Effective Face Verification Systems Based on the Histogram of Oriented Gradients and Deep Learning Techniques

Sawitree Khunthi
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Maha Sarakham, Thailand
sawitri0212@gmail.com

Pichada Saichua
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Maha Sarakham, Thailand
pichadakt@gmail.com

Olarik Surinta
Multi-agent Intelligent Simulation Laboratory (MISL), Department of Information Technology, Faculty of Informatics, Mahasarakham University
Maha Sarakham, Thailand
olarik.s@msu.ac.th

***Abstract*—In this paper, we proposed a face verification method. We experiment with a histogram of oriented gradients description combined with the linear support vector machine (HOG+SVM) as for the face detection. Subsequently, we applied a deep learning method called ResNet-50 architecture in face verification. We evaluate the performance of the face verification system on three well-known face datasets (BioID, FERET, and ColorFERET). The experimental results are divided into two parts; face detection and face verification. First, the result shows that the HOG+SVM performs very well on the face detection part and without errors being detected. Second, The ResNet-50 and FaceNet architectures perform best and obtain 100% accuracy on the BioID and FERET dataset. They also, achieved very high accuracy on ColorFERET dataset.**

Keywords— face verification systems, face detection, face verification, ResNet-50, FaceNet

I. Introduction

Face verification is part of the face recognition system that focuses on the one-to-one matching problem [1] to compare whether it is the same person or not the same person. For this reason, face verification is much used in security, surveillance, and immigration, for example, to search for people from closed circuit television (CCTV) or to check if the person is a criminal by comparison of a face captured on camera with faces from a database. Many problems, such as images, low-light images, blurred image, and flare on an image resulting from stray light entering the camera lens, will occur depending on the quality and location of the camera. These effects are of concern for the researchers working on face recognition.

Face verification systems perform two main tasks. The first task is face detection and is essential to any face verification system because the system cannot process if the face is not detected. Many researchers focus on developing algorithms for face detection such as edge detection [2], Haar-cascade classifier [3][4], and histogram of oriented gradients (HOG) [5–7]. These algorithms allow us to find faces even in low-light and blurred images. Moreover, convolutional neural networks (CNNs) that have been proposed [8][9] provide a robust method to detect a face in many conditions such as a small faces, occlusion, or images that do not show the entire face.

The second task of face verification, is the extraction of information from the face (called face encoding) which is sent to the similarity function to calculate and compare the unknown face and detected face. A high similarity value shows that the two faces are the most similar face. Many algorithms have been proposed for the face encoding such as local directional number pattern [10], local binary patterns [11], common encoding feature discriminant [12] and supervised feature encoding [13] are proposed. Nowadays, deep learning approaches are successful in encoding the face, including VGGNet [14], DeepFace [15], FaceNet [16] and ResNet [17].

Contribution: In this paper, we evaluate the performance of face verification systems on three well-known face datasets (BioID, FERET, and ColorFERET). It is quite challenging to verify faces from the ColorFERET because this dataset consists of 3,553 face images of 474 subjects. We divided the experiment into two parts; face detection and face verification. In the face detection part, four different face techniques, including the histogram of oriented gradients combined with the linear support vector machine (HOG+SVM), max-margin object detection with convolutional neural network (MMOD-CNN) [18][19], Haar-Cascade Classifier [20][21] and Faced techniques were evaluated on the BioID dataset. The experiments showed that the HOG+SVM performs very well and without errors of face detection. Moreover, in the face verification part, three robust deep CNN architectures called VGG16, FaceNet, and ResNet-50 architectures were used as the face encoding. The experimental results showed that the ResNet-50 and FaceNet performed best and obtained 100% accuracy on the BioID and FERET dataset. Additionally, both architectures achieved very high accuracy on the ColorFERET dataset.

Paper outline: This paper is organized as follows: In Section II, the face verification systems are described in detail. In Section III, three well-known face image datasets are explained. The experimental results of face detection and verification are presented in Section IV. The last section is the conclusion and suggestions for future work.

II. Face Verification Systems

In the following, we describe the face verification systems used in the experiments; the histogram of oriented gradients and linear support vector machine aimed for face detection. Two face encoding methods; FaceNet and ResNet-50, are computed.

A. Face Detection

For face detection, the Viola-Jones face detector [20][21] is a well-known method that was first proposed for object and then for pedestrian detection. Nowadays, this technique, called Haar-cascade classifier, has become a standard technique for face detection. The Viola-Jones face detector computes feature vector based on the Haar feature. It calculates from the rectangle detector or sub-window. The detector scans through the image. Then, the set of the feature vector is given to the AdaBoost classifier, which is the weak classifier. This approach can process in real-time and get high precision. However, this approach performs not very well on the BioID dataset.

We proposed to use the histogram of oriented gradients and the linear support vector machine, called HOG+SVM, in face detection experiments.

First, the well-known HOG [22] is proposed to compute a feature vector from sub-images that scans over the whole image. With this method, the oriented gradients are computed using a gradient detector. Then the oriented gradients of each sub-image are weight to the orientation bins and used as a feature vector [23]. The gradient detector is calculated as follows:

$$G_x = I(x+1,y) - I(x-1,y) \quad (1)$$

$$G_y = I(x,y+1) - I(x,y-1) \quad (2)$$

where G_x is the horizontal and G_y is the vertical components of the gradients.

The gradient magnitude (M) and the oriented gradients (θ) are computed as:

$$M(x,y) = \sqrt{\left(G_x^2 + G_y^2\right)} \quad (3)$$

$$\theta(x,y) = tan^{-1}\frac{G_y}{G_x}d \quad (4)$$

where $M(x,y)$ is the gradient magnitude and $\theta(x,y)$ is the orientation of the gradients at the location (x,y).

Consequently, orientation bins are selected based on oriented gradients. The gradient magnitudes for each oriented gradient are weight and summed up to each orientation bin. Then, the orientation bins for each sub-image are normalized using the L2 normalization.

Second, the support vector machine (SVM) [24] algorithm with a linear kernel is proposed in this paper due to the two-class classification. With the SVM algorithm, the hyperplane, which is the maximum distance to the training points, is used to separate training data. The training points that are closest to the calculated separating hyperplane are called support vectors. So, the best hyperplane is the distance between the closest data points of both classes and the hyperplane [25]. The optimal hyperplane is calculated as;

$$g(x) = W^TX + b \quad (5)$$

where W is the weight vector and b is the bias. The decision rule is

$$y = \begin{cases} 1 \; if g(x) > 0 \\ 0 \; otherwise \end{cases} \quad (6)$$

B. Face Encoding

In this research, two deep learning architectures for face encoding; ResNet-50 and FaceNet are proposed as the face encoding.

1) ResNet-50

The residual network architecture, which is a very deep network, was invented by He et al. [27], called ResNet architecture. The deep residual network creates simple stack layers, therefore the network can be set up as 18, 34, 50, 101, and 152-layer. This architecture is quite different from the original convolutional neural network (CNN) that each layer feedforward to the next layer. A deep residual learning block is implemented in the ResNet architecture (see Fig. 1). Hence, each layer allows to feed the output to feed into the next layer and directly into the next 2-3 forward blocks. This architecture known as shortcut connections.

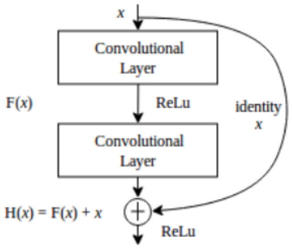


Fig. *1*. The residual network [26].

In this paper, we applied ResNet architecture with 50 layers for the face encoding (called ResNet-50). The lower-level features, which are more specific to the training data, are extracted from the face image. To encode a feature vector; we applied the flatten after the average pooling layer, which is the last layer of the ResNet-50. This architecture encodes 2,048 features and uses them as a feature vector.

1) FaceNet

FaceNet architecture was invented by Schroff et al. [16] to solve the problem of face recognition and clustering. This architecture is invariant to illumination and pose. Firstly, in this technique, the deep CNN architecture, which is inspired by Inception network, is used as a black box. The size of the parameters in FaceNet architecture is 7.5M. The small mini-batch size of around 40 faces per identity (in total, around 1,800 examples) are fed to the deep CNN. These direct to increase convergence while optimizing the network with Stochastic Gradient Descent (SGD).

Secondly, the output from the deep CNN architecture is normalized using L2 normalization and sent to the face embedding process. The embedding process is embeds in a face image into a dimensional space using the Euclidean function. This method guarantees the identity that the face image of person A is closer to other face images of the person A than closer to other face images of other persons.

Finally, the triplet selection is the last process of FaceNet. This process is given the face image of person A to compare other face images from the mini-batch to avoid poor training.

From this process, two parameters are selected, argmax and argmin, which are the hardest positive image of the same person and the hardest negative image of a different person, are selected.

In this paper, we applied FaceNet architecture using Inception network as the core network. This architecture encodes 512 features and used as a feature vector.

III. Face Image Datasets

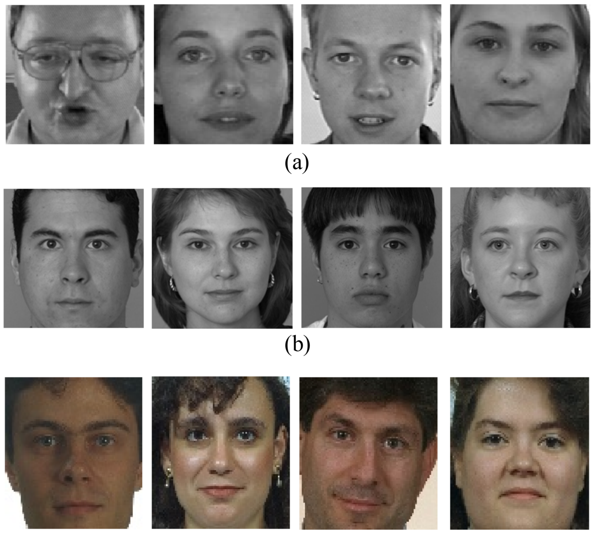
Many face image datasets were invented for face verification systems. In this paper, we select three face image datasets; the BioID, FERET, and ColorFERET dataset for evaluating the face detection and face verification.

A. BioID Face Dataset

The BioID face dataset used in the face detection experiment includes 1,513 frontal view images [27]. In this dataset, the image resolution is 384x286 pixels and stored on the grey level. Additionally, the number of people (subject) used in the face verification experiment is 21 subjects from 1,507 face images. The BioID dataset is shown in Fig. 2(a).

B. FERET and ColorFERET Datasets

The face recognition technology (FERET) dataset and ColorFERET were published in 1993 by J. Phillips and P. Rauss [15-16]. These datasets consist of 1,199 subjects, and the total number of the face images is 14,126 images with an image resolution of 384x256 pixels. In our experiments, we have used the FERET and ColorFERET for face verification. As for the FERET dataset. We selected 1,372 images from 196 subjects from the FERET dataset (See Fig. 2(b)). and 3,553 images from 474 subjects from the ColorFERET dataset (Fig. 2(c)).


(a)

(b)

(c)

Fig. 2. Sample of face images in the (a) BioID, (b) FERET, and (c) ColorFERET datasets.

IV. Experimental Results

A. Evaluation Methods

We have used two methods to evaluate the face verification system. The first evaluation method is face detection accuracy which is given by:

$$Accuracy = Acc - Err \tag{1}$$

where

$$Acc = \frac{c*100}{N} \tag{2}$$

$$Err = \frac{e*100}{N} \tag{3}$$

where c is the number of the face images after applying face detection method, and e is the number of the error face images N is the total number of the face images of the face dataset.

The second method is the accuracy of face verification.

1) We used the cosine similarity function to compare a feature vector extracted from the face image. The most similarity face is given the highest value. Then the correct prediction is that if the label of the highest value is the same as the test image. The cosine similarity function is computed as follows:

$$cos(\theta) = \frac{A.B}{\|A\|\|B\|} \tag{4}$$

where $A.B$ is the dot product of feature vector A and B.

2) To calculate the accuracy, the total number of correct predictions is multiplied by 100 and then divided by the total number of faces in the dataset.

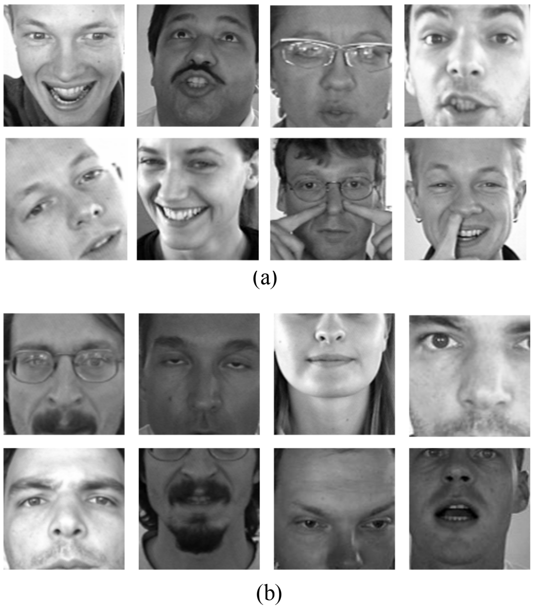
B. Results

In this section, we show the experimental results of face detection techniques and face verification accuracies of CNN face encoding architectures.

1) Face Detection Results

To illustrate the results of face detection, Fig. 3(a) shows face images cropped so as to leave the entire face visible and Fig. 3(b) shows error due to poor cropping that results in the face being only partly visible. In this paper, when calculating the accuracy of the face detection method, we carefully reject the error face images by calculating the error (Err), as shown in Equation 3.

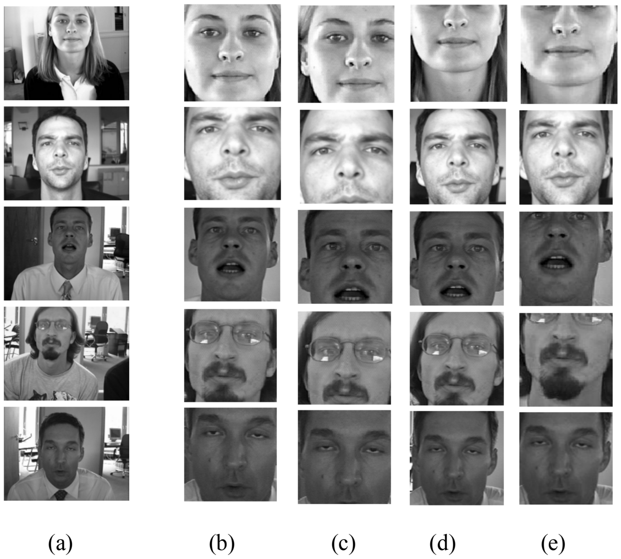
Table I show the experimental results of four different face detection techniques; HOG+SVM, MMOD-CNN, Haar-Cascade, and Faced techniques. Here, the histogram of oriented gradient combined with the linear support vector machine (HOG+SVM) is the only one face detection method that detects face without any error. The performance of HOG+SVM technique obtained on the BioID face dataset is 99.60%. The accuracy obtained from all face detection techniques was over 90%, except for the Faced technique. The face detection results are shown in Fig. 4.



(a)

(b)

Fig. 3. Sample results of the face images after applying face detection method. (a) entire faces and (b) error faces.



(a) (b) (c) (d) (e)

Fig. 4. Face detection results after applying face detection techniques. (a) BioID images, (b) HOG+SVM, (c) MMOD-CNN, (d) Haar-Cascade, and (e) Faced techniques.

TABLE I. PERFORMANCE OF FACE DETECTION TECHNIQUES ON BIOID DATASET

Methods	Number of face detected	Number of error detected	Accuracy (%)
HOG+SVM	1,507	0	**99.60**
MMOD-CNN	1,513	40	97.36
Haar-Cascade	1,459	40	93.79
Faced	1,449	107	88.70

2) Face Verification Results

For the face encoding techniques, we evaluated the performance of three deep convolutional neural networks, including VGG16, FaceNet, and ResNet-50. The image resolution used in the experiments was 224x224 pixels. In the experiments, the VGG16 extracts the highest feature dimension with 25,088 features, followed by ResNet-50 and FaceNet architectures. The image resolution and size of the feature vector are shown in Table II.

In this paper we found that HOG+SVM was the best face detection method based on our experiments on the BioID dataset. We then chose the HOG+SVM method for detecting faces from three face datasets; BioID, FERET, and ColorFERET. As a result, the number of face images detects from the BioID, FERET, and ColorFERET were 1,507, 1,372, 3,553 face images, respectively. This was quite challenging because of the number of subjects in the ColorFERET (474 subjects) was 20 times higher than in the BioID dataset (only 21 subjects). The number of face images and the number of subjects are shown in Table III.

TABLE II. THE RESOLUTION OF FACE IMAGES REQUIRES FOR CNN METHODS AND THE NUMBER OF FEATURES EXTRACTS FROM THREE CNN FACE ENCODING TECHNIQUES

Parameters	Method		
	VGG16	FaceNet	ResNet-50
Image resolution	224x224	224x224	224x224
Feature vector	25,088	512	2,048

TABLE III. FACE VERIFICATION ACCURACIES (%) AND STANDARD DEVIATIONS OF THREE CNN FEATURE EXTRACTION METHODS. THE EXPERIMENTAL RESULTS ARE COMPUTED USING THREE FACE DATASETS

Dataset	Number of image	Number of subjects	Accuracy (%)		
			Vgg16	FaceNet	ResNet-50
BioID	1,507	21	99.74±0.38	100	100
FERET	1,372	196	83.93±0.77	100	100
Color FERET	3,553	474	74.96±1.26	99.32±0.32	99.60±0.46

In this paper, five random fold cross-validations are applied to evaluate the performance of the different face encoding methods. In our experiments, the best deep convolutional neural network (CNN) architecture for face encoding was ResNet-50 and FaceNet architectures because these two architectures obtain an accuracy of 100% on BioID and FERET face datasets. We particularly note that ResNet-50 outperforms other deep CNN architectures when experimenting on the ColorFERET dataset which consists of 3,553 face images with 474 subjects. The ResNet-50 and Facenet architectures had highly accuracies of 99.60% and 99.32%, respectively.

II. CONCLUSION

The key factor in achieving the highest accuracy in face verification systems consists of face detection and the face encoding process. In this paper, we have presented an effective face verification systems. First, the histogram of oriented gradients method combined with the linear support vector machine (HOG+SVM) was applied as the face detection process. The experimental results showed that the HOG+SVM method outperformed other face detection methods; CNN, Haar-Cascade, and Faced methods. There is no error while detecting faces in the BioID dataset with this method. Second, the FaceNet and the Resnet-50 architectures, which are the deep convolutional neural network (CNN), are proposed to use as the face encoding methods. Surprisingly, these two deep CNN architectures obtained an accuracy of 100% on the BioID and FERET datasets. Moreover, ResNet-50 architecture was slightly better than FaceNet architecture. The ResNet-50 and FaceNet architectures obtain very high verification accuracy on ColorFERET dataset, with accuracy of 99.60% and 99.32%, respectively.

References

[1] D. Li, H. Zhou, and K. M. Lam, "High-Resolution face verification using pore-scale facial features," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2317–2327, 2015.

[2] A. Singh, M. Singh, and B. Singh, "Face detection and eyes extraction using Sobel edge detection and morphological operations," in *Conference on Advances in Signal Processing (CASP)*, 2016, pp. 295–300.

[3] C. Li, Z. Qi, N. Jia, and J. Wu, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," in *IEEE 13th International Conference on Electronic Measurement and Instruments (ICEMI)*, 2017, pp. 483–487.

[4] E. K. Shimomoto, A. Kimura, and R. Belem, "A faster face detection method combining bayesian and Haar cascade classifiers," in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2015, pp. 7–12.

[5] A. Ade-ibijola and K. Aruleba, "Automatic attendance capturing using histogram of oriented gradients on facial images," in *IST-Africa Week Conference (IST-Africa)*, 2018, pp. 1–8.

[6] H. X. Jia and Y. J. Zhang, "Fast human detection by boosting histograms of oriented gradients," in *Proceedings of the Fourth International Conference on Image and Graphics Fast (ICIG)*, 2007, pp. 683–688.

[7] H. ChunYang and X. A. Wang, "Cascade face detection based on histograms of oriented gradients and support vector machine," in *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, 2015, pp. 766–770.

[8] H. Shu, D. Chen, Y. Li, and Shengjin Wang State, "A highly accurate facial region network for unconstrained face detection," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 665–669.

[9] L. Pang, Y. Ming, and L. Chao, "F-DR Net:Face detection and recognition in one net," in *International Conference on Signal Processing (ICSP)*, 2018, pp. 332–337.

[10] A. R. Rivera, J. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: face and expression recognition," *IEEE Trans. image Process.*, vol. 22, no. 5, pp. 1740–1752, 2013.

[11] F. Juefei-Xu and M. Savvides, "Encoding and decoding local binary patterns for harsh face illumination normalization," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3220–3224.

[12] D. Gong, Z. Li, W. Huang, X. Li, and D. Tao, "Heterogeneous face recognition: a common encoding feature discriminant approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2079–2089, 2017.

[13] A. Majumdar, R. Singh, and M. Vatsa, "Face verification via class sparsity based supervised encoding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1273–1280, 2017.

[14] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015, pp. 1–12.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.

[16] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 815–823, 2015.

[17] K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy, "Pose-Robust Face Recognition via Deep Residual Equivariant Mapping," in *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 5187–5196.

[18] D. E. King, "Max-margin Object Detection," 2015.

[19] O. Surinta and S. Khruahong, "Tracking people and objects with an autonomous unmanned aerial vehicle using face and color detection," in *International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*, 2019, pp. 206–210.

[20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[21] P. Viola and M. Jones, "Robust real-time object detection," *Vingtieme Siecle Rev. d'Histoire*, vol. 57, pp. 1–25, 2007.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[23] M. Karaaba, O. Surinta, L. Schomaker, and M. A. Wiering, "Robust face recognition by computing distances from multiple histograms of oriented gradients," in *IEEE Symposium Series on Computational Intelligence, (SSCI)*, 2015, pp. 203–209.

[24] V. N. Vapnik, *Statistical Learning Theory*. 1998.

[25] O. Surinta, M. F. Karaaba, L. R. B. Schomaker, and M. A. Wiering, "Recognition of handwritten characters using local gradient feature descriptors," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 405–414, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[27] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," in *Lecture Notes in Computer Science book series (LNCS)*, 2001, pp. 90–95.

[28] P. J. Phillips, P. J. Rauss, and S. a. Rizvi, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[29] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, pp. 295–306, 1998.