Visual Observation Confidence based GMM Face Recognition robust to Illumination Impact in a Real-world Database

Anh Tuan TRAN†[1], Jin Young KIM†‡[2], Asmatullah CHAUDHRY†[3], The Bao PHAM†[1] Hyoung-Gook Kim[4]

[1] Faculty of Mathematics and Computer Science, University of Science Ho Chi Minh City, Vietnam,
[e-mail: tuantran261083@gmail.com; ptbao@hcmus.edu.vn]
[2]Department of Electronics and Computer Engineering, Chonnam National University
Gwangju, 500-757, South Korea
[e-mail: beyondi@jnu.ac.kr]
[3] HRD, PINSTECH, P.O. Nilore Islamabad, Pakistan
[e-mail: asmatullah.chaudhry@gmail.com]
[4]Department of Electronics Convergence Eng., Kwangwoon University
[e-mail:hkim@kw.ac.kr]
*Corresponding author: Jin Young KIM, Hyoung-Gook Kim

Received February 11, 2015; revised June 2, 2015; accepted March 3, 2016; published April 30, 2016

Abstract

The GMM is a conventional approach which has been recently applied in many face recognition studies. However, the question about how to deal with illumination changes while ensuring high performance is still a challenge, especially with real-world databases. In this paper, we propose a Visual Observation Confidence (VOC) measure for robust face recognition for illumination changes. Our VOC value is a combined confidence value of three measurements: Flatness Measure (FM), Centrality Measure (CM), and Illumination Normality Measure (IM). While FM measures the discrimination ability of one face, IM represents the degree of illumination impact on that face. In addition, we introduce CM as a centrality measure to help FM to reduce some of the errors from unnecessary areas such as the hair, neck or background. The VOC then accompanies the feature vectors in the EM process to estimate the optimal models by modified-GMM training. In the experiments, we introduce a real-world database, called KoFace, besides applying some public databases such as the Yale and the ORL database. The KoFace database is composed of 106 face subjects under diverse illumination effects including shadows and highlights. The results show that our proposed approach gives a higher Face Recognition Rate (FRR) than the GMM baseline for indoor and outdoor datasets in the real-world KoFace database (94% and 85%, respectively) and in ORL, Yale databases (97% and 100% respectively).

Keywords: GMM-based face recognition, Visual Observation Confidence, Flatness Measure, Centrality Measure, Illumination Normality Measure

This research was supported by a research grant from NRF, the Korean government [2009-0077345] and .the research grant of Kwangwoon University in 2015.

http://dx.doi.org/10.3837/tiis.2016.04.020 **ISSN : 1976-7277**

1. Introduction

In biometric recognition, auditory or visual information such as voice color, iris shape, fingerprint, finger vein pattern and facial structure are main clues for person recognition. Recently, a great deal of performance enhancement has been achieved and biometric systems show almost perfect performance for ideal auditory or visual signals without noises, channel distortions, or visual distortions. However, real biometric signals suffer from moderate or severe distortions. In visual cases, visual additive and quantization noises or various illumination mismatches are the main causes of performance degradation.

In the face recognition area, many feature-based or holistic approaches have been proposed (Nixon, 1985 [14]; Reisfeld, 1994 [16]; Graf et al ., 1995 [5]; Nallammal and Radha, 2012 [13]; Demers and Cottrell, 1993 [2]; Li et al., 2004 [12]). However, they still have a problem with robustness with a mismatched or biased illumination condition. Particularly, deep shadows and highlights are typical factors that can significantly degrade the system performance. To overcome the illumination problem, two types of approaches have been suggested. The first approach involves obtaining shadow-free or highlight-free images, which can yield a higher recognition rate. Shadow-free and highlight-free approaches include oriented local histogram equalization (Lee et al., 2012 [11]), lighting aware preprocessing (LAP) method (Han et al., 2010 [6]), and the shadow compensation method based on Fourier analysis (Choi and Jeong, 2011 [1]). The second approach involves obtaining illumination-robust features. A representative approach was proposed by Sanderson et al. (2005) [19]. They applied discrete cosine transform (DCT) and added more delta coefficients to feature vectors: DCT-mod, DCT-mod-delta, and DCT-mod2. While the results were promising, most illumination changes were artificial and do not represent most situations in the real world.

Lately, a new idea was developed based on observation confidence in the speaker identification area. Kim et al. (2007) [10] introduced signal confidence concepts and suggested an auditory confidence measure embedded in GMM-based speaker identification. The key idea is to manage frame-based feature vectors unequally in the GMM training and recognition stages. This is because each frame suffers from a different amount of distortion in stationary noise environments. The different amount of distortion is measured in segmental SNR. The observation confidence acts as a weighting factor for observation probability. Also, Jiang (2005) [8] presented a survey on observation confidence. However, all of the mentioned methods are commonly applied for audio processing. In video processing, the observation confidence is considered less than audio confidence. That's because it is not easy to define the VOC due to the complexity of the illumination problem. For this reason, the confidence values for evaluating observations in a facial image are necessary to improve the system performance.

In some novel approaches, the approaches of using graph theory and statistic theory to recognize a face become more general. Such as in a research of the author M.Kafai in 2014 [22], he proposes to apply a Reference Face Graph (RFG) where a reference face is a node representing a single individual. Each reference face has multiple images with various poses, expressions, and illumination. Obviously, we can see that the performance of this approach depends on a large number of images with various poses, expressions, and illumination for constructing a basis set of RFG. In real world database, we may not have enough such a number of images with various poses, expressions, and illumination. About statistic theory,

the author Yi Sun mentions an approach to use Joint Bayesian based on Deep hidden IDentity features (DeepID) [23]. Highly compact 160-dimensional DeepID at the end of the cascade that contain rich identity information and directly predict a much larger number (e.g., 10, 000) of identity classes. But in this paper, the author does not mainly discuss about illumination. He just focus on the problem with a large number of face identities. In 2015, another author A. Punnappurath and A.N. Rajagopalan propose a methodology for face recognition in the presence of space-varying motion blur comprising of arbitrarily-shaped kernels [24]. The illumination in their paper is handled by estimating 9 light source directions from minimizing a cost function. This approach seems to be effective but requires the consuming time for minimizing a cost function to determine 9 light source directions.

In this paper, we propose a visual observation confidence (VOC) to counteract the impact of shadows and highlights in a real-world database. The VOC is calculated for the decomposed blocks of a face image. The proposed VOC considers flatness, distance from the face center, and intra-class variance as measures for confidence estimation. These correspond to the Flatness Measure (FM), Centrality Measure (CM), and Illumination Normality Measure (IM), respectively. Subsequently, the VOC values accompany the feature vectors in the modified-GMM training process to determine the subject optimal models.

The remainder of this paper is organized as follows. In Section 2, we describe our main idea about the VOC-based approach with FM, CM, and IM. Section 3 is an overview of the proposed GMM based recognition system. Some modifications in the EM algorithm and classification process using VOC are also mentioned. In Section 4, some experiments on the KoFace, ORL, and Yale databases are discussed to evaluate the contributions of measurements (FM, CM, and IM) in our recognition system. Concluding remarks are presented in Section 5.

2. The visual observation confidence-based approach

Observation confidence is a measure representing the degree of reliability of the signal. As we discussed in the introduction, signals are corrupted by outer interferences. In the visual signal case, additive noises, biased illuminations, and coding noises are causes of signal corruption. Even though a captured signal is free from any corruption, visual image blocks may not be reliable at the point of discrimination ability while performing face recognition. For example, image blocks from hairs or cheeks are in-discriminable compared with those of the eyes, nose, and lips areas. Therefore, in the face recognition domain, VOC refers to the differentiating power of image blocks between subject faces. In this paper, the main aim is to enhance performance degradation from biased or excessive illumination. We consider the three factors of flatness, centrality, and illumination normality for estimating VOC as indirect measurements. The idea of flatness derives from spectral flatness or Wiener entropy. Centrality measures the distance between each image block and the face center. The idea of centrality is developed based on the fact that outer areas of the face contains less discriminant features more than inner areas of a human face. Regarding illumination normality, we need to decide that image blocks belong to highlight or deep-shadow regions. That is, illumination normality checks the rate of excessiveness or deficiency of lighting. For flatness, centrality, and illumination normality, we develop FM, CM, and IM respectively. Therefore, the final VOC is represented by a linear combination of the three measures,

$$VOC = w_{FM}*FM + w_{CM}*CM + w_{IM}*IM \quad (1)$$

where $w_{FM} + w_{CM} + w_{IM} = 1$. This constraint ensures that the VOC will be in the range 0 and 1 when FM, CM and IM are in the same range.

2.1 Flatness Measure

Discrimination measurement reflects the ability to recognize the difference between two or more faces. Based on the observations, some facial features such as the eyes, nose, mouth, cheekbones, and jaw are distinctive features that differ in each person. In addition, it is easily noticed that most of these distinctive features are present in high frequency or uneven areas within the human skin region. Also, deficient or excessive lighting makes the face look flatter.

Kim *et al.* (2004) [9] proposed the Flatness Measure (FM), which is useful for calculating the flatness degree of grayscale intensity in an image as follows:

$$FM = \frac{1}{1+e^{-Flatness\ Value}} \tag{2}$$

$$Flatness\ Value = 10\log_{10}\frac{G_m}{A_m},\ G_m = \left[\prod_{(x,y)\in block} Lu(x,y)\right]^{\frac{1}{T}},\ and\ A_m = \frac{1}{T}\left(\sum_{(x,y)\in block} Lu(x,y)\right) \tag{3}$$

where G_m and A_m are the corresponding geometric and arithmetic means of a block, respectively, while T and Lu are the number of pixels in a block and luminance channel Y in YCbCr, respectively. The sigmoid function in Eq. (2) aims to refine all FM values into the range of 0 to 1. Using the sigmoid function, we the change flatness measure from the binary selection (Kim *et al.*, 2004) to a measurement as a confidence value in our paper.

After we calculate the FM value on each block in a facial image, the blocks that have a higher FM value have less distinctive ability, while a lower FM value indicates more distinctive ability than other blocks. **Figs. 1** and **Figs. 2** describe the ability of FM to represent the discrimination information of a face. **Figs. 1** shows the FM value in an image representation and **Figs. 2** shows the FM value after cutting out 20%, 40%, 60%, and 80% of the blocks with the highest FM values.

2.2 Centrality Measure

The main components of the face are the eyes, nose, and lips, all of which are located around the center of the face. Even though the face outline gives us information about the identity of the face, the eyes, nose, and lips are more distinctive for face identification. Therefore, face image blocks around the center should be treated with more weight than the outer face blocks such as the hair and neck. Also, the outer regions are easily distorted by illumination compared

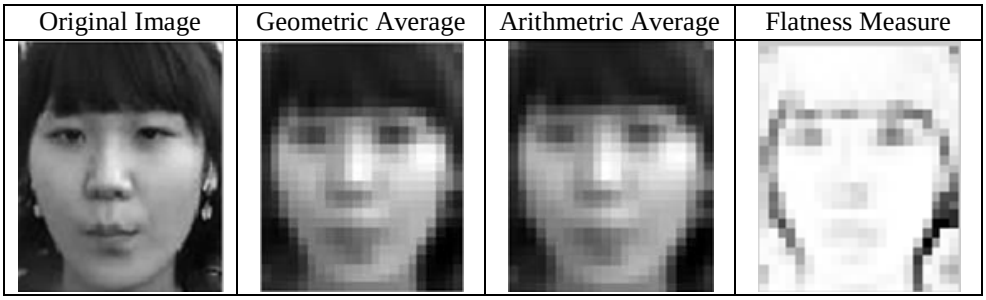

Fig. 1. Y image – G_m image – A_m image – FM image representation

Flatness Measure	20%	40%	60%	80%
				

Fig. 2. Cutting blocks based on FM values. Blocks with high value are cut, and the remaining blocks are characteristic blocks.

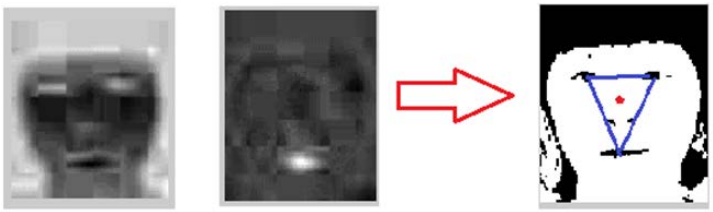

Fig. 3. Eye and mouth map – the center of the face.

with the center regions. Thus, we can say that face image blocks near the face center are more specific for face recognition.

The center of a face can be found by detecting the eyes and mouth based on the chrominance and luminance properties in the YCbCr color space.

$$EM = \left(\frac{1}{3}\left(C_b^2 + (1 - C_r)^2 + \frac{C_b}{C_r}\right)\right) \times \left(\frac{Y_{dil}}{Y_{er} + 1}\right) \tag{4}$$

$$MM = C_r^2 . \left(C_r^2 - \eta . \frac{C_r}{C_b}\right)^2 \text{ with } \eta = 0.95 \frac{average(C_r^2)}{average(C_r / C_b)}$$

Here, Y_{dil} and Y_{er} are the dilation and erosion in the luminance channel Y. Eq. (4) was suggested by Hsu (2002) [7]. **Fig. 3** describes the result of the eye and mouth map.

In YC_bC_r color space, the luminance Y contains mainly illumination information. So the EM (Eye Map) will be affected by illumination because it has Y dilation and Y erosion. However, since we make the division of Y dilation over Y erosion in Eye Map equation, the illumination effects will be reduced. About the MM (Mouth Map), since it only has Cb and Cr in its equation, so, MM will be not much effected by illumination.

In order to construct an eye and mouth map, we aim to find the center point of the face; then, we can define the CM based on this point (**Fig. 3**). In order to refine this value to a range between 0 and 1, we divide the Euclidean distance of the center point of a block (P_{block}) and the center point of a face (P_{center}) by half of the diagonal length of the face. This diagonal length is the Euclidean distance of the top-left point ($P_{top\text{-}left}$) and bottom-right point ($P_{right\text{-}bottom}$) in a rectangle of the face (see Eq. (5)). **Fig. 4** shows that CM helps to remove further unnecessary blocks in the hair, neck, and background areas than **Fig. 2**.

$$CM = \frac{2 * \left\| P_{block} - P_{center} \right\|}{\left\| P_{top-left} - P_{right-bottom} \right\|} \tag{5}$$

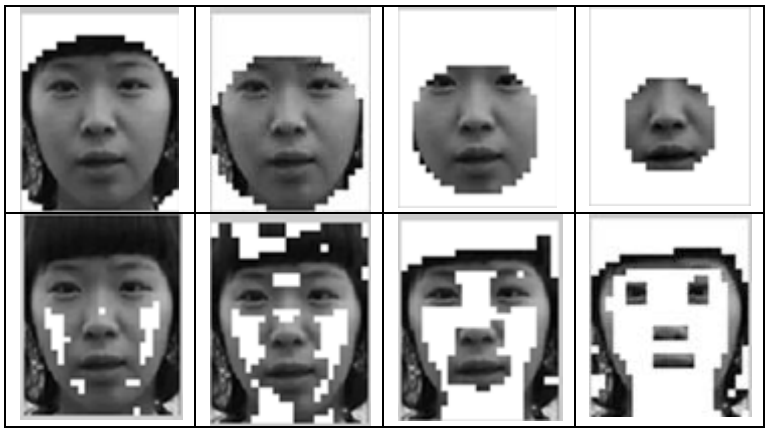

Fig. 4. Block representation in cutting 20%, 40%, 60%, and 80% of the blocks. The first row images are block representations in CM and the second row images are block representations in both CM and FM.

2.3 Illumination Normality Measure

In a real-world database, it is more robust to define a measurement to evaluate exactly how much illumination affects a face than modifying or changing the content of the image. This is because we cannot control the level of accuracy when modifying or changing the content of the image for compensation or reduction.

The illumination in the KoFace database is quite diverse. Shadows and highlights are commonly representative effects in this database. Directly applying a recognition algorithm without considering illumination will result in poor performance. Instead, we need an evaluation measurement.

Let IM be an illumination normality measure. Three steps are needed to measure illumination effectively and robustly. Firstly, we utilize intra-class variance to construct an intra-class variance graph (Otsu, 1979) [15]. Secondly, based on the graph, we detect shadow and highlight thresholds to segment their regions. Lastly, we define the IM value from the shaded and highlighted regions from the segmentation step.

In Eq. (6), the intra-class variance f_{IV} is employed in the luminance channel Y based on the bimodal histogram of the background and foreground class. A threshold t varies from 0 to 1 to compute this variance. All pixels in the background class have intensities less than t, whereas the other pixels belong to the foreground class.

$$f_{IV}(t) = p_{foreground}(t)\sigma^2_{foreground}(t) + p_{background}(t)\sigma^2_{background}(t) \tag{6}$$

where p and σ represent the probability and variance of each class, respectively.

Fig. 5 shows a graph of f_{IV}. Analyzing this graph will help to detect the shadow and highlight points or thresholds for segmentation. Generally, the intra-class variance graph consists of 5 different parts. Assume that the background is the 1st class and the foreground is the 2nd class. The shape of each part is explained as follows:

- Part 1: the 1st class consists of hair or eye areas. Therefore, the variance of the 1st class is 0, and f_{IV} is only the variance of the 2nd class.

- Part 2: the 1st class includes more deep shadows, and its variance increases slightly. In contrast, the 2nd class loses these areas and the variance decreases substantially. This leads the decreasing of f_{IV}.

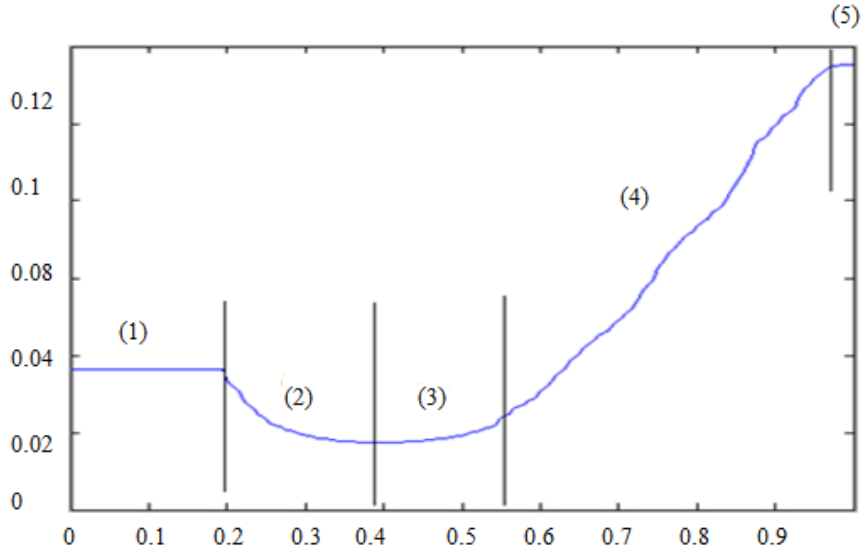

Fig. 5. Intra class variance of background and foreground class intuitively divided into 5 parts.

- Part 3: the 1st class includes more soft shadows, and its variance increases substantially because soft shadows are quite distinctive from the eye, hair or deep shadows. In this case, the variance of the 2nd class loses only a soft shadow and its variance decreases slightly. As a result, f_{IV} increases.

- Part 4: the 1st class includes a larger skin area, so its variance increases significantly, while the 2nd class only loses some skin area. Therefore, f_{IV} increases significantly.

- Part 5: The 1st class includes all the skin areas and this leads to a slightly increasing variance of the 1st class. The variance of the 2nd class will become 0, because it only has highlight areas. In this case, f_{IV} is the variance of the 1st class, and has quite a stable value.

It is obvious that Part 4 contains all of the human skin information. Therefore, if we can find the beginning and end points of this part, the shadow and highlight points or thresholds will be these points. To do that, we firstly find the minimum and maximum points (t_{min} and t_{max}) of the graph, and the inflection point ($t_{inflection}$) by the first and second-order differential operation. And then, the slope comparison will help to obtain the shadow and highlight threshold as in Eq. (7) where $\varepsilon = 0.05$.

$$
\begin{aligned}
t_{shadow} &= \{\, t_1 \in [t_{\min}, t_{\text{inflection}}] \\
&\quad \left| \frac{f_{IV}(t_1) - f_{IV}(t_{\text{inflection}})}{t_1 - t_{\text{inflection}}} \approx \frac{f_{IV}(t_{\text{inflection}}) - f_{IV}(t_{\text{inflection}} - \varepsilon)}{\varepsilon} \right\} \\
t_{highlight} &= \{\, t_2 \in [t_{\text{inflection}}, t_{\max}] \\
&\quad \left| \frac{f_{IV}(t_{\text{inflection}}) - f_{IV}(t_2)}{t_{\text{inflection}} - t_2} \approx \frac{f_{IV}(t_{\text{inflection}} + \varepsilon) - f_{IV}(t_{\text{inflection}})}{\varepsilon} \right\}
\end{aligned}
\tag{7}
$$

The segmentation is very important for defining the illumination normality measure. If a block has a luminance average less than the shadow point or higher than the highlight point, the block is labeled as an illumination-affected block. In order to define the IM value, we employ the Tukey function as a mapping function, as in Eq. (8):

$$f_{tukey}(x)=\begin{cases}\frac{1}{2}(1+\cos(\frac{2\pi}{r}(x-\frac{r}{2}))) & 0\le x<\frac{r}{2}\\ 1 & \frac{r}{2}\le x<1-\frac{r}{2}\\ \frac{1}{2}(1+\cos(\frac{2\pi}{r}(x-1+\frac{r}{2}))) & 1-\frac{r}{2}\le x\le 1\end{cases} \tag{8}$$

In this function, the first and last r/2 percent of samples is equal to parts of a cosine function. r/2 represents the shadow and highlight points (t_{shadow} and $t_{highlight}$), as shown in **Fig. 6**.

Finally, IM is defined by Eq. (9) to reverse the value from the f_{Tukey} function, because a higher IM value means a greater illumination effect on a block.

$$IM=\frac{\max f_{Tukey}-f_{Tukey}(A_m)}{\max f_{Tukey}} \tag{9}$$

where A_m is the arithmetic mean of each block (in Eq. (3)).

We performed two experiments on two images. One image is a face affected by highlight illumination and the other image is a face affected by shadow illumination. The IM value is calculated for each image and 20%, 40%, 60%, and 80% of the blocks with the highest IM value are cut off. The results are shown in **Figs. 7** and **Figs**. **8**.

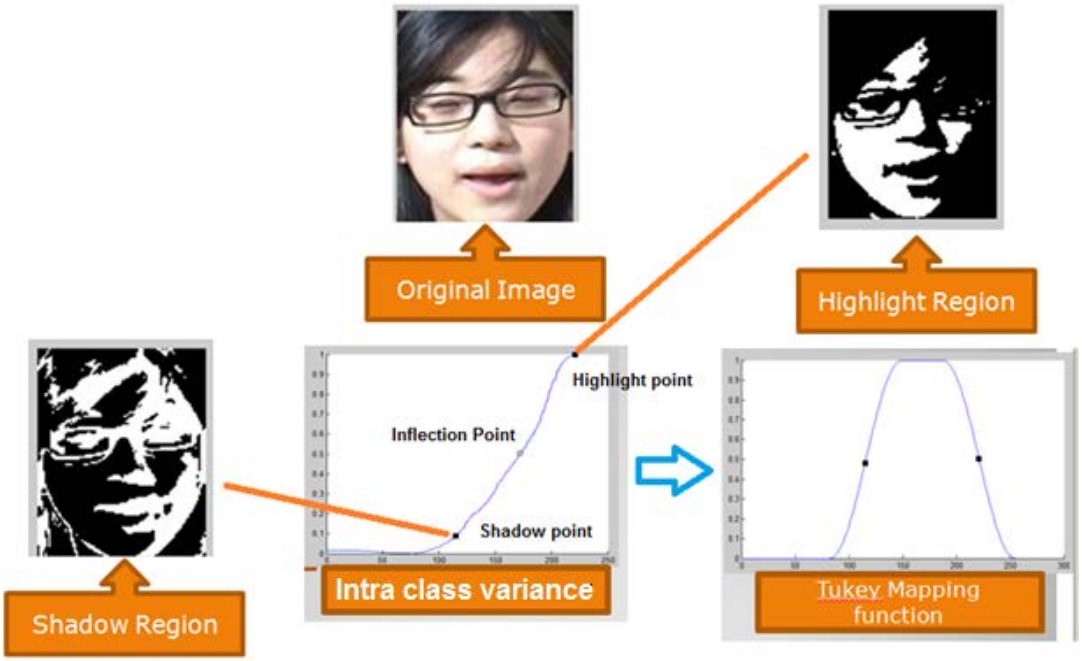

Fig. 6. Tukey function maps intra-class variance into IM, based on shadow and highlight points.

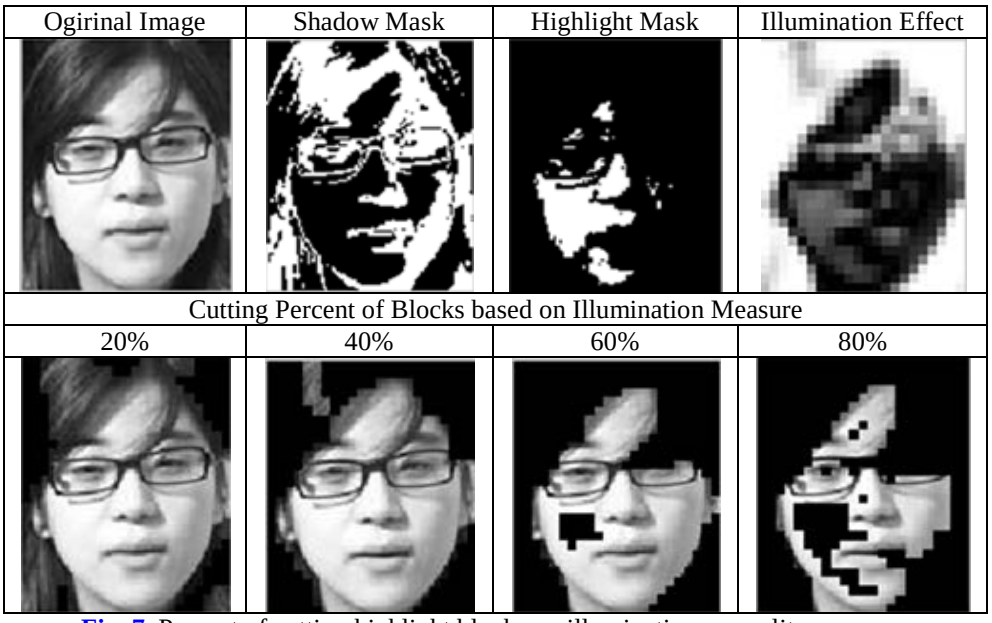

Fig. 7. Percent of cutting highlight blocks on illumination normality measure.

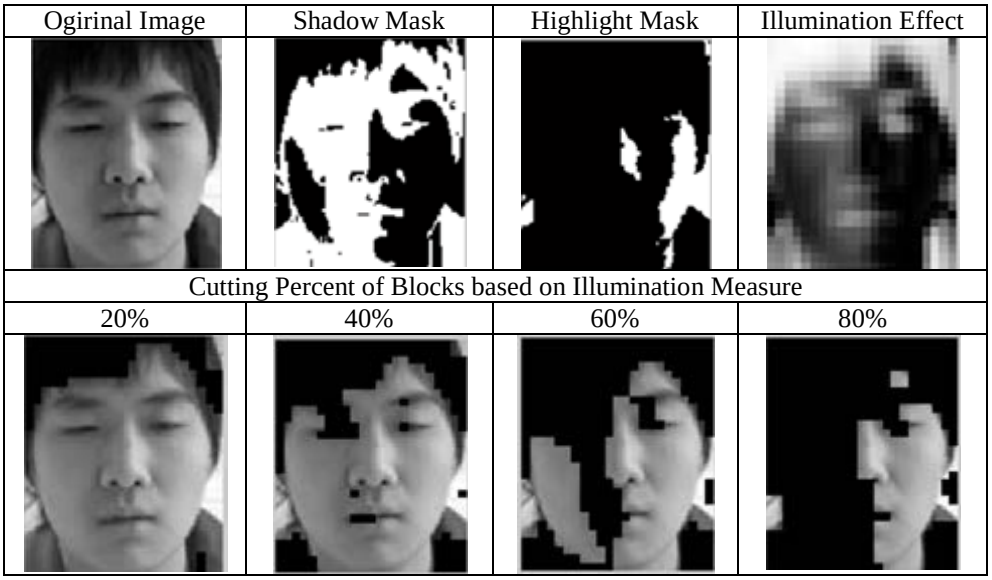

Fig. 8. Percent of cutting shadow blocks on illumination normality measure.

3. The Proposed GMM-based Face Recognition System

3.1 Face Recognition Block Diagram using VOC

For the purpose of this paper, we would like to build a system which is robust to illumination changes while achieving a high recognition rate. The main contribution in this approach is that we define an optimum combination of FM, CM, and IM representing the discrimination, distance, and illumination information respectively, for each feature vector.

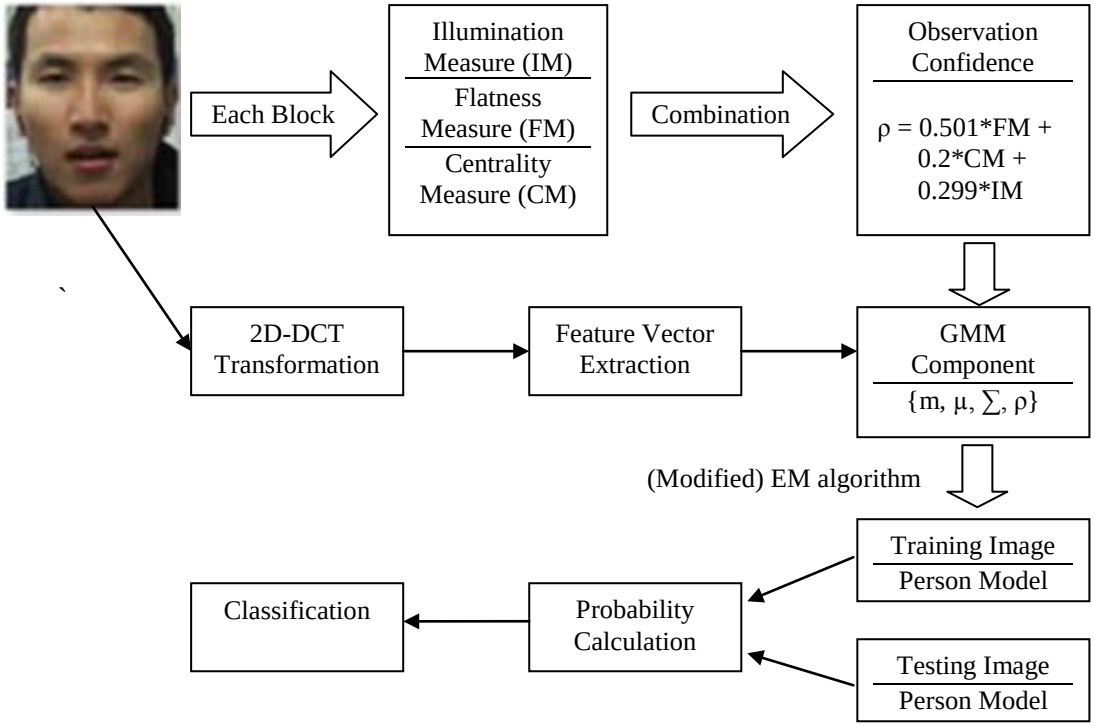

Fig. 9. Block diagram of the proposed face recognition system

Based on the traditional GMM recognition process, we add further modifications in order to integrate the visual observation confidence value into the system. The overall block diagram is shown in **Fig. 9.** In the diagram, the input image is a facial image in the RGB color space. Therefore, we transform the image into YCbCr, take the luminance channel as a grayscale image for the 2D-DCT transformation and computation of the FM and IM values, and take the chrominance channel to find the CM values.

About optimal GMM parameters, we need to collect all GMM components for the training process. In our approach, GMM components are feature vectors accompanying the VOC value. Following the process of EM estimation, one training and one testing face model are calculated. Using probability calculation on the training and testing face models, a person can be classified into a corresponding face subject. The VOC value is the combination following Eq. (1) with weight values α, β, and γ as 0.501, 0.2, and 0.299, respectively. These values are determined by a VOC optimization process which is mentioned in Section 4.2. In addition, further detail about each step in this diagram will be described in the next section.

3.2 Preprocessing Step

The KoFace database is composed of facial images, which are taken in a real-world environment. Each image contains a human face with other unnecessary information in the background. As such, we need a preprocessing step to segment the human face, and draw a rectangle around it, and scale it to the same size for block-based feature extraction.

Fig. 10. Face extraction from the whole input image.

Human face segmentation can be treated as skin-tone segmentation using a special color space. YCbCr is a familiar color space used to detect and segment skin-tone information because the skin color is more compact in this space. Some constraints are defined based on luminance and chrominance information in order to cluster all skin color information (Garcia and Tziritas, 1999; Hsu *et al.*, 2002) [4]. The largest skin-tone region is determined as the main human face in single face images. After the human face is detected, we draw a rectangle around it. And then, we need to normalize all rectangles into the same size. In the KoFace database, the distance from the camera to each face is nearly the same, so the normalization does not significantly affect each face structure. Our desired size for each face after normalization is 92 x 112 pixels (**Fig. 10**).

3.3 Feature Extraction

The feature extraction stage plays an important role in recognition systems because a good extracted feature vector will give a good representation that will provide good training for classification models. The Discrete Cosine Transform (DCT) was suggested as an efficient and robust approach to give a compact feature representation and desired dimensionality reduction (Sanderson *et al.*, 2005 [19]; Ekenel and Stiefelhagen, 2006 [3]). In our approach, we perform the DCT transform on blocks divided from a facial image.

Each image is divided block by block with a size of 8 x 8 pixels in an overlap of neighboring blocks by 50%. **Fig. 11** shows an example of overlapping blocks in 8x8 red squares. If we denote *b(x,y)* as a block in the location *(x,y)* and *N = 8* as the size of blocks, the 2-D DCT of a block is defined as follows.

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} b(x,y)\cos[\frac{(2x+1)u\pi}{2N}]\cos[\frac{(2y+1)v\pi}{2N}] \quad (10)$$

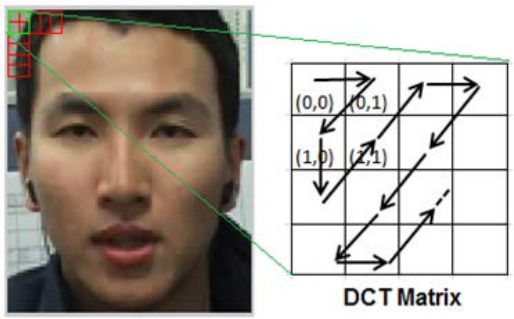

Fig. 11. An overlap of neighboring blocks by 50% in red squares – Zigzag scanning pattern to order values in terms of high frequency (or discrimination information).

For $u, v = 0, 1, \ldots, N-1$ where $\alpha(v) = 1/N$ for $v = 0$, $\alpha(v) = 2/N$ for $v = 1,2,\ldots N-1$.

After DCT transformation, zig-zag scanning is processed to collect all values for a feature vector in one block to arrange values in order of high frequency (or discrimination information). In **Fig. 11**, the zig-zag scanning pattern is described and we choose $M = 18$ highest values as an M-dimensional local feature vector for one block.

3.4 EM Algorithm using VOC

The expectation-maximization (EM) algorithm (Xu and Jordan, 1996 [21]) is a well-known iterative parameter estimation scheme used to find the maximum-likelihood estimation of parameters in statistical models. One main modification is to apply the visual observation confidence ($\rho_n = VOC_n$) accompanying each feature vector in the M-step (Kim *et al.*, 2007 [10]). This combination helps to focus on these characteristic feature vectors and to reduce the effects of unnecessary or illumination-affected feature vectors to find optimal parameters.

Expectation step: The posteriori probability for component i is calculated.

$$\Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j) = \frac{\omega_i^j f(x_n \mid \mu_i^j, \Sigma_i^j)}{\sum_{k=1}^{M} \omega_k^j f(x_n \mid \mu_k^j, \Sigma_k^j)} \tag{11}$$

Maximization step: Mixture weights, mean vectors, and covariance matrices are updated as:

$$\omega_i^{j+1} = \frac{\sum_{n=1}^{N} \rho_n \Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j)}{\sum_{i=1}^{M}\sum_{n=1}^{N} \rho_n \Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j)} \tag{12}$$

$$\mu_i^{j+1} = \frac{\sum_{n=1}^{N} \rho_n x_n \Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j)}{\sum_{n=1}^{N} \rho_n \Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j)} \tag{13}$$

$$\Sigma_i^{j+1} = \frac{\sum_{n=1}^{N} \rho_n \Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j)(x_n - \mu_i^{j+1})(x_n - \mu_i^{j+1})^T}{\sum_{n=1}^{N} \rho_n \Pr(i \mid x_n, \omega_i^j, \mu_i^j, \Sigma_i^j)} \tag{14}$$

3.5 GMM-based Classifier using VOC

The Gaussian Mixture Model is a type of density model that comprises a certain number of Gaussian functions. Using GMM, feature vectors can be modeled to perform real-time face recognition. Basically, we divide an image into overlapping blocks and generate feature vectors by DCT transformation. Then, GMM can be defined as the following equation.

$$\lambda = \{m_i, \mu_i, \Sigma_i\}, i = 1, \ldots, N_M \quad and \quad p(x \mid \lambda) = \sum_{i=1}^{N_M} m_i f(x \mid \mu_i, \Sigma_i) \tag{15}$$

where x is a feature vector, λ is the GMM model, and N_M is the number of GMM components. The density function $f(x|\mu_i,\Sigma_i)$ can be represented as a D-variate Gaussian pdf with μ_i and Σ_i:

$$f(x \mid \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right\} \tag{16}$$

In Eq. (15), the value m_i represents how much data exists on each Gaussian component satisfying $\sum_{i=1...M} m_i = 1$ that can be derived from K-means on training feature vectors.

In the GMM approach, for a given set of N independent observations $\boldsymbol{X}=\{x_i\}_{i=1..N}$, an objective function (or GMM likelihood) represents the degree of possibility that observation $\boldsymbol{X}$ matches the GMM model λ, which is trained from the training process as follows.

$$L(X \mid \lambda) = \log(p(X \mid \lambda)) = \sum_{i=1}^{N} \log(p(x_i \mid \lambda)) \tag{17}$$

Eq. (17) shows that the objective function treats all observations equally. The summary of probabilities means that VOC values are not considered. In the case of no illumination effect, the recognition rate is rather high and accurate. However, in the case of high or low illumination effects, some observations are severely contaminated, and the equation has no information about observation confidence information to perform sufficiently and effectively. In order to address this issue, we suggest adding VOC values ($\rho_n = VOC_n$) to each observation, which reflects how much it contributes to the recognition results under the influence of illumination as follows.

$$L(X \mid \lambda) = \log(p(X \mid \lambda)) = \sum_{n=1}^{N} \rho_n \log(p(x_n \mid \lambda)) \tag{18}$$

After the process of EM estimation, one training face model and one testing face model are calculated. We utilize these to calculate the probability and to determine the classified m^{th} person using the equations presented.

$$P^* = \arg\max_m L(X \mid \lambda_m) \tag{19}$$

4. Experimental results

4.1 KoFace Database and Dataset Construction

The KoFace database is composed of 106 face subjects. Each face subject has 5 standard images (i.e., no illumination), 10 indoor, and 10 outdoor images under different illumination conditions from very low to high shadow and highlight effect for testing datasets (**Fig. 12**).

Since KoFace is our constructed real-world database, most of the experiments were conducted on this database. The ORL and Yale databases are well-known databases and we only use them in the comparison of the face recognition rate with the GMM baseline algorithm. As we know, the ORL and Yale databases contain grayscale images, for that reason, the center point cannot be automatically detected by chrominance information. To overcome this problem, we manually locate the center point by clicking on each image with a mouse.

The KoFace database contains three types of conditions and three types of photograph contexts for each face subject. The lighting conditions are moderate (standard), low (shadow), and high (highlight). The photograph contexts are standard (little illumination effect), indoor (normal illumination effect), and outdoor (diverse illumination effects) as in **Figs. 13–15**

Fig. 12. Indoor and outdoor facial images in the KoFace database under diverse illumination conditions.

Fig. 13. Standard facial images under moderate illumination conditions.

Fig. 14. Indoor facial images under low illumination conditions.

Fig. 15. Outdoor facial images under high illumination conditions.

In **Table 1**, we describe the datasets for all experiments conducted in the KoFace, Yale, and ORL database. In the KoFace dataset, we have 40 face subjects and, for each, we choose 5 standard images for the GMM models training process. We also choose 5 indoor images and 5 outdoor images for visual observation confidence optimization in Section 4.2. In testing, when we have GMM models and VOC optimization combination, the other 5 indoor and outdoor images for each face subject are employed in the recognition process.

4.2 Visual Observation Confidence Optimization

Under this requirement, we need to combine FM, CM, and IM measurements in a linear manner to construct such a visual confidence value. A large dataset used in this process includes 40 face subjects in the KoFace database. With 5 indoor and 5 outdoor images, overall,

we have 400 facial images for this VOC optimization.

Table 1. Dataset description for experiments on the Yale, ORL, and KoFace databases.

Dataset	Face Subjects	Image Type		GMM Models Training	VOC Optimization	Testing	Total Image for each face subject
KoFace	40	Each Face Subject	Standard	5			5
			Indoor		5	5	10
			Outdoor		5	5	10
ORL	40		Grayscale	5		5	10
Yale	15		Grayscale	5		5	10

In determining an optimum combination, we perform a complete search by setting weight values w_{FM}, w_{CM}, and w_{IM} for FM, CM, and IM, respectively, from 0.001 to 0.999 with an interval of 0.05, in conjunction with the face recognition rate.

Follow the condition $w_{FM} + w_{CM} + w_{IM} = 1$, there will then be 210 combinations of three measurements for each block in 594 blocks on one face subject. Similarly, a massive implementation is continuously applied to other face subjects within the 40 face subjects in our dataset. The GMM-face recognition process is performed using these combinations to determine which combination gives the highest recognition rate. As a result, we found that the $w_{FM} = 0.501$, $w_{CM} = 0.2$, and $w_{IM} = 0.299$ is optimal with recognition rate 98.6667% (**Fig. 16**).

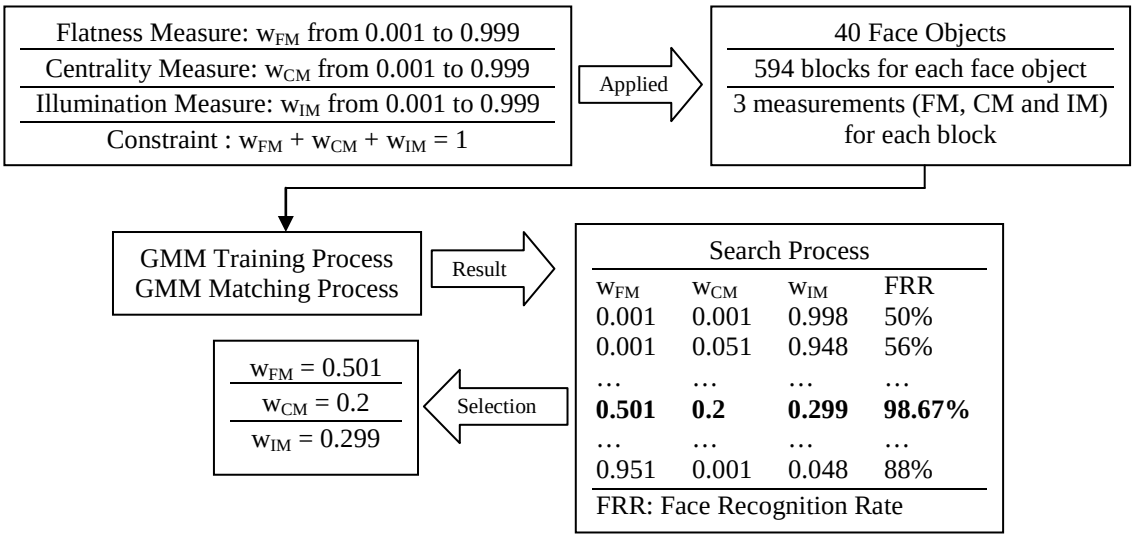

Fig. 16. A complete search process to find an optimum combination of FM, CM, and IM for the VOC

Based on the optimum search process, we defined a visual confidence value from three measurements. According to Eq. (1), the values α, β, and γ are exactly the weight values of w_{FM}, w_{CM}, and w_{IM}, respectively, thus, we have Eq. (20) as a resultant VOC equation:

$$VOC = 0.501 * FM + 0.2 * CM + 0.299 * IM \tag{20}$$

4.3 Face Recognition Performance on Proposed Approach

In this section, we compare our proposed approach using visual observation confidence and the GMM baseline approach, and we examine the face recognition performance based on the Yale, ORL, and KoFace databases.

Table 2 shows that our approach has a higher face recognition rate (FRR) than the GMM baseline approach on all three databases. While the increasing rates are only 1.33% and 1% in the cases of the Yale and ORL databases, respectively, the increasing rate is particularly high at 22.66% on the KoFace database in comparison with the GMM baseline rate. When the GMM baseline was applied to the KoFace database, the FRR was low at 74.67%, but when we utilized FM, CM, and IM, the result was accurate at 97.33% FRR because of the highly diverse illumination in our database. **Fig. 17** illustrates the high face recognition rate (FRR) with standard and real-life databases in a horizontal bar graph. All of them mean that our proposed approach can work better on a real-world database than the GMM baseline approach.

In **Fig. 18**, the Error Rate Reduction graph is shown based on the degree of error decreasing from the GMM baseline to the proposed approach. Error Reduction is calculated by the difference of error rate in baseline method and error rate in proposed method over the error rate in baseline method.The figure shows a high error reduction of 89.46% and 100% in the KoFace and Yale databases, respectively. These values mean that the illumination is very well controlled by the visual observation confidence in these databases. The 25% error reduction in ORL is also a high error reduction rate, and is acceptable because the FRR in this case increases from 96% to 97%.

In order to examine the effectiveness of GMM/VOC under illumination effects and the its comparison to the best other approaches, we conduct three experiments on three dataset (Case 1, Case 2, Case 3) under three different illumination effect conditions. We choose database Extend Yale-B database for the experiments because it's lighting variety and diversity. Each dataset contains 10 images (5 for training and 5 for testing) for each face subject within 30 face

Table 2. Comparison between the GMM baseline and proposed approach on three databases.

Database	GMM Baseline	Proposed Approach	Error Reduction
Yale	98.67%	100%	100%
ORL	96%	97%	25%
KoFace	74.67%	97.33%	89.46%

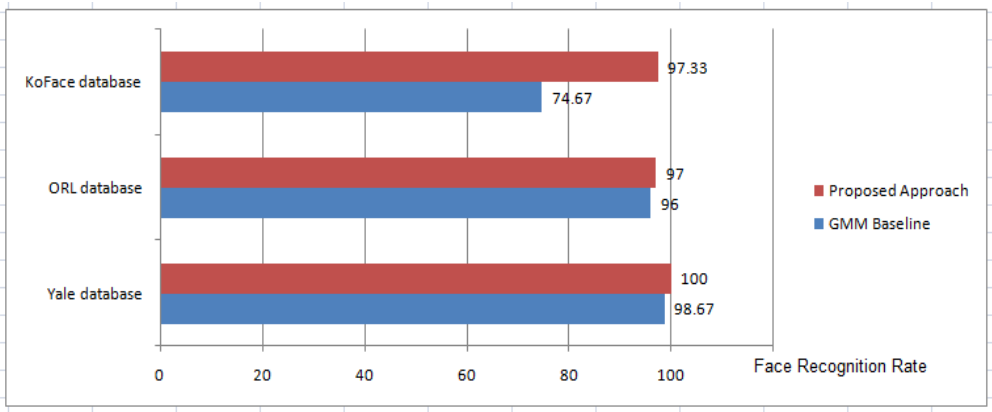

Fig. 17. FRRs on three datasets (KoFace, ORL, and Yale) are illustrated on a horizontal bar graph.

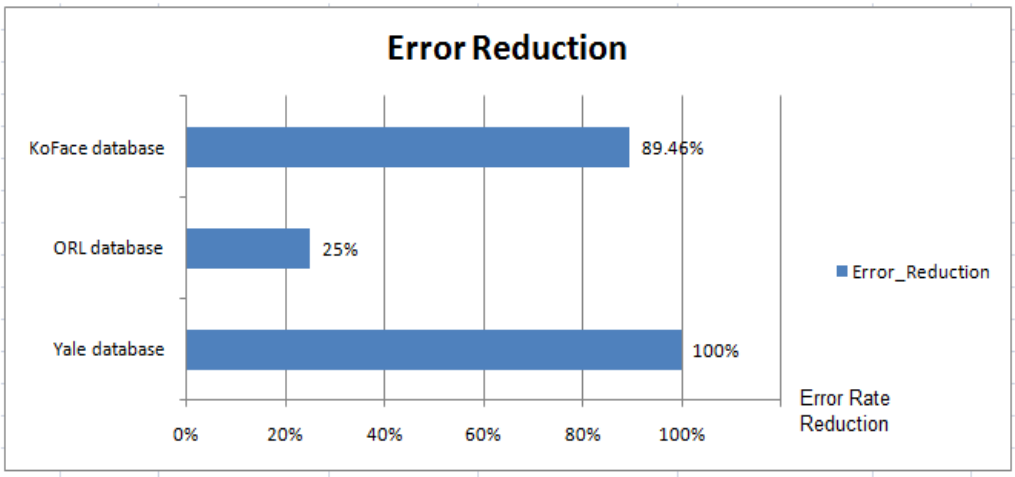

Fig. 18. Error reduction in FRR for each database of the proposed approach and GMM baseline.

subjects. In case 1, the effect of lighting is higher than case 2 and 3. In case 2, the lighting effect changes from low to high and, in case 3, there is a little effect of lighting. In the comparison with other approach, we choose the famous approach named PP + LTP/DT where PP is pre-processing process and LTP/DT is Local Ternary Patterns with distance transform-based similarity metric (Tan and Triggs, 2010 [20]).

From the **Table 3**, we realize that the GMM baseline cannot give a good recognition rate under high illumination effects (56.67 %). In the experiment on our proposed approach, we can see that we got a high rate for all case 1-3. That means our approach deals with the illumination effect robustly and effectively and is highly comparative to one of the best approaches (LTP/DT) with the small difference (about from 1 % to 2.37 %).

From the KoFace database, we also conduct more comparison results between some other illumination techniques with our VOC model. In this comparision, we choose DCT-based normalization, Wavelet-based normalization, Gradientfaces Normalization, DoG filtering based normalization, and Weberface normalization from the INFace toolbox at http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/INFace/. There are 40 subjects in the dataset. The gallery image are standard and well-illuminated frontal images. The probe images are under indoor and ourdoor lighting conditions. The probe suffers from low and high amount of lighting sources. The result in **Table 4** shows that out VOC model can handle the lighting condition better than the other normalization general techniques and gives a high recognition rate in a real-world database such as our KoFace database.

4.4 Contribution Evaluation of Each Measurement

In the aspect of contribution evaluation of each measurement, we conducted 7 experiments on 4 datasets from the KoFace database. The mutual combination of FM, CM, and IM is considered to evaluate the main contribution of them. The four datasets include two indoor and two outdoor datasets as follows: for each face subject, 10 indoor images are divided ½:½ into Indoor 1 and Indoor 2 datasets. 10 outdoor images are also divided ½:½ into the Outdoor 1 and Outdoor 2 datasets.

Table 3. Comparison Experiments between the proposed algorithm and LTP/DT approach.

Dataset(Yale B)	Illumination Effect Condition	PP + GMM Baseline	PP + GMM/VOC	PP + LTP/DT
Case 1	Very high	56.67 %	93.33%	95.7%
Case 2	Low to High	90.67 %	98%	99%
Case 3	Very low	100%	100%	100%

Table 4. All cases of combination from FM, CM , and IM based on four datasets

Techniques	Indoor condition	Outdoor condition
DCT–based Normalization	83.5 %	76 %
Wavelet- based Normalization	82.5 %	76.5 %
Gradientfaces Normalization	67 %	58.5 %
DoG filtering based Normalization	71 %	60.5%
Weberface Normalization	70 %	69.5 %
VOC Model (Proposed method)	94.67 %	84.67 %

Table 5 shows numerical results of the mutual contributions of each measurement. Look at this table, we have some evaluations as:

First evaluation: the GMM baseline gives rather low recognition rates of about 70-80% and 60-65% for indoor and outdoor, respectively. The illumination has a considerably degraded the performance in these cases.

Second evaluation:, FM, CM, and IM are individually employed. The FM proves that discriminate features are the main key to increase the recognition rate to 85-90% for indoor datasets and 75-80% for outdoor datasets. IM gives an even higher recognition rate in these cases. Only CM gives a poor result when employed individually.

Third evaluation: we conducted three more combinations: FM + CM, FM + IM, and CM + IM. Table 4 shows that FM and CM give high FRRs of over 90% for indoor datasets and over 80% for outdoor datasets. However, the rate with the combination of CM and IM is not high, at only 85-90% for indoor and 70-75% for outdoor datasets. This means that FM plays an important role in the combination for the indoor case, and IM plays an important role in the combination for the outdoor case. CM is merely a supplement to help FM obtain a more accurate block. The last experiment is the overall combination of all measurements. With the combination of FM, CM, and IM, we always have the highest recognition rate in indoor and outdoor datasets with rates of about 94-98% and 86-90%, respectively.

Table 5. All cases of combination from FM, CM , and IM based on four datasets.

KoFace Database	Indoor 1	Indoor 2	Outdoor 1	Outdoor 2
Baseline	74.67%	82.67%	61.33%	65.33%
FM	88%	89.67%	77.33%	76%
CM	75%	83.33%	62%	66.33%
IM	81.33%	84%	71.33%	70.67%
FM + CM	94.67%	93.33%	82.67%	80%
FM + IM	96.67%	94%	86.33%	81.33%
CM + IM	88%	86.67%	73.33%	74%
FM + CM + IM	97.33%	94.67%	89.33%	86.67%

The contribution of FM, CM and IM is demonstrated in **Fig. 19** in comparison with the GMM baseline approach. **Fig. 20** more clearly shows the increasing degree of FM, FM + CM,

FM + IM, and FM + CM + IM in each indoor and outdoor dataset. They nearly converge to 100% in the case with little illumination, and to 90% in the case of diverse illumination.

Figs. 21a and **Figs. 21b** illustrate us the ROC curve of our face recognition method with and without the visual observation confidence. The x axis represents the false positive rate and the y axis represents the true positive rate. The experiment for this case was conducted on an indoor dataset with 75 images for the training set (5 images x 15 face subjects).

The ROC curve with the visual observation confidence has an Area Under Curve (AUC) (0.7219) larger than the ROC curve with no visual observation confidence (AUC = 0.5938) in the same case. This means that the GMM classifier using the VOC value is better than the GMM classifier without the VOC value.

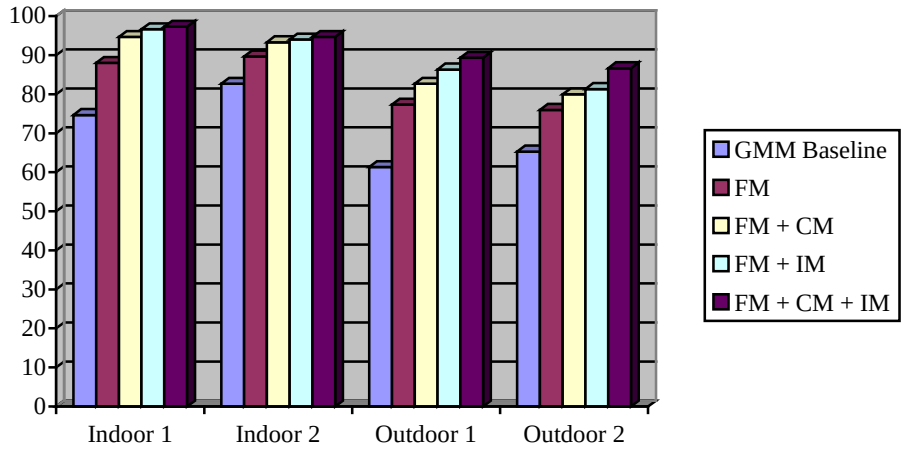

Fig. 19. Contribution of FM and IM in indoor and outdoor datasets

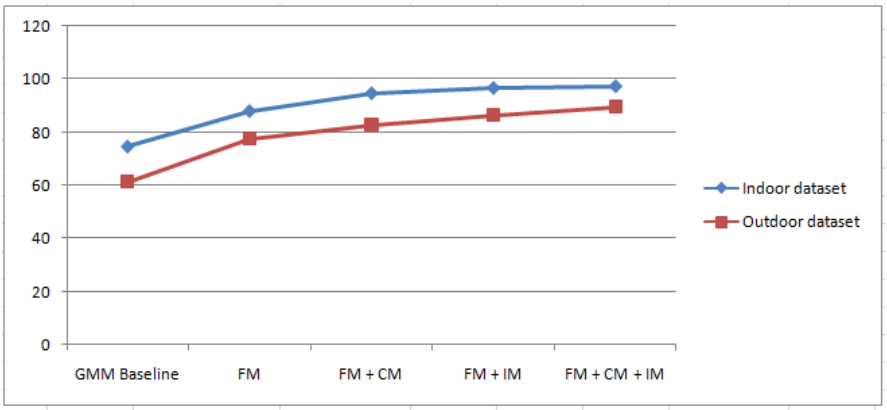

Fig. 20. Convergence of the proposed approach in the case of indoor and outdoor datasets.

In **Fig. 21c**, the relationship between the true positive and the false positive rate and the number of samples of the indoor dataset is shown. The x axis is the number of testing samples from 0 to 75 testing images and the y axis is the true positive and false positive rates. The pink and red points represent the true positive and false positive rates obtained using VOC. Similarly, the blue and green points represent the true positive and false positive rates with no VOC value. The pink points are always higher than the blue points and the red points are

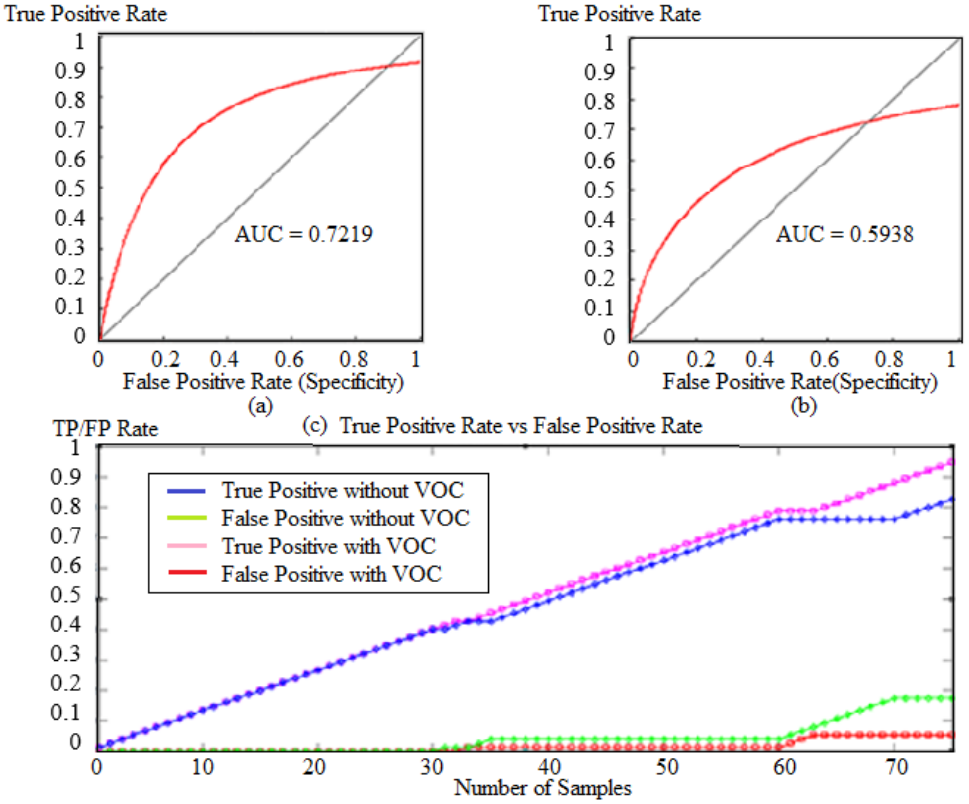

Fig. 21. (a) ROC of indoor dataset with VOC participation. (b) ROC of indoor dataset without VOC participation. (c) Relationship between true positive/false positive rate and the number of samples.

always lower than the green points. This means that the VOC makes a significant contribution to the GMM face recognition process in increasing the true positive rate and reducing the effect of the false positive rate. In summary, using the visual observation confidence is a reliable approach to help GMM-based recognition systems to obtain higher accuracy.

5. Conclusion

We proposed a GMM-based face recognition approach using the visual observation confidence to deal with the problem of illumination impacts effectively and completely. We defined the Flatness Measure (FM), Centrality Measure (CM), and Illumination Normality Measure (IM). These measurements reflect three characteristics of one feature vector: the discrimination, distance, and illumination. And VOC is the linearly optimal combination of these measurements. In GMM-base face recognition approach, we include some modifications in EM algorithms and the classification process. The experimental results showed that the proposed approach can work well with many types of databases.

References

[1] S.I. Choi, G.M. Jeong, “Shadow compensation using Fourier analysis with application to face recognition,” *IEEE Signal Processing Letters*, vol.18, no.1, pp.23-26, 2011.
Article (CrossRef Link)

[2] D. Demers, G.W. Cottrell, "Non-linear dimensionality reduction," Advances in neutral information processing systems, vol.5, pp.580-587, 1993. Article (CrossRef Link)
[3] H. K, Ekenel, R. Stiefelhagen, "Block selection in the local appearance-based face recognition scheme," in *Proc. of the 2006 IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pp.43-50, 2006. Article (CrossRef Link)
[4] C. Garcia, G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Transactions on Multimedia*, vol.1, no.3, pp.264-277, 1999. Article (CrossRef Link)
[5] H.P. Graf, T Chen, E. Petajan, and E Cosatto, "Locating faces and facial parts," in *Proc. of International Workshop on Automatic Face and Gesture Recognition*," pp.41-46, 1995. Article (CrossRef Link)
[6] H. Han, S. Shan, L. Qing, X. Chen, and W. Gao, "Lighting aware preprocessing for face recognition across varying illumination," *European Conference on Computer Vision*, pp.308-321, 2010. Article (CrossRef Link)
[7] R. L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," *IEEE Transaction on PAMI*, vol.24, no.5, pp.696-707, 2002. Article (CrossRef Link)
[8] Jiang, H., 2005. Confidence measures for speech recognition: A survey. *Speech Communication*, p.455 – 470. [http://dx.doi.org/ doi:10.1016/j.specom.2004.12.004]
[9] J.Y. Kim, D.Y. Ko, and S.Y. Na, "Implementation and enhancement of GMM face recognition systems using flatness measure," in *Proc. of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, p.247-251, 2004. Article (CrossRef Link)
[10] J.Y. Kim, S.H. Min, S.Y. Na, and S.H. Choi, "Modified GMM training for inexact observation and its application to speaker identification," *Speech Sciences*, vol.14, no.1, pp.163-175, 2007. Article (CrossRef Link)
[11] P.H. Lee, S.W Wu, and Y.P. Hung, "Illumination compensation using oriented local histogram equalization and its application to face recognition," *Journal of Image Processing*, vol.21, no.9, pp.4280-4289, 2012. Article (CrossRef Link)
[12] H.Q. Li, S.Y. Wang, and F.H. Qi, "Automatic face recognition by support vector machines," in *Proc. of Combinatorial Image Analysis, Lecture Notes in Computer Science*, vol.3322**,** pp.716 – 725, 2004. Article (CrossRef Link)
[13] N. Nallammal, V. Radha, " Performance evaluation of face recognition based on PCA, LDA, ICA and Hidden Markov Model," in *Proc. of ICDEM*, pp.96-100, 2012. Article (CrossRef Link)
[14] M. Nixon, "Eye spacing measurement for facial recognition." *SPIE Proceeding* 0575, p.279-285, 1985. Article (CrossRef Link)
[15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.9, no.1, pp.62-66, 1979. Article (CrossRef Link)
[16] D. Reisfeld, Generalized Symmetry Transforms: Attentional Mechanisms and Face recognition, PhD. Thesis, Tel-Aviv University, 1994.
[17] D.A. Reynolds, "Speaker identification and verification using gaussian mixture speaker models," *Speech Communication*, vol.17, no.1-2, pp.91-108, 1995. Article (CrossRef Link)
[18] C. Sanderson, K.K. Paliwal, "Fast features for face authentication under illumination direction changes," *Pattern Recognition Letters*, vol.24, no.14, pp.2409-2419, 2003. Article(CrossRef Link)
[19] C. Sanderson, M., Saban, and Y. Gao, "On local features for GMM based Verification," in *Proc. of the Third International Conference on Information Technology and Applications*, vol.1, pp.650-655, 2005. Article(CrossRef Link)
[20] X. Tan, B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol.19, no.6, pp.1635-1650, 2010. Article(CrossRef Link)
[21] L. Xu, M.I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixtures," *Neural Computation*, vol.8, no.1, pp.129- 151, 1996. Article(CrossRef Link)
[22] M. Kafai, B. Bhanu, "Reference face graph for face recognition," *IEEE Trans. on Information Forensics and Security*, vol.9, no.12, pp.2132 – 2143, 2014. Article(CrossRef Link)
[23] Y. Sun, X. Wang, X. Tang, "Deep learning face representation from predicting 10,000 classes," in

Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.1891-1898, 2014. Article (CrossRef Link)

[24] A. Punnappurath, A.N. Rajagopalan, "Face fecognition across non-uniform motion blur, Illumination, and Pose," *IEEE Transactions on Image Processing*, vol.24, no.7, pp.2067-2082, 2015. Article (CrossRef Link)

Tran Anh Tuan is a lecturer at Computer Science Department of University of Sciences, Ho Chi Minh city, Vietnam. He received the PhD. degree in Electronics and Engineering Department at Chonnam University, Korea in 2014. His current research interests include medical image processing and analysis, pattern recognition, and object classification.

Jin Young Kim received the Ph.D degree in electronic engineering from the Seoul National University. He worked on speech synthesis at Korea Telecom from 1993 to 1994. Since 1995 he has been a professor in the Dept. of Electronics and Computer Eng., Chonnam National University. His research interests are speech synthesis, speech and speaker recognition, and audio-visual speech processing. (corresponding author)

Asmatullah Chaudhry 1993 : M.Sc. Physics ; 1998 : M.Sc. Nuclear Engg ; 2003 : MS Computer System Engg. 2007 : Ph. D. Computer System Engg. ; 1998 ~ now : Pr. Scientist, PINSTECH, Islamabad, Pakistan ; 2011 ~ now : Postdoc Fellow, School of Electronics and Computer Engineering, Chonnam National University ; Research Interests : Image Processing, Cognitive Radio, Pattern Recognition, and Machine Learning.

Pham The Bao received the Ph.D degree in computer science from the University of Science, Ho Chi Minh city, Vietnam. He worked on Mathematics & Computer Science Faculty, University of Science from 1995. Since 2013 he has been a professor. His research interests are image processing, pattern recognition, and computing.

Hyoung-Gook Kim received the Dr-Ing degree in computer science from the Technical University of Berlin, Germany. From 1998 to 2005, he worked at Daimler Benz and Siemens, Berlin, Germany. From 2005 to 2007, he was a PL at the Samsung AIT, South Korea. Since 2007, he has been a professor in the Dept. of Electronics Convergence Eng., Kwangwoon University, South Korea. His research interests include audio signal processing, audio-visual content indexing and retrieval.