# Face Recognition Performance Improvement using a Similarity Score of Feature Vectors based on Probabilistic Histograms

Gatesakda SRIKOTE, Anupap MEESOMBOON
*Department of Electrical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, 40002, Thailand*
*anupap@kku.ac.th*

***Abstract*—This paper proposes an improved performance algorithm of face recognition to identify two face mismatch pairs in cases of incorrect decisions. The primary feature of this method is to deploy the similarity score with respect to Gaussian components between two previously unseen faces. Unlike the conventional classical vector distance measurement, our algorithms also consider the plot of summation of the similarity index versus face feature vector distance. A mixture of Gaussian models of labeled faces is also widely applicable to different biometric system parameters. By comparative evaluations, it has been shown that the efficiency of the proposed algorithm is superior to that of the conventional algorithm by an average accuracy of up to 1.15% and 16.87% when compared with 3x3 Multi-Region Histogram (MRH) direct-bag-of-features and Principal Component Analysis (PCA)-based face recognition systems, respectively. The experimental results show that similarity score consideration is more discriminative for face recognition compared to feature distance. Experimental results of Labeled Face in the Wild (LFW) data set demonstrate that our algorithms are suitable for real applications probe-to-gallery identification of face recognition systems. Moreover, this proposed method can also be applied to other recognition systems and therefore additionally improves recognition scores.**

***Index Terms*—gaussian mixture model, expectation-maximization algorithm, similarity score, probabilistic histogram, face recognition**.

## I. Introduction

Human face recognition has become one of the most popular biological identification techniques among signature, fingerprint analysis, retinal or iris scan, palm veins, hand geometry or gait because of its friendly interface and easy comprehension [1]. Different types of biometric information are suitable for different kinds of various practical identification applications due to their variations in sensor availability, device availability, intrusiveness, computational cost, accuracy, reliability, and power consumption. Some of the face recognition applications are information security, forensic applications, smart cards, law enforcement, entertainment, and surveillance. Research on still image face recognition began nearly half a century ago. There are two main approaches that have been proposed for illumination of the invariant problem, one is based on appearance which uses holistic templates [2-3] and the other is model-based feature geometry which employs shape and texture of geometrical local features [4-6]. A comparison of the two approaches [7] indicates that the first recognition method is more expensive with regard to computational cost with a higher recognition rate, whereas the second one is fast and has a low computational cost with a lower recognition rate. A general statement of the problem of face recognition scenarios can be classified into two categories, namely face identification and face verification. First, in identification scenarios the system performs a comparison (a one-to-many matching process), the input to the system is an unknown face (also called test face/probe or face/query face), and the system completes identification by returning who has the highest similarity with the test face from the system database of known individuals. Within the verification mode (realistic authentication/one-to-one comparison), the system needs to compare a probe face image against the specific templates stored in a biometric system database in order to verify that the individual is the person they claim to be. One problem in the real-world face verification field that has not yet been solved is a matter of perspective of the two different unseen faces of the same person, but are determined to be a different person due to the environment being uncontrolled. As a result, when poses vary quite dramatically, face detection and alignment will have great influence on the recognition performance [8]. Thus to address these problems, not only does the face recognition system need to construct generic face models using various feature extraction methods, but it also can determine whether the faces are not well aligned frontally. The main contribution of this work is that we propose to address the varying pose problems by extending each pair of statistical frontal face feature comparisons with the similarity score between face feature consideration methods. This paper describes a system for one-to-one face recognition which uses an adapted form of the probabilistic MRH [9]. In this paper, we propose a simple yet effective clustering technique that produces classification maps by the following: 1) considering the spatial-context information and 2) preserving the small areas and details present in the scene. This technique merges the advantages of segmentation, hierarchical methods, similarity score and expectation-maximization (EM) algorithm with a Bayesian framework. Moreover, our proposed method can also be deployed to other recognition systems in order to additionally increase recognition accuracy.

The paper proceeds as follows: Section 2 summarizes

This work was financially supported by Khon Kaen University under the Incubation Researcher Project.

Digital Object Identifier 10.4316/AECE.2016.03015

face verification systems: face localization, feature extraction, Gaussian mixture models, and model parameter estimation. The proposed method is described in Section 3. In Section 4, we describe the empirical evaluation results. Finally, the main findings and conclusions are discussed in Section 5.

## II. Face Verification System

A generic probabilistic based face verification system has the components of face localization, feature extraction, estimation of the model parameters, and face signature comparison [1]. A typical face recognition system is illustrated in Fig. 1 and details of the system components are given below.

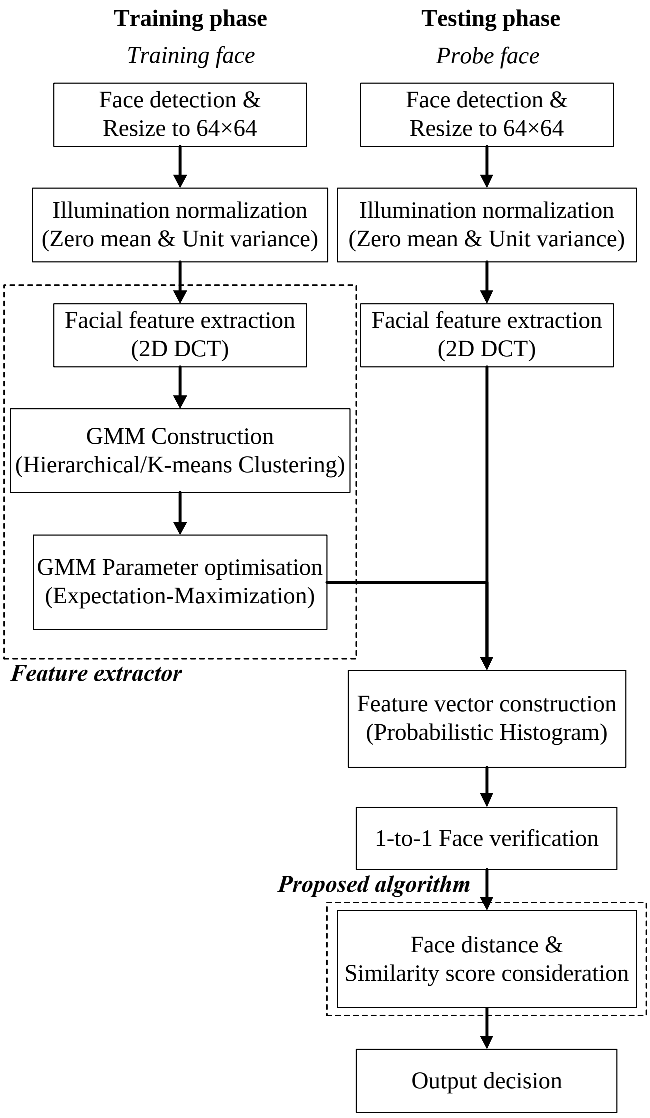


Figure 1. The block diagram of the proposed algorithm

### A. Face Localization

In this section we discuss basic details to implement the face localization technique. The face detection and localization step presented here use the Haar feature-based cascade classifier [10-11]. From Fig. 2, each classifier uses rectangular areas known as Haar features to make a decision about whether the region of the image looks like the predefined Haar features or not. Eyes are located within each face using the frontal face type Haar classifier cascade and only the biggest face is then considered. Finally, the faces are cropped and resized to 64x64 pixels.

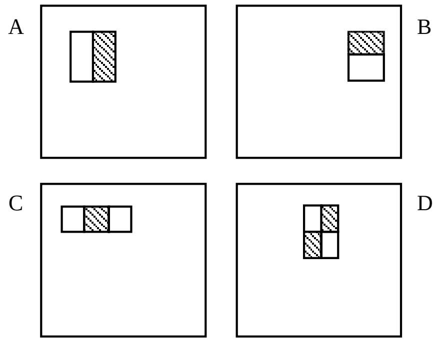


Figure 2. Pattern of Haar features. The edge features are shown in A and B. C is a line feature. D is a special diagonal line feature.

### B. Feature Extraction

One of the most frequently used transformations for the image compression and feature extraction is the two-dimensional discrete cosine transform (2D-DCT) [12].

Reducing the size of data, the magnitudes of the frequency coefficients are extracted as important features of the package. Descriptive features are extracted from each block (or patch) via 2D-DCT decomposition. The face area which is cropped and resized to 64x64 is described as a set of feature vectors. The high dimensional feature vectors are obtained by dividing the face into small overlapping blocks. Each block consists of $8 \times 8$ (64 elements) subsquares with 7 pixels overlapping the neighboring blocks.

Let $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of feature vectors representing a face where $\mathbf{x}_n$ is associated with the 2D-DCT coefficients of each block. These coefficients are ordered according to a zigzag pattern which implies that the amount of information is stored at a given frequency. In this case, the preliminary experiments suggest that using the first 3 low frequency elements of 64 DCT coefficients is sufficient [9]. A face feature vector $\boldsymbol{h}_A$ is then constructed to obtain a high dimensional probabilistic histogram according to Gaussian components of a visual dictionary model as described in the following subsection.

### C. Gaussian mixture models

For each feature vector of the human face, the probability density function of the Gaussian mixture distribution of *k*-components (model probability at $\mathbf{x}_n$) can be written as a linear combination of each Gaussian model as the following:

$$P(\mathbf{x}_n) = \sum_k P_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

where $P_k$ denotes the *k*-th population fraction (or prior class probabilities), $\boldsymbol{\mu}_k$ is the mean vector, $\boldsymbol{\Sigma}_k$ is the covariance matrix which is assumed to be positive definite and $N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the probability density function of the *k*-th component feature vector, which can be expressed as the M-dimensional Gaussian distribution [13]

$$N(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{M/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu}) \right\} \tag{2}$$

where $\mathbf{x}$ is an M-dimensional random vector of DCT coefficients.

The feature extraction approach in this section is motivated by the "visual words" method which is used in image categorization [14] consisting of two important processes. The first process is to build a generic model (a Gaussian mixture model) from a training set of faces and not specific to any person. The mean of each Gaussian distribution in generic models represents a small patch in the face words. The next step is that a given face with $64\times 64$ pixels is divided in to 784 blocks of $8\times 8$ pixels. To compensate for varying illumination and contrast, each block is normalized to have zero mean and unit variance. Then a feature vector for a given face *A* is written as

$$\boldsymbol{h}_A = \left[\sum_{n=1}^{N} P_1 N_1(\mathbf{x}_n \mid \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, \sum_{n=1}^{N} P_k N_k(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right]^T \tag{3}$$

where $P_k$ in $\boldsymbol{h}_A$ is the weight or the *k*-th population fraction for Gaussian model, and $k$-th element is the posterior probability of vector $\mathbf{x}_n$ (descriptive features are extracted from each block using DCT) according to the Gaussian model of a visual dictionary [15-16]. The dictionary model is a Gaussian Mixture Model (GMM) with $k$ components, in which each parameter is optimized by using the EM algorithm [13].

*D. The model parameters estimation using EM algorithm*

The most powerful and widely used method for finding the maximum likelihood of GMM is the EM algorithm. Roughly speaking, the goal of EM is to maximize the likelihood function of the model in respect to each Gaussian parameter. The EM algorithm iterates between two steps which are an expectation (E) step and a maximization (M) step. The process will run until its estimation converges to a local maximum of the likelihood. The EM algorithm is an iterative procedure that consists of two alternation steps. This algorithm obtains the optimized Gaussian parameters by performing the following iterations.

In the Expectation step (E-step), the likelihood function or responsibility of the sample observations is expressed by

$$p_{nk} \equiv P(k \mid n) = \frac{N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) P_k}{P(\mathbf{x}_n)} \tag{4}$$

where $P(k \mid n)$ is denoted as the conditional probability.

Then, in the Maximization step (M-step,) the parameters are restricted using the current expectation value to maximize the log-likelihood function with respect to the set of Gaussian parameters, so the updated equations of the parameters are:

$$\boldsymbol{\mu}_k^{new} = \sum_n p_{nk}\mathbf{x}_n \Big/ \sum_n p_{nk} \tag{5}$$

$$\boldsymbol{\Sigma}_k^{new} = \sum_n p_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \Big/ \sum_n p_{nk} \tag{6}$$

$$P_k^{new} = \frac{1}{N}\sum_n p_{nk} \tag{7}$$

Then we evaluate the log-likelihood

$$\sum_n \ln\left(\sum_k P_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right) \tag{8}$$

and check for convergence of either the parameters or the log-likelihood. If we are not satisfied with the convergence criterion, we return to E-step.

To avoid underflow in practice, the model probability at $\mathbf{x}_n$ ($P(\mathbf{x}_n)$) should proceed as the log-sum-exponential method, which can be written as

$$\ln\left(\sum_k \exp(z_k)\right) = z_{\max} + \ln\left(\sum_k \exp(z_k - z_{\max})\right) \tag{9}$$

where $z_k = \ln\{P_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}$ and $z_{\max} = \max_k\{\ln\{P_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}\}$.

For developing the face recognition model the k-means algorithm must be employed in order to find appropriate parameters, as the EM algorithm requires more iterations and increases computational costs if compared with the k-means algorithm.

## III. Face Signature Comparison using Similarity Score

A decision on whether or not two given faces *A* and *B* are the same person, i.e. the matched pair, or two different people, i.e. the mismatched pair, can be obtained by comparing a similarity measurement between two faces with a threshold. For the training process, the threshold value can be determined from the optimal selection to get a minimum error rate of matching and mismatching pairs. The similarity measurement which is typically used to compare two faces is a metric or distance between two feature vectors. In most research, $\ell_1$- norm based distance [15] is used to measure how far two faces are given by

$$d(\boldsymbol{h}_A, \boldsymbol{h}_B) = \left\|\boldsymbol{h}_A - \boldsymbol{h}_B\right\|_1 = \sum_{i=1}^{k}\left|\boldsymbol{h}_A^{[i]} - \boldsymbol{h}_B^{[i]}\right| \tag{10}$$

where the superscript [i] means the *i*-th element of the feature vector. Normally, the final value of the accumulated absolute difference is used as the distance measurement between two face signatures. Note that the distance (10) will be called the face feature distance in the below sections.

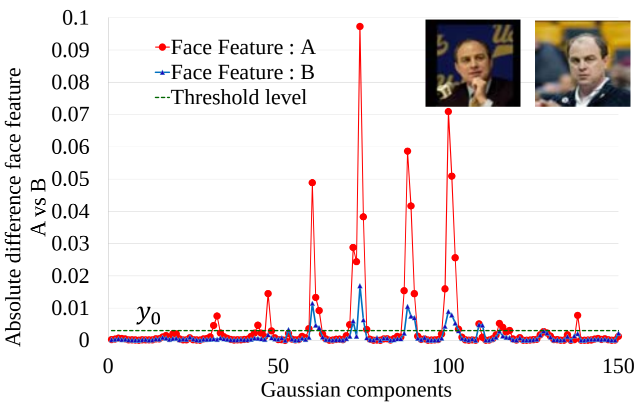


Figure 3. Difference in two face features from the same person

Performing on a test set database, the test protocol is a verification base which classifies whether a pair of previously unseen faces are the same or different. The decision is made by comparing the face feature distance measurement with the threshold value. If the value is greater than the threshold, then it is considered to be a different person. On the other hand, if it is less, then it is considered to be the same person. The disadvantage of this method is the inability to indicate a difference in each of the Gaussian components. As a result, the wrong decision can occur in the case of testing the same person's face because the face feature distance is greater than the threshold, although there are similarities in each Gaussian component, as illustrated in Fig. 3. As a result of these problems, we propose to analyze the face feature distance in conjunction with patterns of similarities. Two-dimensional spatial data are created in order to be able to increase discrimination ability. Thus, a two-dimensional representation of the data will actually be a representation in the full space with no loss of information. A coordinate in another dimension which is created from a value of the similarity score is obtained from the summation of the similarity coefficient as described in the next subsection.

*A. Similarity score*

The simplest similarity measures deal only with present-absent data. The basic concept for calculating similarity coefficients is based on a $2\times 2$ table (see Table I).

Let $a$ be the number of Gaussian components present in both face features A and B, $b$ be the number of Gaussian components present in face feature A but not in B, $c$ be the number of Gaussian components present in face feature B but not in A and $d$ be the number of Gaussian components absent in both face features A and B. The simple matching coefficient is defined by [17]

$$s_{rt} = (a+b)/(a+b+c+d) \tag{11}$$

TABLE I. THE SIMILARITY COEFFICIENT DEFINITION

| | | **Face feature A : No. of Gaussian components** | |
|---|---|---|---|
| | | present | absent |
| **Face feature B : No. of Gaussian components** | present | $a$ | $b$ |
| | absent | $c$ | $d$ |

For the similarities between the feature vectors, this research is particularly interested in cases that occur concomitantly only. This refers to the number $a$ in Table I. In practice, we can neglect the denominator in (11) due to the fact that all of the simple matching coefficients are divided by the same constant, it does not result in the classification performance. Therefore, the operator similarity index (coefficient) $\delta_{li}$ is defined by

$$\delta_{li} = \begin{cases} 1 \;\; if \;\; h_A^{[i]} \; and \; h_B^{[i]} > y_0 \\ 0 \;\; ohterwise \end{cases} \tag{12}$$

where $y_0$ is the similarity detection threshold as depicted as by a green dotted line in Fig. 3, the test number $l$ is an index of a predefined face pair and $i$ is the $i$-th Gaussian component. This means that the similarity score is detected if the probability in $i$-th Gaussian component of faces $h_A^{[i]}$ and $h_B^{[i]}$ is both greater than $y_0$. Then, the similarity score (matching coefficient) $\sigma_l$ is defined by

$$\sigma_l = \sum_{i=1}^{k} \delta_{li} \tag{13}$$

This implies that the similarity score must be within the range $[0,k]$. Fig. 4(a) shows the two-dimensional plot between the similarity score versus the Euclidian distance of face features from two classes of 500 unseen matched and 500 unseen mismatched pairs in the test set denoted by blue circles and red crosses, respectively. As can be seen, the similarity score is helpful in discriminating the X axis (face feature distance) from the Y axis (similarity score), with the consistent advantage of using only face feature distance in only the Y axis as depicted in Fig. 4(b). However, the class separation accuracy also depends on a suitable threshold which will be explained how to obtain in the next section.

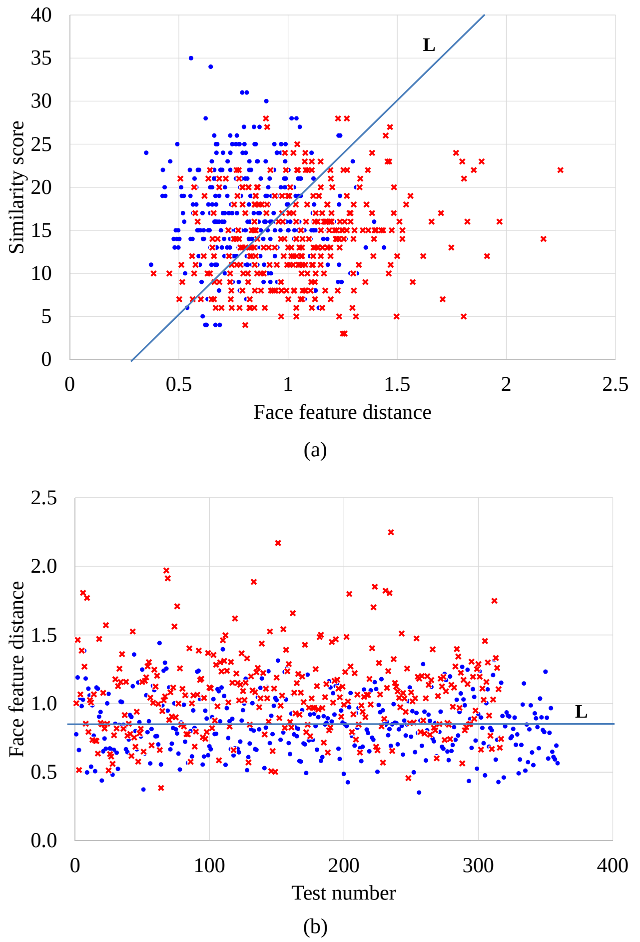


Figure 4. Illustration of the test result for a simple linearly separable data set. We see that the decision boundary lies roughly midway between the clusters of data points. Plot (a) shows the similarity score versus face feature distance in a two–dimensional feature space. Plot (b) shows only face feature distance corresponding to test number.

## IV. EXPERIMENTAL RESULTS

In this section, we evaluate how well the proposed similarity measurement method can identify whether or not a pair of previously unseen faces is the same person in cases of incorrect decision. The training phase is typically accomplished by using supervised approaches that take advantage of the information included in the training set (labeled faces image). In our experiments we also propose to use the Yale face database to construct a visual dictionary due to the fact that it consists of images that have a difference of lighting and facial expressions under a controlled environment. The test protocol was set in a verification base, in which we used a recent and difficult LFW dataset [18] under unconstrained environments collected from the web by the Computer Science Department at the University of Massachusetts in 1994. This dataset is authentic and more suitable for the properties of the proposed approach, which contains 13,233 images of faces with variations in resolution, localization, illumination, expression and pose. There were 1100 matched and 1100 mismatched pairs in each training set and 500 unseen matched and 500 unseen mismatched pairs in the test set. All the proposed algorithms were implemented in C++ platform with OpenCV library and executed on Intel(R) Core(TM) i7-4790 CPU 3.60GHz with 4GB of RAM with only the operating system running.

In the first experiment we studied the effect of increasing the threshold value on the classification results. In statistics, the receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. In general, accuracy is measured by the area under the ROC curve. The best possible prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). The (0,1) point is also called the perfect classification. The accuracy of the test depends on how well the test separates the group into match and mismatch clusters. The threshold was optimized to obtain the highest average accuracy which was averaged over the classification accuracies for 500 matched and 500 mismatched pairs as shown in Fig. 5.

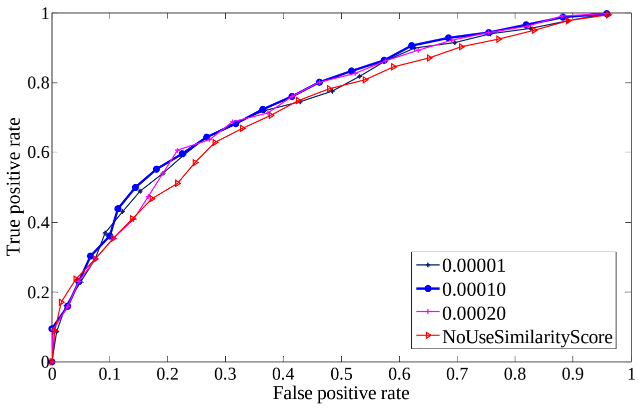


Figure 5. ROC curve

For the maximization of the area under the curve (AUC) with respect to a threshold $y_0$, the threshold was varied from 0.00001 to 0.0002 with a step size 0.00001 as shown in Fig. 6. It is clear that for threshold 0.0001, we had achieved a best classification accuracy of up to 0.7546 of AUC, while on the other hand the method did not consider the similarity score, the AUC was only 0.7223.

In the second experiment we fixed the value of the threshold at 0.0001 and varied the decision boundary to find the most accurate rate. We used the brute-force algorithm to find the best separation hyperplane (the line L in Fig. 4(a)). The purpose of the classification algorithm is to find the line separating most accurately a set of data represented by numerical vectors into two classes. The training data of this experiment were formed by a set of labeled 2D-points that belongs to one of two different classes.

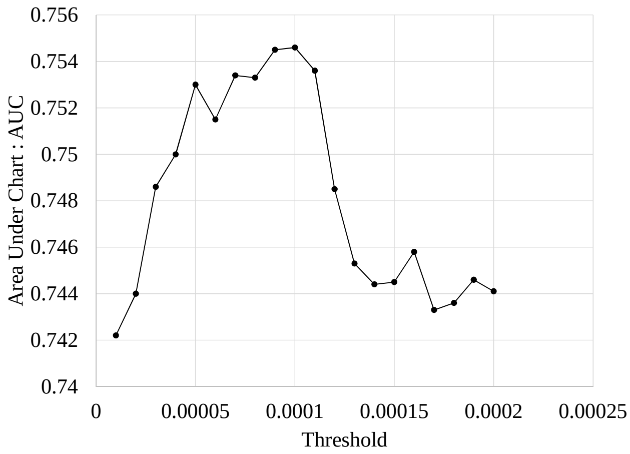


Figure 6. AUC curve of varied threshold value

The output of a classification is a set of two values, i.e. binary classification. The decision line that separates a two dimensional-two class plot is the line that passes through the points ($x_1$, $y_1$) and ($x_2$, $y_2$) as illustrated in Fig. 7. We start ($x_1$, $y_1$) at (0,0) and ($x_2$, $y_2$) at (0,40). In each iteration step, $x_2$ was varied from 0 to 2.5 with a step size 0.1 by holding $y_1$ and $y_2$ constant at 0 and 40, respectively. Then we increase $x_1$ with a step size 0.1 followed by repeating and varying $x_2$ as the previous step ($y_1$ and $y_2$ are still kept fixed at 0 and 40, respectively).

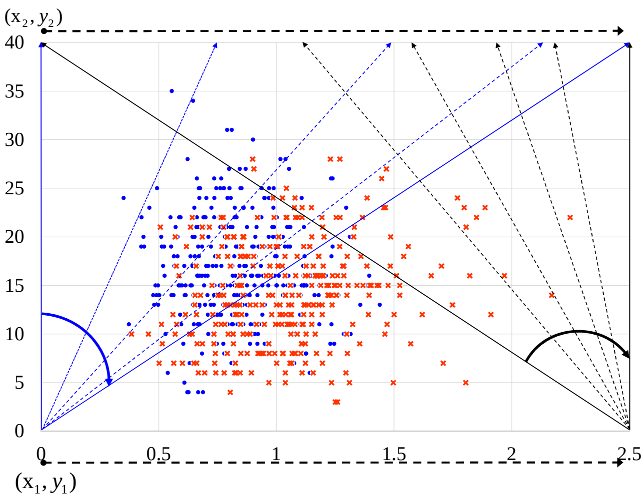


Figure 7. A two-dimensional plot depicting how the decision line changes

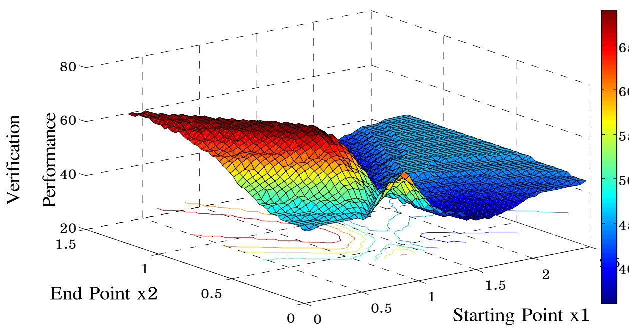


Figure 8. A three-dimensional plot of verification performance versus position of decision boundary

The accuracy was calculated and plotted versus coordinates of the decision boundary line for each iteration step as depicted in Fig. 8. Finally, this process continues until $x_1$ and $x_2$ are both at 2.5. The 3D-surface plot is clear and it is easy to determine the shape and variation in height across the $x_1$ and $x_2$ grid coordinates.

The height for the display is specified by the accuracy corresponding to coordinates of the decision boundary line that passes through each iteration step. Fig. 8 indicates that the starting point of the decision boundary line ($x_1, y_1$) at (0.7, 0) and the end point ($x_2, y_2$) at (1.4, 40) improve accuracy in most cases. Then, the proposed method was evaluated in comparison with PCA with raw distance and normalized distance [19]. The experiments were conducted with the verification scenario. Following the suggestion in [9], 16 eigenfaces were used and the first three eigenfaces were disregarded. Results for the randomized binary trees (RBT) were obtained from the method published in [19] and the results for 3x3 MRH are obtained from [20]. In terms of performance for face verification accuracy, the experiments demonstrate that our proposed similarity based method achieves an average accuracy of up to 1.15% when compared with 3x3 MRH and outperforms those methods using PCA (raw and normalized distance) and the RBT method by 16.87% and 5.42%, respectively. The results suggest that our approach treats the mismatch test problem as an intrinsically two-dimensional (similarity and distance) discrimination problem rather than a one-dimensional (distance) discrimination problem by taking advantage of the discriminative power of two-dimensional characteristic views. As mentioned in [21] many applications of face recognition do not require perfect identification but most require a low false-positive rate. In searching a large database of faces, for example, it may be preferable to obtain a small set of likely matches to present to the user. Our experiments show that the similarity score technique can be made to detect just 1.15% or 11 faces from 1000 which would be overlooked if the similarity score was not used.

## V. Conclusions

In this paper we address the pose mismatch problem and present a similarity score measurement, which is a new approach to face recognition based on a probabilistic histogram of the multivariate Gaussian mixture model. Using the EM algorithm the model parameters are estimated under the maximum-likelihood framework. The face recognition system presented here was evaluated on LFW still face datasets. It has been empirically demonstrated that when applied to mismatching face verification tasks, the similarity between face feature vectors achieves a higher accuracy and greater discrimination and is more widely applicable to biometrics when compared to the conventional distance measurement generally used. This method can be utilized for authentication and surveillance systems with greater accuracy. We have demonstrated that the similarity score in the measurement approach can be effective. Thus the proposed similarity method is a more robust and reliable means of consistently boosting face recognition accuracy.

Future work will consider the use of other face detection methods and the average face to make Gaussian mixture models and work on density estimation using these models will also be conducted.

## References

[1] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, “Face recognition: A literature surve,” ACM Comput Surv , vol. 35, no. 4, pp. 399-458, 2003, doi: 10.1145/954339.924342

[2] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose,” IEEE Trans Pattern Anal Mach Intell, vol. 23, no. 6, pp. 643-660, 2001, doi: 10.1109/34.927464

[3] R. Basri and D. W. Jacobs, “Lambertian reflectance and linear subspaces,” IEEE Trans Pattern Anal Mach Intell , vol. 25, no. 2, pp. 218-233, 2003, doi: 10.1109/TPAMI.2003.1177153

[4] Y. Gao and M. K. Leung, “Face recognition using line edge map,” IEEE Trans Pattern Anal Mach Intell , vol. 24, no. 6, pp. 764-779, 2002, doi: 10.1109/TPAMI.2002.1008383

[5] J. Soldera, C. A. R. Behaine and J. Scharcanski, “Customized orthogonal locality preserving projection with soft-margin maximization for face recognition,” IEEE Trans Instrumentation and Measurement, vol. 64, no. 9, pp. 2417-2426, 2015, doi: 10.1109/TIM.2015.2415012

[6] M. Jian and K. M. Lam, “Simultaneous Hallucination and recognition of low-resolution faces based on singular value decomposition,” IEEE Trans Circuits and Systems for Video Technology, vol. 25, no. 11, pp. 1761-1772, 2015, doi: 10.1109/TCSVT.2015.2400772

[7] R. Brunelli and T. Poggio, “Face recognition: Features versus Template,” IEEE Trans Pattern Anal Mach Intell, vol. 15, no. 10, pp. 1042-1052, 1993, doi: 10.1109/34.254061

[8] G. B. Huang, V. Jain and E. L. Miller, “Unsupervised joint alignment of complex images,” In: Proceedings of International Conference on Computer Vision, pp. 1-8, 2007, doi: 10.1109/ICCV.2007.4408858

[9] Y. Wong, S. Chen, S. Mau, C. Sanderson and B. C. Lovell, “Multi-region probabilistic histograms for robust and scalable identity inference,” In: Proceedings of International Conference on Biometrics, Lecture Notes in Computer Science (LNCS), pp. 199-10, 2009, doi: 10.1007/978-3-642-01793-3_21

[10] P. A. Viola and M. J. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004, doi: 10.1023/B:VISI.0000013087.49260.fb

[11] S. Zafeiriou, C. Zhang and Z. Zhang, “A survey of face detection in the wild: Past, present and future,” Computer Vision and Image Understanding, vol. 138, pp. 1-24, 2015, doi:10.1016/j.cviu.2015.03.015

[12] R. Gonzales and R. Woods. Digital Image Processing. 3rd ed. New Jersey: Prentice Hall, pp. 472-473, 2007.

[13] C. Bishop. Pattern Recognition and Machine Learning. Berlin: Springer, pp. 435-439, 2006.

[14] E. Nowak, F. Jurie and B. Triggs, “Sampling strategies for bag-of-features image classification,” In: European Conference Computer Vision (ECCV), Part IV, Lecture Notes in Computer Science (LNCS), pp. 490-503, 2006, doi: 10.1007/11744085_38

[15] C. Sanderson and B. C. Lovell, “Multi-region probabilistic histograms for robust and scalable identity inference,” In: International Conference on Biometrics, Lecture Notes in Computer Science (LNCS), pp. 199-208, 2009, doi: 10.1007/978-3-642-01739-3_21

[16] F. Bashar, A. Khan, F. Ahmed and H. Kabir, “Face recognition using similarity pattern of image directional edge response,” Advances in Electrical and Computer Engineering, pp. 69-76, 2014, doi: 10.4316/AECE.2014.01011

[17] P. N. Tan, M. Steinbach and V. Kumar. Introduction to Data Mining. Boston: Pearson Addison Wesley, p. 73, 2006.

[18] G. B. Huang, M. Ramesh, T. Berg and E. L. Miller. Labeled Faces in the Wild:A database for studying face recognition in unconstrained environments. Faces in Real-Life Images Workshop in European Conference on Computer Vision (ECCV), 2008.

[19] E. Nowak and F. Jurie, “Learning visual similarity measures for comparing never seen objects,” In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2007, doi: 10.1109/CVPR.2007.382969

[20] S. Chen, S. Mau, M. T. Harandi, C. Sanderson, A. Bigdeli and B. C. Lovell, “Face recognition from still images to video sequences: A local-feature-based framework,” EURASIP Journal on Image and Video Processing, pp. 1-14, 2011, doi: 10.1155/2011/790598

[21] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991, doi: 10.1162/jocn.1991.3.1.71