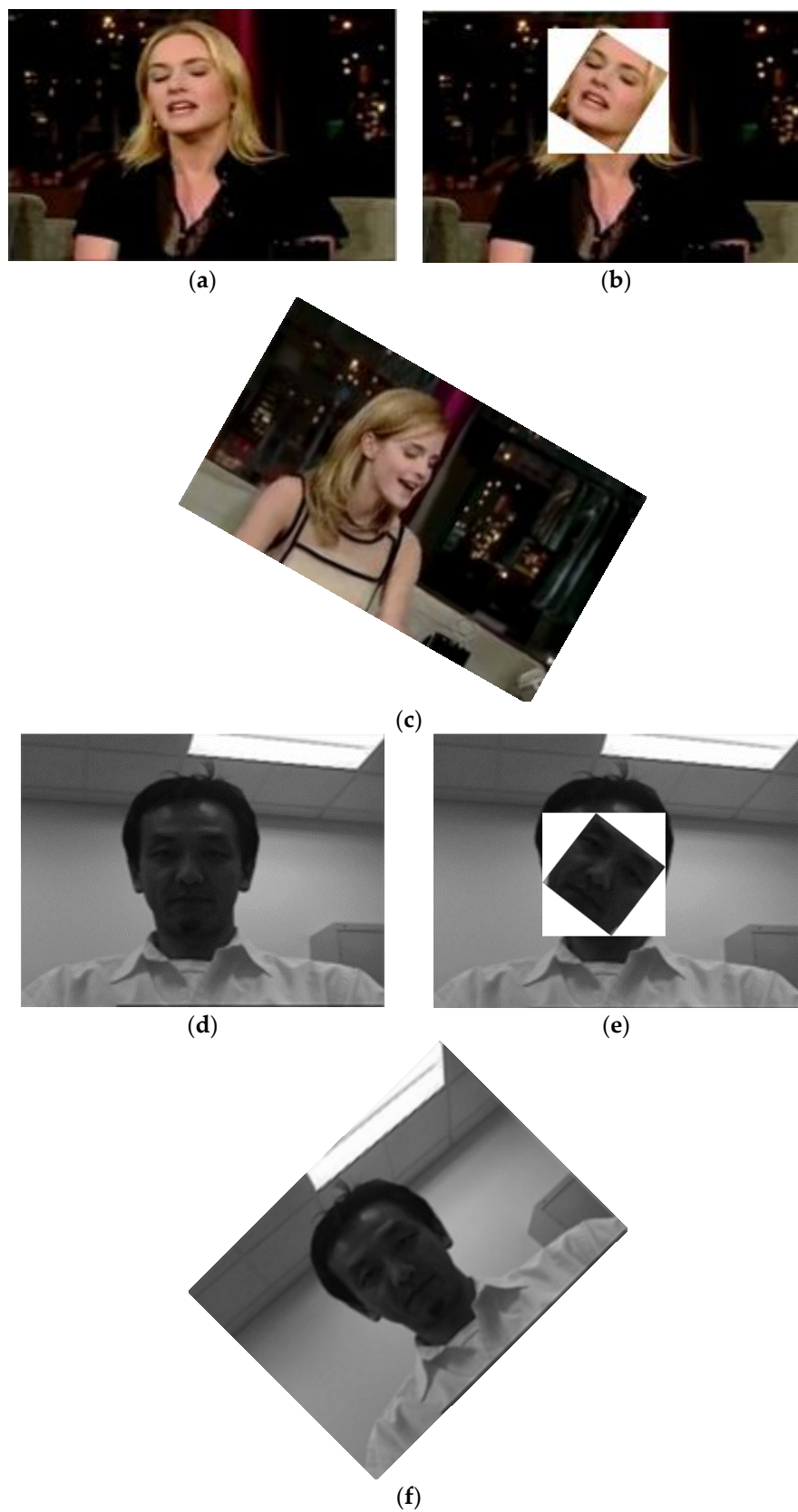


Figure 14. Example of images of YouTube and Honda/UCSD databases. (a) Original image of YouTube database [66]; (b) Image where cropped face area is rotated and discontinuous region around face area exists with (a); (c) Rotated image of YouTube dataset; (d) Original image of Honda/UCSD database [67]; (e) Image where cropped face is rotated and discontinuous region around face area exists with (d); (f) Rotated image of (d).

4. Conclusions

In this paper, we proposed a new fuzzy-based face recognition algorithm that is robust to in-plane rotation. Among the multiple candidate face regions detected by image rotation and the Adaboost face detector, a single correct face region is selected by a fuzzy system and used for recognition. Experimental results using two databases show that our method outperformed previous ones. Furthermore, the performance of our method was not affected by changes in the Z distance, sitting position, or number of people in each image. By using a non-training-based fuzzy system, our method does not require a time-consuming training procedure, and the performance of our method is less affected by the kinds of databases on which it is tested.

As future work, we plan to research a way to combine our fuzzy-based method with a training-based one such as neural networks, SVMs, or deep learning. In addition, we would research a method of enhancing the accuracy of face recognition based on other similarity metrics (such as human vs. machine d-prime) instead of the chi-square distance. In addition, the metric validity would be also checked based on spatial-taxon contours instead of precision and recall when measuring the accuracies of face detection.

Acknowledgments: This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1014) supervised by the IITP (Institute for Information & communications Technology Promotion), and in part by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP (NRF-2016M3A9E1915855).

Author Contributions: Hyung Gil Hong and Kang Ryoung Park designed the overall system for fuzzy-based face detection. In addition, they wrote and revised the paper. Won Oh Lee, Yeong Gon Kim, Ki Wan Kim, and Dat Tien Nguyen helped in the experiments and in the analyses of experimental results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Choi, J.Y.; Plataniotis, K.N.; Ro, Y.M. Face Feature Weighted Fusion Based on Fuzzy Membership Degree for Video Face Recognition. *IEEE Trans. Syst. Man Cybern. Part. B Cybern.* **2012**, *42*, 1270–1282.
2. Taigman, Y.; Yang, M.; Ranzato, M.A.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, DC, USA, 23–28 June 2014; pp. 1701–1708.
3. Masi, I.; Rawls, S.; Medioni, G.; Natarajan, P. Pose-Aware Face Recognition in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 4838–4846.
4. Sun, Y.; Wang, X.; Tang, X. Sparsifying Neural Network Connections for Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 4856–4864.
5. Kemelmacher-Shlizerman, I.; Seitz, S.M.; Miller, D.; Brossard, E. The MegaFace Benchmark: 1 Million Faces for Recognition at Scale. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 4873–4882.
6. Ghazi, M.M.; Ekenel, H.K. A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 102–109.
7. Qin, H.; Yan, J.; Li, X.; Hu, X. Joint Training of Cascaded CNN for Face Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 3456–3465.
8. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. WIDER FACE: A Face Detection Benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, CA, USA, 26 June–1 July 2016; pp. 5525–5533.
9. Vezzetti, E.; Marcolin, F.; Fracastoro, G. 3D Face Recognition: An Automatic Strategy Based on Geometrical Descriptors and Landmarks. *Robot. Auton. Syst.* **2014**, *62*, 1768–1776. [[CrossRef](#)]

10. Beumier, C. 3D Face Recognition. In Proceedings of the IEEE International Conference on Industrial Technology, Mumbai, India, 15–17 December 2006; pp. 369–374.
11. An, K.H.; Chung, M.J. Cognitive Face Analysis System for Future Interactive TV. *IEEE Trans. Consum. Electron.* **2009**, *55*, 2271–2279. [[CrossRef](#)]
12. Mlakar, T.; Zaletelj, J.; Tasic, J.F. Viewer Authentication for Personalized iTV Services. In Proceedings of the 8th International Workshop on Image Analysis for Multimedia Interactive Services, Santorini, Greece, 6–8 June 2007; pp. 63–66.
13. Zuo, F.; de With, P.H.N. Real-time Embedded Face Recognition for Smart Home. *IEEE Trans. Consum. Electron.* **2005**, *51*, 183–190.
14. Lee, S.-H.; Sohn, M.-K.; Kim, D.-J.; Kim, B.; Kim, H. Face Recognition of Near-infrared Images for Interactive Smart TV. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, Dunedin, New Zealand, 26–28 November 2012; pp. 335–339.
15. Lin, K.-H.; Shiue, D.-H.; Chiu, Y.-S.; Tsai, W.-H.; Jang, F.-J.; Chen, J.-S. Design and Implementation of Face Recognition-aided IPTV Adaptive Group Recommendation System Based on NLMS Algorithm. In Proceedings of the International Symposium on Communications and Information Technologies, Gold Coast, Australia, 2–5 October 2012; pp. 626–631.
16. Parveen, P.; Thuraisingham, B. Face Recognition Using Multiple Classifiers. In Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence, Arlington, VA, USA, 13–15 November 2006; pp. 179–186.
17. Viola, P.; Jones, M.J. Robust Real-Time Face Detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [[CrossRef](#)]
18. Jones, M.; Viola, P. *Fast Multi-View Face Detection*; TR2003–96; Mitsubishi Electric Research Laboratories: Cambridge, MA, USA, August 2003.
19. Xiao, R.; Li, M.-J.; Zhang, H.-J. Robust Multipose Face Detection in Images. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 31–41. [[CrossRef](#)]
20. Liu, C. Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 572–581. [[PubMed](#)]
21. Ban, J.; Féder, M.; Jirka, V.; Loderer, M.; Omelina, L.; Oravec, M.; Pavlovičová, J. An Automatic Training Process Using Clustering Algorithms for Face Recognition System. In Proceedings of the 55th International Symposium ELMAR-2013, Zadar, Croatia, 25–27 September 2013; pp. 15–18.
22. Asthana, A.; Marks, T.K.; Jones, J.M.; Tieu, K.H.; Rohith, M.V. Fully Automatic Pose-Invariant Face Recognition via 3D Pose Normalization. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 937–944.
23. Joo, Y.H.; An, K.H.; Park, J.W.; Chung, M.J. Real-Time Face Recognition for Mobile Robots. In Proceedings of The 2nd International Conference on Ubiquitous Robots and Ambient Intelligence, Daejeon, Korea, 2–4 November 2005 ; pp. 43–47.
24. Kim, D.H.; An, K.H.; Chung, M.J. Development of a Cognitive Head-Eye System: Real-time Face Detection, Tracking, Recognition and Visual Attention to Human. In Proceedings of the 13th International Conference on Advanced Robotics, Jeju Island, Korea, 21–24 August 2007; pp. 303–307.
25. Ryu, Y.G.; An, K.H.; Kim, S.J.; Chung, M.J. Development of An Active Head Eye System for Human-Robot Interaction. In Proceedings of the 39th International Symposium on Robotics, Seoul, Korea, 15–17 October 2008.
26. Mekuz, N.; Bauckhage, C.; Tsotsos, J.K. Face Recognition with Weighted Locally Linear Embedding. In Proceedings of the 2nd Canadian Conference on Computer and Robot Vision, British Columbia, Canada, 9–11 May 2005; pp. 290–296.
27. Eleyan, A.; Özkaramanli, H.; Demirel, H. Complex Wavelet Transform-Based Face Recognition. *EURASIP J. Adv. Signal. Process.* **2008**, *2008*, 1–13. [[CrossRef](#)]
28. Choi, W.-P.; Tse, S.-H.; Wong, K.-W.; Lam, K.-M. Simplified Gabor Wavelets for Human Face Recognition. *Pattern Recognit.* **2008**, *41*, 1186–1199. [[CrossRef](#)]
29. Albiol, A.; Monzo, D.; Martin, A.; Sastre, J.; Albiol, A. Face Recognition Using HOG–EBGM. *Pattern Recognit. Lett.* **2008**, *29*, 1537–1543. [[CrossRef](#)]
30. Anvar, S.M.H.; Yau, W.-Y.; Nandakumar, K.; Teoh, E.K. Estimating In-Plane Rotation Angle for Face Images from Multi-Poses. In Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management, Singapore, 16–19 April 2013; pp. 52–57.

31. Du, G.; Su, F.; Cai, A. Face Recognition Using SURF Features. In Proceedings of the 6th International Symposium on Multispectral Image Processing and Pattern Recognition, Yichang, China, 29–31 October 2009; pp. 749629–1–749628–7. [[CrossRef](#)]
32. Lee, W.O.; Kim, Y.G.; Hong, H.G.; Park, K.R. Face Recognition System for Set-Top Box-Based Intelligent TV. *Sensors* **2014**, *14*, 21726–21749. [[CrossRef](#)] [[PubMed](#)]
33. Cao, X.; Wei, Y.; Wen, F.; Sun, J. Face Alignment by Explicit Shape Regression. *Int. J. Comput. Vis.* **2014**, *107*, 177–190. [[CrossRef](#)]
34. Kazemi, V.; Sullivan, J. One Millisecond Face Alignment with an Ensemble of Regression Trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874.
35. Athana, A.; Zafeiriou, S.; Tzimiropoulos, G.; Cheng, S.; Pantic, M. From Pixels to Response Maps: Discriminative Image Filtering for Face Alignment in the Wild. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1312–1320. [[CrossRef](#)] [[PubMed](#)]
36. Chetverikov, D. GLDH Based Analysis of Texture Anisotropy and Symmetry: An Experimental Study. In Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994; pp. 444–448.
37. Chen, X.; Rynn, P.J.; Bowyer, K.W. Fully Automated Facial Symmetry Axis Detection in Frontal Color Images. In Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies, Buffalo, NY, USA, 17–18 October 2005; pp. 106–111.
38. Song, Y.-J.; Kim, Y.-G.; Chang, U.-D.; Kwon, H.B. Face Recognition Robust to Left/Right Shadows; Facial Symmetry. *Pattern Recognit.* **2006**, *39*, 1542–1545. [[CrossRef](#)]
39. Wang, F.; Huang, C.; Liu, X. A Fusion of Face Symmetry of Two-Dimensional Principal Component Analysis and Face Recognition. In Proceedings of International Conference on Computational Intelligence and Security, Beijing, China, 11–14 December 2009; pp. 368–371.
40. Klir, G.J.; Yuan, B. *Fuzzy Sets and Fuzzy Logic*; Prentice-Hall: Upper Saddle River, NJ, USA, 1995.
41. Adeli, H.; Sarma, K.C. *Cost Optimization of Structures: Fuzzy Logic, Genetic Algorithms, and Parallel Computing*; John Wiley and Sons: West Sussex, UK, 2006.
42. Siddique, N.; Adeli, H. *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*; John Wiley and Sons: West Sussex, UK, 2013.
43. Broekhoven, E.V.; Baets, B.D. Fast and Accurate Center of Gravity Defuzzification of Fuzzy System Outputs Defined on Trapezoidal Fuzzy Partitions. *Fuzzy Sets Syst.* **2006**, *157*, 904–918. [[CrossRef](#)]
44. Ross, T.M. *Fuzzy Logic. with Engineering Applications*, 3rd ed.; John Wiley and Sons: West Sussex, UK, 2010.
45. Mamdani, E.H. Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis. *IEEE Trans. Comput.* **1977**, *C-26*, 1182–1191. [[CrossRef](#)]
46. Zadeh, L.A. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 28–44. [[CrossRef](#)]
47. Ahonen, T.; Hadid, A.; Pietikäinen, M. Face Recognition with Local Binary Patterns. In Proceeding of the 8th European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 469–481.
48. FEI Face Database. Available online: <http://fei.edu.br/~cet/facedatabase.html> (accessed on 30 October 2015).
49. Minear, M.; Park, D.C. A Lifespan Database of Adult Facial Stimuli. *Behav. Res. Methods* **2004**, *36*, 630–633. [[CrossRef](#)]
50. AR Face Database. Available online: <http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html> (accessed on 30 October 2015).
51. The Japanese Female Facial Expression (JAFPE) Database. Available online: <http://www.kasrl.org/jaffe.html> (accessed on 30 October 2015).
52. Wolf, L.; Hassner, T.; Maoz, I. Face Recognition in Unconstrained Videos with Matched Background Similarity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 20–25 June 2011; pp. 529–534.
53. Lee, K.-C.; Ho, J.; Yang, M.-H.; Kriegman, D. Visual Tracking and Recognition Using Probabilistic Appearance Manifolds. *Comput. Vis. Image Underst.* **2005**, *99*, 303–331. [[CrossRef](#)]
54. Gorodnichy, D.O. Video-based Framework for Face Recognition in Video. In Proceedings of the 2nd Canadian Conference on Computer and Robot Vision, British Columbia, BC, Canada, 9–11 May 2005; pp. 1–9.

55. Dongguk Face Database (DFace-DB1). Available online: <http://dm.dgu.edu/link.html> (accessed on 29 April 2016).
56. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [[CrossRef](#)]
57. Turk, M.; Pentland, A. Eigenfaces for Recognition. *J. Cogn. Neurosci.* **1991**, *3*, 71–86. [[CrossRef](#)] [[PubMed](#)]
58. Li, S.Z.; Hou, X.W.; Zhang, H.J.; Cheng, Q.S. Learning Spatially Localized, Parts-based Representation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; pp. 207–212.
59. Kim, S.-K.; Park, Y.J.; Toh, K.-A.; Lee, S. SVM-based Feature Extraction for Face Recognition. *Pattern Recognit.* **2010**, *43*, 2871–2881. [[CrossRef](#)]
60. Froba, B.; Ernst, A. Face Detection with the Modified Census Transform. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 17–19 May 2004; pp. 91–96.
61. OpenCV. Available online: <http://opencv.org/> (accessed on 11 July 2016).
62. Wu, B.; Al, H.; Huang, C.; Lao, S. Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost. In Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 17–19 May 2004; pp. 79–84.
63. Rowley, H.A.; Baluja, S.; Kanade, T. Rotation Invariant Neural Network-Based Face Detection. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, USA, 23–25 June 1998; pp. 38–44.
64. Labeled Faces in the Wild. Available online: <http://vis-www.cs.umass.edu/lfw/> (accessed on 29 April 2016).
65. Kim, B.-S.; Lee, H.; Kim, W.-Y. Rapid Eye Detection Method for Non-glasses Type 3D Display on Portable Devices. *IEEE Trans. Consum. Electron.* **2010**, *56*, 2498–2505. [[CrossRef](#)]
66. YouTube Faces DB. Available online: <http://www.cs.tau.ac.il/~wolf/ytfaces/> (accessed on 29 April 2016).
67. The Honda/UCSD Video Database. Available online: <http://vision.ucsd.edu/~leekc/HondaUCSDVideoDatabase/HondaUCSD.html> (accessed on 29 April 2016).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).