International Journal of Inventive Engineering and Sciences (IJIES)
ISSN: 2319–9598, Volume-1 Issue-11, October 2013

Face Recognition Based on Local Image Descriptor and Non-linear Features Extraction

Mina Hojjati

Abstract— ***In this paper, we introduce EABF (Extraction Analysis of Bsif Features) new method to face recognition based on extraction and analysis of binary sif features (BSIF). In our proposed algorithm, FABF eliminates some objections that led to many problems in the previous algorithms, such as a large query space and different quality and the size of images due to different time conditions for imaging and it removes the disadvantages of ELPDA (Nearby Local Discriminating Analysis) methods as a between-class-Scatter by using the Scatter matrix. This matrix introduces and updates the nearest neighbors to the outer class (K) through the samples. In addition, one of the advantages of the EABF is the high-speed face recognition by reducing the size of feature matrix and using NLPCA (Non-Linear Locality Preserving Analysis). Finally, the experiments results on the FERET data base indicate the impact of proposed method on the face recognition.***

Keywords—Face Recognition, Non-linear Features, Linear Features, Local Image Descriptor.

I. INTRODUCTION

In the recent years, the subspace learning methods have been considered with respect to their positions in the machine vision, the machine training and the pattern classification. Over the past two decades, many subspace learning methods have been proposed for face recognition. Generally, these methods can be divided into two groups: First, Unsupervised learning methods; and second, supervised learning methods [2-11].

In this paper, we introduce a new method to analysis the linear features extraction. This method reduces the feature matrix to 3*N (N is the number of image samples). Also, in order to show the size problem in the smaller samples, our objective function includes an inner-class scatter discriminating matrix. The results of the FERET data base indicate the impact of EANF method.

In this paper, we extract a feature of tested images based on BSIF binary models, so that we can make a classification by reducing the dimension and extracting the best feature of face.

II. THE FEATURE EXTRACTION BASED ON BSIF

This algorithm provides a method for the local image descriptor that encodes the model data effectively and it is very suitable for representing the images area. By lining up the local image segments in a subspace, this method defines a binary code for each pixel that resulted from the unit vectors of natural images analysis against the independent combined analyses as well as binarizing the unit vector coordinates against thresholding.

The length of binary code strings is determined by the numbers of unit vector and the image regions are obtained by the chart for binary codes of pixels. Our method has derived from the other descriptors, such as local phase quantization and local binary pattern that produce the binary codes. However, instead of the credit code structures, the proposed approach is based on the statistics of natural images and increase of modeling capacity. Having a part of image X sized L*L pixel and a Wi linear filter with the same size, the answer of filter si would be achieved using the following equation:

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) = \mathbf{w}_i^\top \mathbf{x},$$

In order to pass the important features through the filter and to create independent features, the images filters are designed through some images of face. The image 3 is an example of an image that is passed from BSIF algorithm. As you can see, varying lighting conditions have no impact on the image features extraction.

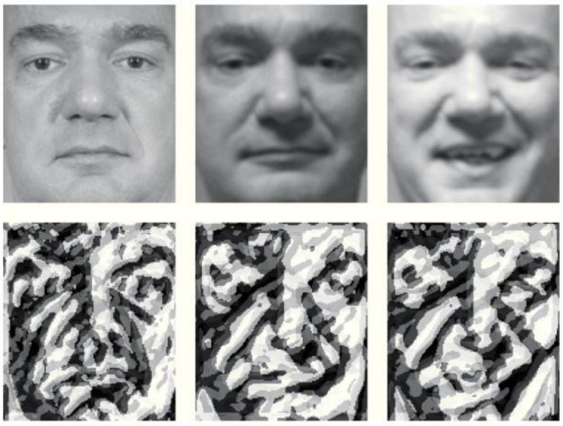

Fig.1. BSIF algorithm applied to the image

III. LOCAL NEARBY DISCRIMINATING FEATURE ANALYSIS

For example, when xi ∈ Rn, consider set X = [x1, x2, • • • , xN] from class C { ω1, ω2,…, ωC}. The subspace learning methods try to find the transfer function Φ, as it is possible to transfer from the next space n to the next space d (d≪n) by minimizing and maximizing the objective function. When yi ∈ Rd, the transfer function Φ is equal to yi = ΦT xi, so that it reduces the space dimension for the interclass scatter matrix and it increases the space dimension for the intraclass scatter matrix. The structures of interclass scatter matrix and intraclass scatter matrix are as follow, respectively:

Manuscript received October, 2013.
Mina Hojjati, Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University,Qazvin, Iran.

$$S_b^L = \sum_{ci=1}^{C} \sum_{cj=1}^{C} (\bar{x}c_i - \bar{x}c_j) W_{c_i c_j} (\bar{x}c_i - \bar{x}c_j)^T \quad (1)$$

$$S_W^L = \sum_{c=1}^{C} \sum_{xi,xj \in w0}^{1} (x_i - x_j) W^{(c)}{}_{ij} (x_i - x_j)^T \quad (2)$$

where, $\bar{x}\, c_i$ is the average amount of vector ωci. The weight wij(C) for both Cth interclass data, x_i and x_j, is as follow:

$$= \begin{cases} \exp\left(-\frac{\|\bar{x}c_i - \bar{x}c_j\|^2}{2\sigma^2}\right) \\ 0, \end{cases} \text{Wci cj} \quad (3)$$

$$= \begin{cases} \exp\left(-\frac{\|xc_i - xc_j\|^2}{2\sigma^2}\right) \\ 0, \end{cases} W_{ij}^{(c)} \quad (4)$$

σ is also determined experimentally.

Due to using the linear PCA, the ELPDA method has two problems, i.e. the high volume of feature vector and the feature matrix error in returning the primary value of matrix, that the problem would be solved by using NLPCA.

This function chooses all of the non-linear and linear points and leads to remove the additional points and to loss the accuracy in choosing components. As a result, we can recognize the face faster. [23,24]

In differentiating data that their classes are not initially discriminated as linear ones, linear transformations would be done weakly. The non-linear Method of Principal Components Analysis (NLPCA), similar to PCA, is used to determine and decrease the data correlation. PCA method indicates the linear correlation between the features, but NLPCA indicates the linear and non-linear correlation between the features without respect to the linear nature of data. In NLPCA method, a neuron system is trained to determine the nonlinear registration. [20,23-25]

Nonlinear P (principal) Components Analysis (NLPCA) is a nonlinear generalization for the PCA (linear) standard method. So far, many of the generalizations relied on a kind of learning. In this paper, in order to recognize the face, we propose an algorithm at which extending PCA to NLPCA would be made through a hierarchical learning. [24,24]

When using any kind of linear and nonlinear analyses (PCA), it is important that there is a differentiation to use them in decreasing the dimension and identifying correctly a special set of features that are based on the certain criterions. In the first set of applications, by emphasizing on the data compaction and a non-noisy state, only a subspace with a high power of description would be appeared. It is unnecessary to be alternative each of the features. It is just necessary to explain the mean square error (MSE) by the related subspace as much as the data information.

Implementing a PCA hierarchical algorithm has two important characteristics, i.e. scalability and stability. The first refers to its ability to respond to the increase of workload conveniently or it indicates the system ability to increase workload. For example, the scalability refers to a system ability to increase the total performance when adding resources (e.g. dimensions), and the second means that the ith characteristics of feature n have i solutions in return for each m feature, as (m ~= n), thus it would be stable and we can detect the primary matrix. The autoencode can extract the remaining features on the error variance by training the hierarchical features. Nevertheless, for the nonlinear data, this method is not good enough. The variance reminder can't be considered without respect to the nonlinear registration [15].

Linear (PCA)

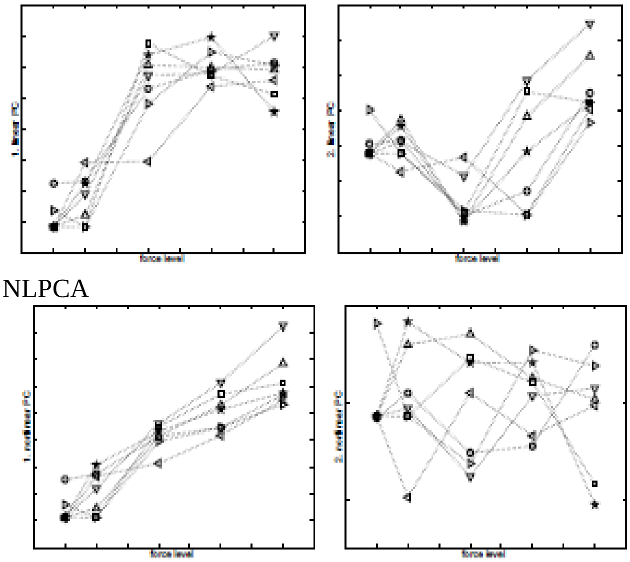

Fig.1. Linear and non-linear PCA features in 5 force levels.

As you can see in the Fig.1, the linear PCA loses some areas and it chooses a large range that increases complication. [17]

IV. TRANSFORMING THE LINEAR PCA TO THE NONLINEAR PCA

Linear autoencoder development includes nonlinear registrations that would be generated by adding hidden nonlinear layers. The Fig.2 indicates the strategy that there is a NLPCA (Nonlinear Principal Component Analysis) in its center. [18,19]

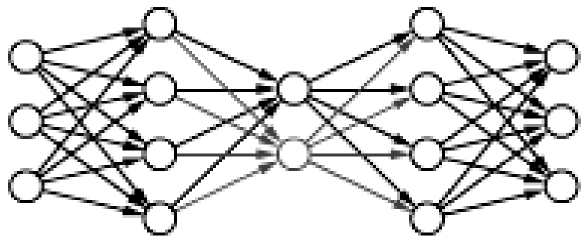

Fig.2. Encoding and decoding the network nodes trough registration [4-3-2-3-4]

There are two general solutions to introduce the nonlinear features extraction method in the feature space. First, similar to the linear PCA, the ith feature is be used in order to calculate the highest ith variance. The second strategy is to search the main data space for the smallest mean square error (MSE) in return for each i feature. Of course, finding a solution for the first strategy is harder than the second strategy. The formula to calculate MSE is as follow [21]:

$$E = \frac{1}{dN} \sum_{n}^{N} \sum_{k}^{d} (x_k^n - x^{\wedge}{}_k^n)^{2!} \quad (5)$$

where X and ^x are the pricipal data and supplementary data respectively. n is the number of sample and d is its dimension. All of the results can be generalized to any dimension. E1 and E1,2 are mean errors when they be calculated by using one or two features.

To perform h-NLPCA, we need to the minimum amount of E1 and E1,2, and we can make a hierarchical method by minimizing the error.

EH = E1 + E1,2 (6)

At the present, we should make a balance between the weight amounts of E1 and E1,2 and parameter α:

E H = αE1 + E1,2 ; α∈ (0,∞) (6)

Nevertheless, in order to select the optimal component α through the computational costs, …….(page 4). As shown in the Fig. 3.

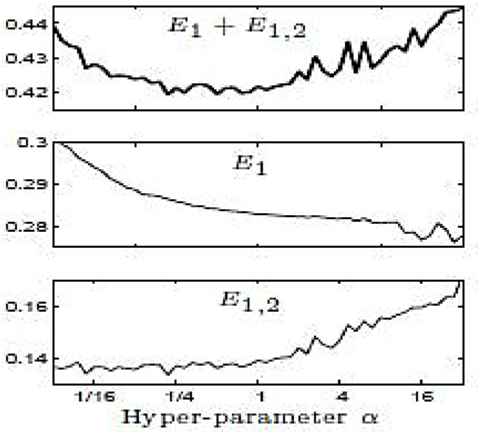

Fig.3. The correlation between the errors via the component α. In the error E1, the component α is the zero and in the error E1,2 , α has an infinite value.

In order to train the h-NLPCA through the gradient descent, we would calculate the amounts of E1 and E1,2 errors in each training independently. In the s-NLPCA, this method would be calculated through the network with one or two neurons in the features layer. As the gradient rEH is equal to the sum of rEH = rE1 + rE1,2 and the weight reduction would be calculated as follows:

E = EH + ν Pi w2i (8)

In most experiments, ν = 0/001 is a good selection. In addition, in order to achieve a more optimal result, the weight of nonlinear layer has been initialized. Thus, in a linear system, the sigmoidal works as a nonlinear one, that this is related to starting the h-NLAPCA network using a simple solution of PCA.

V. ACCURACY OF CLASSIFICATION

The experiment conducted includes the NLPCA features for classifying. Here, we have a set of fifty-dimensional samples that belong to two classes with two levels of classification including the image 1 and the image 2 and the samples of the classes A and B.

We use the nonlinear quadratic equation to transfer the nonlinear data. [27]

Table 1.The error rate of the experiment data through linear and nonlinear PCA

	Classification from 1 to 2				
Features	**1**	**2**	**3**	**10**	**20**
PCA	40/6	30/3	30/3	-	-
NLPCA	48/5	38/3	-	-	-

	Classification from A to B				
Features	**1**	**2**	**3**	**10**	**20**
PCA	48/3	50/0	32/0	-	-
NLPCA	50/0	50/0	-	-	

With respect to the Table 1, it is obvious that each of them should have 2 features of each image and the other three rows are related to the general characteristics of the feature matrix. In both of these classification methods, the NLPCA gives the best amount of the smallest component. As mentioned above, the linear PCA has made a weak classification. [22,26,28]

(Linear) PCA NLPCA

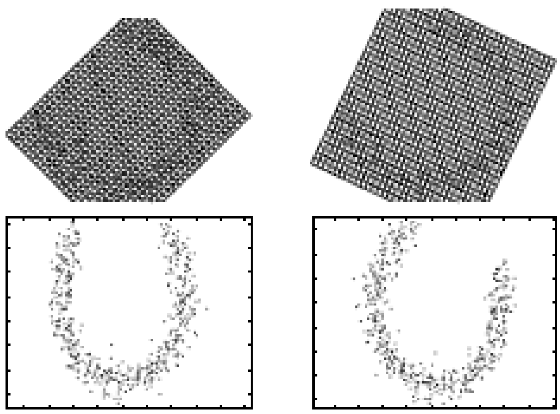

Fig.4. The output of linear and linear PCA method is the image noises.

VI. THE EANF METHOD EXPERIMENT TO CLASSIFY THE FACE

In this section, we would compare the EANF method performance introduced in the standard face data base of the FERET. The FERET face data base is a standard data base to evaluate the techniques of face recognition, including 14126 images of 1199 people. Also, our selective subset includes 1131 full face images of 229 people. [14]

All of the images were turned and evaluated in two subsets. Thus, the eyes centers have been set on the special pixels and then they have been transformed in to the 32 x 32 pixel images. The number of training stages includes 20 repetitions and the simplest classification is related to calculate the nearest neighbor.

Fig.5. FEREST data base images

A. Selection of parameter eabf

In the EABF, in order to make the interaclass scatter matrix, the number of neighbors near to the outer class K must be determined and each image has a different tag that it increases the efficiency. In order to demonstrate this method, we examine the nearby K impact on the basic rate of recognition in the FERET data base: In this experiment, we would select four samples for each chosen training class.

As you can see in the Figure, the recognition rate curve can be resulted by changing K from 1 to 50, and when it would be in a maximum level that k=1. [13]

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In similar experiments with different experimental samples on the FERET data base, the EABF method always indicates the most important recognition rate when k=1.

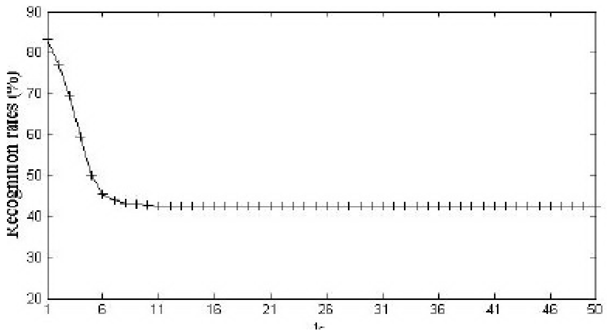

Fig. 6. Parameter k and the recognition rate chart

B. The experiment on the face recognition subsets

In this section, we randomly use the proposed EABF with five subspace learning algorithms (i.e. LDA, PCA [1], NLDA, ELPDA, LPDA and LDA: empty space) and some different images of each person for the experiment. The following Figure shows the comparison results for the recognition rate of six algorithms on the FERET data base. Table 2 indicates the beat recognition rate of six algorithms regarding the same numbers of dimensions.

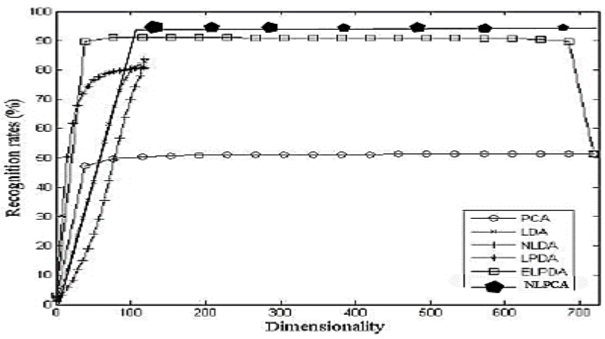

Fig. 7. The FERET Data base Images' Recognition Rate Chart

Fig. 2. The FERET Data base Images' Recognition Chart

Training	**PCA**	**LDA**
2	54/9±1/39(457)	61/8±1/95(109)
3	62/5±1/92(556)	68/9±2/44(228)
Training	**LPDA**	**ELPDA**
2	67/7±1/60(61)	69/0±1/62(289)
3	78/7±1/87(85)	82/5±1/56(186)
Training	**NLDA**	**EABF**
2	70/5±1/47(228)	77/7±1/60(61)
3	78/2±1/96(228)	89/5±1/56(186)

VII. CONCLUSIONS

In this paper, we introduce a new method based on the BSIF features and extraction and a set features selection by using the nonlinear PCA (NLPCA). So we can improve the features calculations speed, the dimension reduction and the high accuracy more than the nonlinear and linear method. In addition, we can classify the images with a small error, high speed and with any size and no additional computations over the other algorithms. As we showed before, this algorithm contains a better result than the other methods.

REFERENCES

1. P. Baldi and K. Hornik. Neural networks and principal component analysis:Learning from examples without local minima. Neural Networks, 2:53 – 58,1989.
2. T. Hastie and W. Stuetzle. Principal curves. JASA, 84:502–516, 1989.
3. M. Kramer. Nonlinear principal component analysis using auto-associative neural networks. AIChE Journal, 37(2):233–243, 1991.
4. D. Mewett, K. Reynolds, and H. Nazeran. Principal components of recurrence quantification analysis of EMG. Proceedings of the 23rd Annual IEEE/EMBS Conference, Oct.25-28 2001.
5. E. Oja. The nonlinear PCA learning rule in independent component analysis. Neurocomputing, 17(1):25 – 6,1997.
6. B. Sch¨olkopf, A. Smola, and K.-R. M¨uller. Nonlinear component analysis as akernel eigenvalue problem. Neural Computation, 10:1299–1319, 1998.
7.] M. Scholz. Nonlinear PCA based on neural networks. Diploma Thesis, Dep. ofComputer Science, Humboldt-University Berlin, in preparation. In German.
8. J. Stock and M. Stock. Quantitative stellar spectral classification. Revista Mexi-cana de Astronomia y Astrofisica, 34:143 – 156, 1999.
9. V. Vapnik. The nature of statistical learning theory. Springer, New York, 1995.
10. C. Webber Jr and J. Zbilut. Dynamical assessment of physiological systems andstates using recurrence plot strategies. J. of Appl. Physiology, 76:965 – 973, 1994.
11. M. Turk and A. Pentland, “Eigenfaces for recognition,” J.Cognitive Neurosci., vol. 3, no. 1, pp. 71–86, 1991.
12. X. He, S. Yan, Y. Hu, P. Niyogi, and Z. Hongjiang, “Facerecognition using Laplacianfaces,” IEEE Trans. Pattern Anal.Mach. Intell., vol. 27, no. 3, pp. 328–340, 2005.
13. X. He, D. Cai, S. Yan, and H. Zhang, “Neighborhood preserving embedding,” in Proc. IEEE Int. Conf. Comput. Vis., 2005, pp. 1208–1213.
14. S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, pp. 2323–2326, 2000.
15. J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A globalgeometric framework for nonlinear dimensionality reduction,”Science, vol. 290, pp. 2319–2323, 2000.
16. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Anal. Mach. Intell.,vol. 19, no. 7, pp. 711–720, 1997.
17. S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, “Grapha embedding and extensions: A general framework for dimensionality reduction,” IEEE Trans. Pattern Anal. Mach Intell., vol. 29, no. 1, pp. 40–51, 2007.
18. L. Yang, W. Gong, X. Gu, W. Li, and Y. Liu, “Bagging null space locality preserving discriminant classifiers for face recognition,” Pattern Recognit., vol. 42, no. 9, pp. 1853–1858,2009.
19. L.-F. Chen, H.-Y. Mark Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, “A new LDA-based face recognition system which can solve the small sample size problem,” Pattern Recognit.,vol. 33, pp. 1713–1726, 2000.
20. D. Tao, X. Li, X. Wu, and S. J. Maybank, “Geometric mean for subspace selection,” IEEE Trans. Pattern Anal. Mach.Intell., vol. 31, no. 2, pp. 260–274, 2009.
21. Z. Li, D. Lin, and X. Tang, “Nonparametric iscriminant analysis for face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 4, pp. 755–761, 2009.
22. T. Zhang, K. Huang, X. Li, J. Yang, and D. Tao, Discriminative orthogonal neighborhood-preserving projections for classification,” IEEE Trans. Syst. Man Cy. B, vol. 40, no. 1,pp. 253–263, 2010.
23. A. M. Martinez and R. Benavente, “The AR face database,”CVC Technical Report #24, Tech. Rep., 1998.
24. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face recognition.