Face Recognition Framework Based on Correlated Images and Back-Propagation Neural Network

Mohannad Abuzneid

Department of Computer Science and Engineering
University of Bridgeport
Bridgeport, USA
Mohannad@my.bridgeport.edu

Ausif Mahmood

Department of Computer Science and Engineering
University of Bridgeport
Bridgeport, USA
Mahmood@bridgeport.edu

***Abstract*— The human face facial complexity and the face changes make the face recognition system a challenging task to design and difficult to implement. The correlation between the training images which has a high impact on the accuracy of the face recognition system never considered by researchers. In this paper, we presented an enhanced framework to improve the face recognition using the classical conventional Principal Component Analysis (PCA) and the Back-Propagation Neural Network (BPNN). A key contribution of this work is based on obtaining a robust training dataset called the T-Set using the correlation between all the images in the training dataset not based on the image density which adds a distinct layer between the dataset. We used the PCA descriptor for features extraction and dimension reduction to show that there is a promising enhancement even with using traditional algorithms. We combined five distance methods (Correlation, Euclidean, Canberra, Manhattan, and Mahalanobis) to obtain the T-Set using the square-root of the sum of the squares to achieve higher accuracy. We added a strength factor to each of the distance methods, and we achieved higher face recognition accuracy than the current approach. Our experimental results on YALE and ORL datasets demonstrate that the approach we proposed improved the accuracy of face recognition system with respect to the existing methods.**

***Keywords*- Principal component analysis (PCA); Back-Propagation neural network (BPN); K-Nearest-Neighbor; Image classification.**

I. Introduction

Face detection and recognition have turned into an exciting field for researchers. Some of the common areas that researchers mostly focus on are computer vision, image processing, and deep neural networks. The motivation behind the enormous demand is the need to improve the accuracy of the real-time applications such as the surveillance systems and to improve the static recognition system like Transportation Security Administration (TSA) and DMV systems. Countless methodologies have been acknowledged and presented in the past years. The illumination, facial expression, poses, appearance, aging variation, and the occlusion are some of the factors that make it a challengeable task to design and implement a robust face recognition system as shown in Fig. 1[1]. Human face personality changes have less effect comparing to the other challenges [2].

In the face recognition, image processing, and machine learning, feature extraction is needed to reduce the classification computing time which helps the system to converge faster and more accurate. Karhunen-Loeve expansion (PCA) [3] is one of the common methods used for features extraction and dimension reduction. We achieved a strong representation of the human face by retaining the most dissimilarities in the image data after reducing the dimensionality of the image which is called Eigen-Faces.

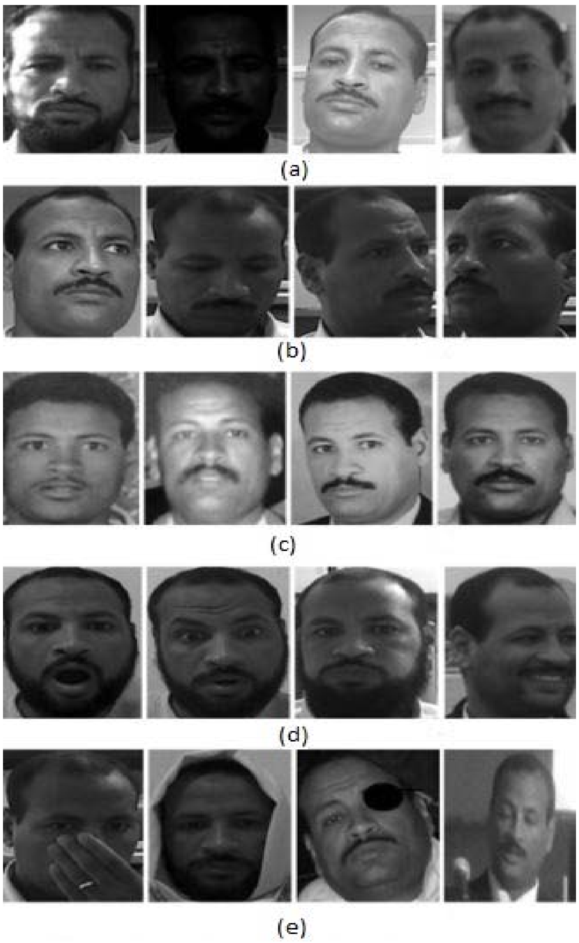

Fig. 1. Example of the challenges. (a) illumination variation (b) pose variation (c) ageing variation (d) expression variation (e) occlusion

Fig. 2 shows the classical face recognition system steps which start with some of the preprocessing methods such as, face detection, cropping, resizing and normalization. The second step using a powerful features extraction and dimension reduction algorithm to retain robust features and removing redundant features. PCA is one of these statistical approach transformations [4]. In addition, numerous algorithms introduced in the computer vision and object recognition field as Diagonal PCA [5], and Curvelet-based PCA [6]. Yang et al. [7] proposed Kernel PCA and Kernel FLD for object recognition. The modular PCA approach has achieved a high

978-1-5386-9120-5/18/$31.00 ©2018 IEEE

accuracy than PCA for the extreme changes in expressions, pose variations, and poor illumination [8].

Independent Component Analysis (ICA) was introduced by Bartlett et al. [9] and Draper et al. [10]. ICA is more powerful method than PCA because a better basis of the face images can be found which is a statistical independence of the estimated factors. However, Moghaddam claimed that ICA does not provide a significant advantage over PCA method [11]. Kernel PCA is an extension of the PCA, and Yang [12] showed that the Kernel PCA method outperforms the classical PCA by dividing the points into random groups. However, the computation time in the Kernel PCA and ICA is more expensive than PCA. 2-DPCA was introduced as a new approach for features extraction and representation by Yang et al. [13]. One advantage of the 2-DPCA is straightforward to implement because the features extraction is based on the image matrix. Another advantage of the 2-DPCA is that it achieves higher recognition accuracy than the PCA with less computational time. However, 2-DPCA needs more storage space than PCA because it needs more coefficients to represent the face image. Linear Discriminate Analysis (LDA) method was introduced as features extraction method for object recognition by Lu et al. [14]. Nonlinear problems were handled by a numerous kernel based classifier methods [15-18]. Martez et al. [19] claimed that the LDA outperform the PCA when the number of the training samples per class is large, otherwise the PCA will exceed the LDA.

Finally, classification method that uses a relevant set of features to characterize the dataset of classes. There are different types of classification methods: simple classifiers such as Nearest-Neighbor, Euclidean distance classifier, and Mahalanobis distance classifier. Powerful classification methods handle problems with many parameters such as deep neural networks, support vector machine, decision trees, and extreme learning machine [20-26].

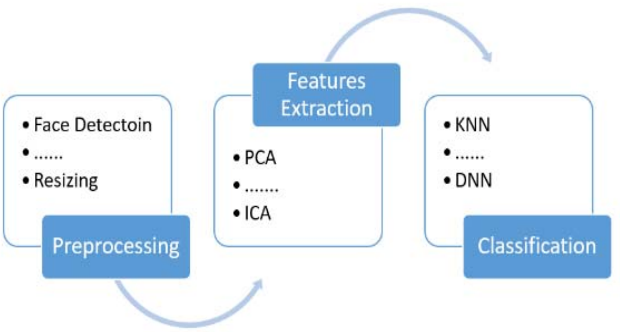

Fig. 2. Classical face recognation system steps.

In this paper, we propose a new framework of the face recognition system to enhance the recognition accuracy rate. The main contribution of this work is by adding another step after the features extraction to obtain a strong training data set called the T-Set. The T-Set have a robust distinction pattern which helps the Neural Network (NN) to converge faster and more accurate. The correlation between the training images has a high impact therefor; we considered it in our work, unlike the existing methods.

We followed the classical face recognition steps. However, we added an extra step to generate the T-Set after the features extraction step. Some of the preprocessing methods used such as face detection, cropping and resizing which helps to reduce the processing time. PCA is used to extract the important features from the image and to reduce the image's dimension. We used five distance methods and combined them to obtain the T-Set which we used as an input of the NN. Finally, we trained the NN offline and tested the framework on Yale and ORL datasets. We achieved a competitive accuracy rate compared to the existing methods.

This paper is organized as follows. In section II, we present the related work. In section III, we present an overview of the methods. In section IV, we explain the proposed framework. In section V, we present our experimental results. Section VI we present our conclusion of this paper and the future work.

II. Related Work

Lu et al. [27] used three holistic technique classifiers in their work: PCA, ICA, and LDA. The authors combined the output vectors of the three methods using the sum rule called matching scores. The concatenated vectors used as an input to the radial basis function (RBF) network to make the final decision. The drawback of this framework is, the three methods provide redundant futures which could lead to an over-fitting problem. Whereas in our proposed framework, we used one method to extract the important features and based on that, we generated our T-Set which we used to train the NN.

Eleyan et al. [28] proposed two approaches using different features extraction methods followed by a simple NN. First, they used the PCA to reduce the images dimension. Then, they replaced the PCA with the LDA, and the LDA with the NN method outperformed their proposed PCA with NN.

Lawrence et al. [29] used a local image sampling, convolution neural networks (CNN) and a self-organizing map (SOM) to propose their hybrid neural network framework. The drawback of this framework: SOM is invariance to minor changes in the images, and the classification process is based on the dimension reduction only using SOM. They achieved 96.2% accuracy on the ORL dataset. However, we outperformed their framework by achieving 96.9% accuracy rate based on the correlation between the training images.

Lone et al. [30] proposed a hybrid framework by combing different methods: PCA, Discrete Cosine Transform (DCT), Iterative Function System (PIFS) and Template Matching using correlation (Corr). They combined two methods using PCA-DCT, three methods using PCA-DCT-Corr, and finally four methods using PCA-DCT-Corr-PIFS. The proposed framework result shows combing the four methods outperformed combining two or three methods by achieving 86.6% accuracy on the ORL dataset.

Liong et al. [31] proposed a deep technique to extract a deep representation of the image which improved the face recognition accuracy. The proposed framework called the deep PCA because they used two layers and each layer has whitening and PCA.

Based on recent proposed works, it is obvious that combining multiple features extraction methods helps to improve the accuracy rate of the face recognition system. However, could lead to other problems. Therefore, we decided to use a single features extraction method and obtain a robust T-Set depending on the correlated training dataset. The new T-Set is used to train our NN.

III METHODOLOGY

A. *Principal Component Analysis (PCA).*

Face recognition and detection needs a huge resource of storage and a powerful system to reduce the computing time. Therefore, the dimension reduction and image re-representation are needed as an essential step in any face recognition systems. PCA is one of the common statistical transform methods. PCA reduces the image dimension by analyzing the image and identifying distinction patterns which can be used as a new representation of the image without losing an enormous information content from the image. Face dimension reduction is obtained by applying the PCA and finding the highest K Eigen-Values and their corresponding Eigen-Vectors. The face image can be re-represented using only 15% of the Eigen-Values with a minimal loss of information [32]. PCA is relying on the variance-covariance matrix. Therefore, the images are not a significant comparison to the number of face images in the training dataset. The advantages of the PCA are its low noise sensitivity, eliminating the data redundancy by providing the orthogonal components and reducing the image complexity. The disadvantage of the PCA method is its difficulty to compute the variance-covariance matrix and invariance unless the training dataset explicitly provides the information. Fig 3 shows an example of PCA methodology. In Fig. 3(a), the face data are randomly distributed, and Fig. 3(b) shows the correlated data grouped in the same coordinate face space.

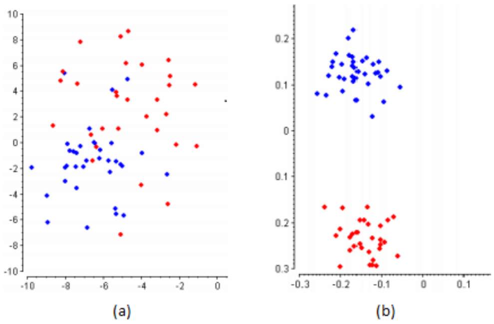

Fig. 3. An example of PCA (a) Original data. (b) Correlated data

The PCA process starts with standardizing the image scale by obtaining the same face size vector $\boldsymbol{\Gamma i}$ for all of the training images ***I1, I2… IM***, . The training dataset of ***M*** faces written as ***I = (I1, I2,…, IN)*** and the average image $\boldsymbol{\Psi}$ is found by equation (1).

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma i \tag{1}$$

Then, centralize each training image by subtracting the mean, which is the average across all dimensions from each image and finds the vector $\boldsymbol{\Phi = \Gamma i - \Psi}$. PCA uses equation (2) to calculate the covariance-matrix ***C***, which is used to find the Eigen-Value.

$$C = \frac{1}{M}\sum_{M=1}^{M}\Phi m\Phi m^{T} = AA^{T} \quad (M^2xM^2\ matrix) \tag{2}$$

$$where\ A = [\Phi_1\Phi_2 \ldots\ldots\ldots \Phi_M] \quad (N^2xM\ matrix)$$

We calculate the Eigen-Values $\boldsymbol{\mu i}$ and the Eigen-Vectors $\boldsymbol{vi}$ from the covariance matrix since the matrix is square using equation (3).

$$A^{T}Avi = \mu i\ vi \tag{3}$$

PCA significantly orders the Eigen-Values from highest to lowest. We then ignore low significance Eigen-Values to reduce the face domain, and we obtain the corresponding Eigen-Vectors. PCA loses some of the information from the image. However, this will not affect the recognition since the most of the data discrepancy exists in the first 15% of the face dimension. Eigen-Faces are obtained by transpose the Eigen-Vectors and multiply them by the original faces dataset. Eigen-Faces appears as ghostly faces in Figure 4.

$$\textit{Eigen-Faces} = (\textit{Original dataset})\ X\ (\textit{Eigen-Vectors}^{-1}) \tag{4}$$

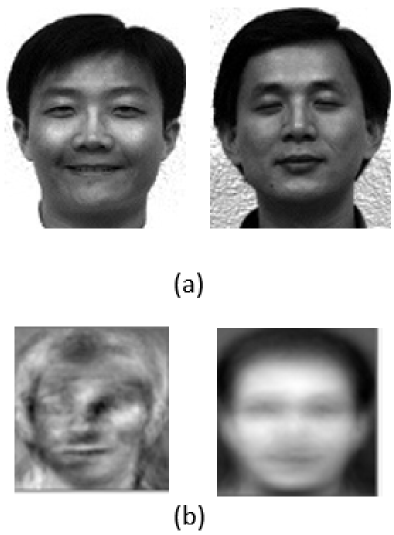

Fig.4. (a) Original images (b) The corresponding Eigen Faces

B. *K-Nearest-Neighbors (KNN) and the distance methods*

The KNN is used for both classification and regression cases and, it is considered as a simplest classification method. We used the KNN to find the expected output, and we chose the K=1 to avoid the majority voting which could lead to the over-fitting issue.

Distance method is used to measure the mismatches between two vectors. There are many of distance methods and each one has an advantage over the other in a different direction. Therefore, we used five of them in this paper to improve the recognition rate by combining them. The methods which we used:

$$Correlation\ (a,b) = \frac{\mathrm{covariance}\ (a,b)}{\sigma_a\sigma_b} \tag{5}$$

where $\boldsymbol{\sigma_a}$ and $\boldsymbol{\sigma_b}$ are the standard deviations of ***a*** and ***b***.

$$Euclidean(a,b) = \sqrt{\sum_{i=1}^{N} (a_i - b_i)^2} \quad (6)$$

$$Canberra(a,b) = \sum_{i=1}^{N} \frac{|a_i - b_i|}{|a_i| + |b_i|} \quad (7)$$

$$Manhattan(a,b) = \sum_{i=1}^{N} |a_i - b_i| \quad (8)$$

$$Mahalanobis(a,b) = \sqrt{(a_i - b_i)^T S^{-1} (a_i - b_i)} \quad (9)$$

where $\boldsymbol{S}$ is the covariance matrix.

C. Back-Propagation Neural Network (BPNN)

The Neural Network (NN) is one of the common classification method used in computer vision and machine learning. The BPNN is one of the supervised learning methods used in the face recognition system to achieve higher accuracy rate by using the gradient descent optimizer. It is called back-propagation because the error is calculated starting from the output layer and backward through the network layers. The BPNN has three-layer types; input, hidden and output layers. It is considered as a deep neural network when the BPNN has more than one hidden layer as shown in Fig. 5 [33]. It is easy to implement the BPNN, but it is a challenging task to select the optimal number of hidden layers and number of neurons of each layer.

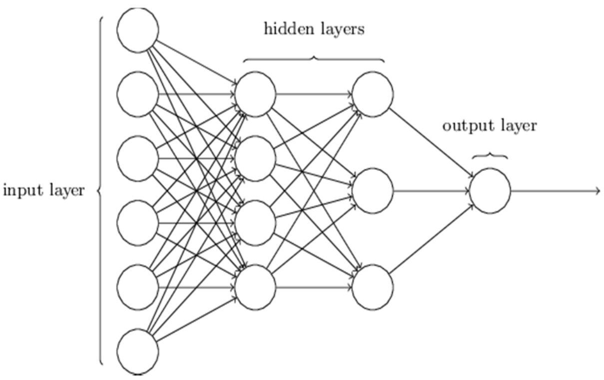

Fig. 5. BPNN with two hidden layers

The BPNN algorithm process is as follows:

1. Collect the training dataset to use it as an input of the BPNN and calculate the expected output of the system.
2. Determine the number of the hidden layers, the number of the neurons of each layer, the threshold and the number of iterations.
3. Randomly initialize the weights $\boldsymbol{\omega_i}$ and the biases $\boldsymbol{b_i}$ of all the hidden and the output layers in a small range.
4. Do step 5-12 while the stop condition is false.
5. For each training image, do step 6-11
6. For each hidden layer neurons:
   -Sum all the neuron input weight signals.
   -Calculate the output of each neuron using the activation function.
   - Send the output of each hidden neuron to all the neurons in the next hidden layer.
7. For each output layer neurons:
   -Sum all the neuron input weight signals.
   -Calculate the output of each neuron using the activation function.
8. Compute the error of the each of the output layer neuron then calculate the $\Delta\boldsymbol{\omega_i}$ and the $\Delta\boldsymbol{b_i}$ and send the updated info to the below layer.
9. Sum all the input weight signals of the output layer neurons then calculate the output using the activation function.
10. For each hidden layer neurons:
    -Sum all the neuron input signals from the above layer.
    -Calculate the error using the derivative of the activation function
    - Calculate the new weights $\boldsymbol{\omega_i}$ and the biases $\boldsymbol{b_i}$.
11. Update the $\boldsymbol{\omega_i}$ and the biases $\boldsymbol{b_i}$ of each neuron in the hidden and output layers.
12. Stop if the stop condition is true.

IV. Proposed Method

We proposed in this paper an enhanced Face Recognition Framework Based on Correlated Images and Back- Propagation Neural Network. The main contributions of our work are:

- Using five distance methods and combining them will provide a clear pattern which helps the NN to converge faster and more accurate.
- Obtaining the T-Set based on the correlation between the training dataset will provide a robust data which we used as an input of the NN.
- Each distance method performs well in a different direction. Therefore, adding a strength factor helps to improve the accuracy rate.

The proposed framework is divided into five steps as shown in Fig. 7.

1) Preprocessing step: Face recognition needs a huge storage and CPU resources. Therefore, we applied few of the preprocessing operations to reduce the computing time as shown in Fig. 6. Haar-cascade detection is used to detect the face then we cropped the face to reduce the background effect. We converted the images to a gray-scale image then we applied a histogram equalizer to reduce the noise effect. Finally, we resized the images to the size we preferred.

Fig. 6. Prepressing step: Cropping, resizing and histogram equalizer

2) Features extraction: We used the PCA algorithm to reduce the dimensionality of the images by eliminating the redundant data between the training images while retaining the variation between them. The PCA is transforming the dataset to a new set of variables which called the principal components (PCs). The first PC retains the maximum variation in the dataset. The PCA sorts the Eigen-Vectors and selects to top K values to reduce the dimensions. The training dataset much be scaled and the complexity of calculating the covariance matrix are some of the drawbacks of the PCA. However, we used the PCA to prove that there is a potential accuracy improvement using the traditional methods by adding an extra step (step 3) to obtain the T-Set based on the correlated training dataset images.

3) Obtaining the T-Set: We added this step to obtain a correlated dataset which we called the T-Set and use it as an input of our NN. The T-Set has strong distinction patterns which improved the overall accuracy rate of the face recognition system. The next steps are applied to each of the training images to obtain the T-Set:

- Based on the reduced dimension of each training image from step 2 and using Mahalanobis, Manhattan, Correlation, Canberra and Euclidean distance methods; we separately computed the distance between each training image and all other images.

- First, we trained our NN using each method individually, and we achieved different accuracy rates as shown in Table 1. Therefore, we decided to combine the five distance methods using equation (10).

$$RSS_\alpha = \sqrt{\sum_{i=1}^{5} \alpha_i DIS_i^2} \qquad (10)$$

where α_i is the strength factor of each distance method, and the sum of them is equal to 1.

Table 1 shows that Manhattan and Mahalanobis produced a higher accuracy rate comparing to the other methods. Therefore, we assign the strength factors as: Euclidean and Correlation = 0.1, Manhattan and Mahalanobis = 0.3 and finally, Canberra = 0.2.

- We used the KNN to find the expected output of the training images.

Table 2 shows how we calculated the T-Set (column 6) and the expected output (column 7) of one of the ORL training images (IMG Y). We used 200 images from the ORL dataset to train the NN.

4) Set up and train the NN: We start the training after we set up the NN parameters such as the number of the hidden layers, the number of the neurons of each layer, number of the iteration, threshold value, setup the input matrix and finally, setup the output matrix.

5) Testing the system: We found the reduced data of the testing image using the Eigen-Vectors which we obtained from step 2. Then, we fed the testing image the trained NN to calculate the predicted output label.

We computed the accuracy based on comparing the predicted label and the expected label. Finally, we calculated the overall accuracy rate of the framework.

TABLE 1
Experiment results using PCA descriptor and BPNN on the Yale and ORL Datasets.

Distance Method	**Recognition Rate (%)**	
	Yale	ORL
Canberra	90	83
Euclidean	79	86.2
Canberra	91.1	89.3
Manhattan	93	93.3
Mahalanobis	95	95.4
Proposed: $\sqrt{\sum_{i=1}^{5} \alpha_i DIS_i^2}$	**97**	**96.9**

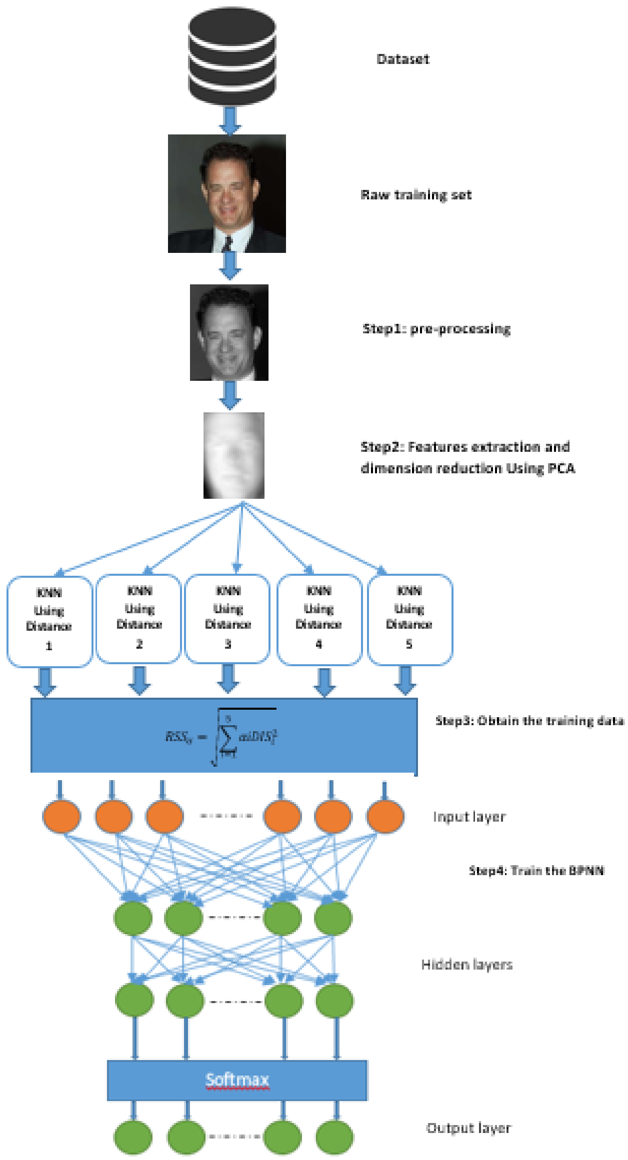

Fig. 7. The proposed framework

TABLE 2
Obtaining the new T-Set (column 6) and the expected out (column 7) for one of the training images (IMG Y).

		Column1	Column2	Column3	Column4	Column5	Column6	Column7
	Distance Between IMG Y and	**Using Correlation**	**Using Euclidean**	**Using Canberra**	**Using Manhattan**	**Using Mahalanobis**	**Combine the result using:** $RSS_\alpha = \sqrt{\sum_{i=1}^{5} \alpha_i DIS_i^2}$	**expected Output Based on the KNN (K=1)**
person1	Image1	0.39	0.34	0.71	0.34	0.56	1.091	1 (best match)
	Image2	0.31	0.32	0.67	0.42	0.49	1.033	
	Image3	0.41	0.28	0.77	0.29	0.39	1.037	
	Image4	0.29	0.29	0.64	0.27	0.48	1.533	
	Image5	0.72	0.54	0.85	0.48	0.79	1.544	
Person2	Image6	0.19	0.77	0.61	0.56	0.32	1.190	0
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Person40	Image197	0.34	0.41	0.57	0.32	0.74	1.121	0
	Image198	0.51	0.57	0.74	0.21	0.64	1.259	
	Image199	0.41	0.27	0.78	0.41	0.71	1.233	
	Image200	0.39	0.54	0.92	0.69	0.69	1.497	

V. EXPERMENTS SETUP AND DISCUSSION

To compare our results with the existing face recognition systems, we applied our framework to two well-known human face datasets, Yale and Olivetti Research Laboratory Human Face datasets.

Olivetti Research Laboratory Human Face Dataset (ORL) [34] has 400 of 92 X112 grayscale images which, represent 40 persons with ten pictures per person. We randomly divided the 400 images into two sets: Training dataset with a total of 200 and the remaining 200 used as a testing dataset. The images were captured under different expressions, pose, and gender.

The Yale face database [35] has a total of 165 face images represent 15 persons. Each person has 11 grayscale images with variant setups, poses, lighting, and expressions. We randomly divided the images into two sets: Training dataset with a total of 75 and the remaining 90 used as a testing dataset.

We obtained the T-Set and the predicted output as we mentioned in the proposed method section to use them as an input and output of the BPNN. In the ORL dataset experiment, we set up the BPNN as: The input matrix is 200 X 200 because we have 200 training images and each image represented by 200 points that we obtained from step 3. The predicted output of the ORL is 40 classes because the dataset represents 40 persons. We used two hidden layers in our BPNN with a 100 neurons per layer. The training rate is 0.01, and we trained the NN to a 100 epochs. In all hidden layers, we used sigmoid activation function, and we used a soft-max classifier at the output layer.

We used the setup in the Yale dataset experiment. However, the numbers are different because we used 75 training images that represents 15 persons. Therefore, the input matrix is 75 X 75, the output is 15 points per image, and the number of the neurons is 40 per layer with two hidden layers.

Table 1 shows that we achieved a variant recognition rate based on the individual distance method with the PCA and the BPNN. However, we achieved a higher accuracy rate when we combined five distance methods and added a strength factor to them to obtain the T-sat based on the correlation between the images. We achieved 97% average accuracy rate in the Yale dataset experiment with only three mismatching of 165 images and 96.9% average accuracy rate in the ORL dataset experiment with six mismatching.

VI. CONCLUSION

We proposed in this paper an enhanced framework for face recognition using the principal component analysis and the back-propagation neural network. We achieved the main contribution by obtaining the T-Set based on the correlation between the images which provided robust features. Combining five distance methods and adding a strength factor provided distinction patterns which helped the BPNN to converge faster and more accurate. We used a traditional features extraction such as PCA to prove that there is a potential improvement in the accuracy rate as Table 1 shows by considering the correlation between the images. Therefore, we will replace the PCA with a stronger features extraction method such as Local binary patterns (LBP) or the Convolution Neural Network (CNN) to compare it to our result in the future work.

REFERENCES

[1] M. Hassaballah and S. Aly, "Face recognition: challenges, achievements and future directions," in *IET Computer Vision*, vol. 9, no. 4, pp. 614-626, 8 2015.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces versus fisherfaces: Recognition using class specific linear

projection", IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 711–720, 1997.

[3] M. A. Turk and A. P. Pentland, "Eigenfaces for Recognition," J. Cogn. Neuroscience. vol. 3, pp. 71–6, 1991.

[4] L. Sirovich and M. Kirby, ªA Low-Dimensional Procedure for the Characterization of Human Faces,º J. Optical Soc. Am. A, vol. 4, no. 3, pp. 519-524, 1987

[5] D. Zhang, Z. Zhou, and S. Chen, "Diagonal Principal Component Analysis for Face Recognition," Pattern Recognition, vol 39, pp. 140142, 2006.

[6] T. Mandal, Q.M.J. Wu, and Y. Yuan, "Curvelet based face recognition via dimension reduction," Elsevier, Signal Processing, vol 89, no. 3, pp. 2345-2353, 2009.

[7] M.H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition (RGR'02), pp. 215-220, May 2002.

[8] R. Gottumukkal and V.K. Ansri, "An improved face recognition technique based on Mudular PCA approach," Pattern Recognition letters, vol 25, pp. 429-436, 2004.

[9] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Trans. Neural Networks, vol. 13, no. 6, pp. 1450-1464, 2002.

[10] B.A. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, "Recognizing Faces with PCA and ICA," Computer Vision and Image Understanding: special issue on face recognition, Academic press, 2003,pp. 115-137.

[11] B. Moghaddam. Principal manifolds and bayesian subspaces for visual recognition. In Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 1131{1136, 1999.

[12] P.C. Yuen and J.H. Lai, "Face representation using Independent Component Analysis," Pattern Recognition, vol. 35, no. 6, pp. 12471257, 2002.

[13] J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 131-137, Jan. 2004.

[14] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, " Face Recognition using LDA based Algorithms," IEEE Trans. on Neural Networks 14 (1) , pp 195-200, 2003.

[15] W. Zheng, L. Zhao, C. Zou A modified algorithm for generalized discriminant analysis. Neural Computation, 16(6): 1283–1297, 2004.

[16] J. Yang, A. F. Frangi, J. Y. Yang, D. Zhang, Z. Jin, KPCA Plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(2): 230–244, 2005.

[17] Wen-Sheng Chen and P. C. Yuen, "Interpolatory Mercer kernel construction for kernel direct LDA on face recognition," 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, 2009, pp. 857-860.

[18] J. Huang, P. C. Yuen, W. S. Chen, J. H. Lai, Face Recognition Using Kernel Subsapce LDA Algorithm. Proceeding of Asian Conference on Computer Vison (ACCV04), 1:61–66, 2004.

[19] A. Martez, A. Kak, "PCA versus LDA," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.23, no. 2, February 2001, pp.228-233.

[20] J. R. Solar, and P. Navarreto, "Eigen space-based face recognition: A comparative study of different approaches", IEEE Transactions on Systems man And Cybernetics- part c: Applications, vol. 35, no.3, 2005.

[21] Duda R, Hart P, Stork D. "Pattern Classification," 2nd Edition., Wiley Interscience, New York, (2001).

[22] N. A. Mahmon, N. Ya'acob, and A. L. Yusof, "Differences of Image Classification Techniques for Land Use and Land Cover Classification" IEEE Trans. CSPA, pp.90-94, 2015.

[23] M. J. Er, W. Chen, and S. Wu, "High speed face recognition based on discrete cosine transform and RBF neural network", IEEE Trans. on Neural Network, vol. 16, no.3, pp. 679-691, 2005.

[24] H. Othman, and T. Aboulnasr, "A separable low complexity 2D HMM with application to face recognition", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 25, no.10, pp. 1229-1238, 2003.

[25] K. Lee, Y. Chung, and H. Byun, "SVM based face verification with feature set of small size", Electronic letters, vol. 38, no.15, pp. 787-789, 2002.

[26] A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, and M.A. Sid-Ahmed, "Human face recognition based on Multidimensional PCA and extreme learning machine," Pattern Recognition, vol 44, pp. 2588-2597, 2011.

[27] X. Lu, Y. Wang, and A. K. Jain, "Combining classifiers for face recognition," in Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, 2003, pp. III-13-16 vol. 3.

[28] A. Eleyan and H. Demirel, "PCA and LDA based face recognition using feedforward neural network classifier," in Multimedia Content Representation, Classification and Security, ed: Springer, 2006, pp. 199-206.

[29] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," Neural Networks, IEEE Transactions on, vol. 8, pp. 98-113, 1997

[30] M. A. Lone, S. Zakariya, and R. Ali, "Automatic face recognition system by combining four individual algorithms," in Computational Intelligence and Communication Networks (CICN), 2011 International Conference on, 2011, pp. 222-226.

[31] V. E. Liong, J. Lu, and G. Wang, "Face recognition using Deep PCA," in Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on, 2013, pp. 1-5.

[32] M. Kirby and L. Sirvoich, "Application of the Karhunen-Loeve Procedure for the characterization of human faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[33] Ian Goodfellow Yoshua Bengio and Aaron Courville, "Deep learning," Book in preparation for MIT Press, 2016.

[34] ORL database: http://www.camorl.co.uk.

[35] Yale Face Database: http://cvc.yale.edu/projects/yalefaces/.