Volume 3, Issue 7, July 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Gabor Phase Features Selection for Facial Recognition

Salima Nebti
Ferhat Abass University, Department of Computer Science,
Algeria, Sétif, 19000

Abstract— In this work, we first used binary particle swarm optimization (BPSO) for Gabor Phase features selection and then we used Euclidean distance matching or Fuzzy Support Vector Machines (FSVM) for classification to improve performance of face recognition. Results have been compared based on different space reduction algorithms namely PCA, LDA, KPCA and KFA. Results on ORL database show the robustness of the proposed approach and the effectiveness of FSVM for face recognition.

Keywords— Gabor phase, features selection, binary PSO, Fuzzy SVM.

I. INTRODUCTION

Face recognition has received an increasing interest as it represents an efficient way to protect personal data in electronic transactions. A recognition system helps to prevent the fraud of identity cards, to control the access to banks and airports and to pursuit criminals by external systems of cameras. These innovations appeared thanks to the development of new 3d cameras able to solve specific problems such as illumination and face detection. Face recognition: is the automatic processing of numerical person faces in order to identify, to verify or to categorize these persons. Face recognition methods can be distinguished in two large axes: recognition of predetermined images or recognition of extracted images from video sequences. The recognition of video faces is more useful since the use of simultaneous temporal and space information helps in many security systems such as pursuit of criminals or stealers. For example, a face recognition system compares the stored criminal faces with those of suspects, which justify the difficulty of its automation. In spite of the high quality results obtained in face recognition, a robust recognition system remains a challenging problem. The present methods are effective when assay and learning data faces are similar. However, the variability of illumination, the changes of face expression and pose cause serious problems for many existent recognition systems. Also, working on automatic recognition requires taking into account, not only, the face position and shape but also the possibility to take glasses or moustache [1]. A solution to this problem, is the use of three-dimensional information, this type of face representation is invariant to changes of illumination and pose [2]. An overview of the various problems encountered in face recognition and their processing may be found in [3].

Automatic face recognition involves several stages [4]:

a) Image acquisition: is the process of faces capturing and their conversion to numerical format. b) Face detection: is the process of face localization; a set of pre-processing operations such as binarization and segmentation are often necessary to detect the presence or absence of faces. Also, the system must be able to track faces from an image to the next in the case of video sequences. c) Normalization: is the process of conversion in standard dimensions d) Features extraction: is the process of determination of distinctive features between person's faces. e) Storing: if the system faces a new person face, so it will be saved for later comparisons. f) Comparison: is the process of similarity measurement between a face and others previously recorded in the system. The aim of comparison is face identification, verification or classification.

Face recognition methods can be distinguished into three families: The first are global methods; they are the most popular and use all the face area as input to the recognition system. The most largely used are eigenfaces, PCA, Linear Fisher discriminant analysis (LDA), DLDA, support vector machines (SVMs) …etc. The global representation saves all the texture information and useful forms for faces discrimination. The second ones are local methods: these latter extract local characteristics, and then use their local statistics as input of recognition system. An example of this category is LFA algorithm (local feature analysis) [5]. The local methods can provide additional information based on local parts. Furthermore, for each type of local characteristics, we can choose the appropriate classifier. The hybrid algorithms combine both local and global methods such as modular eigenfaces [6]. They are based on the idea that local and global features can offer complementary information and thus very useful for classification. Based on these old methods, enhanced and more robust methods have been proposed such as kernel based solutions [7,8], which allowed a significant improvement of recognition and LBP method (local binary pattern), which greatly improved the performance and speed of recognition [9, 10].

This work is devoted to improve recognition of human faces through bio-inspired features selector and a robust classifier namely the Binary Particle Swarm Optimizer (BPSO) applied to select the most important Gabor Phase features and FSVM used for classification. Thus, the selected subset is considered as training set of Fuzzy Support vector machines classifier, results have been found very promising. In the literature, there exist abundant works on Swarm intelligence based features selection for facial recognition to generate excellent recognition performance with small subset of features such as: the work in [21] where PSO-based feature selector is used to search the optimal subset of the extracted coefficients by two types of features: the discrete cosine transform (DCT) and the discrete wavelet transform (DWT). In [22] a feature selection algorithm based on Bacteria Foraging Optimization (BFO) is applied to coefficients extracted by discrete cosine transforms (DCT). In the work published in [23] A Binary Particle Swarm Optimization (BPSO)-based feature selection is used to search the optimal subset of Block-Based Discrete Cosine Transform (BBDCT) features.

The rest of paper is organized as follows: in section 2, we present how face recognition is solved through the phase of Gabor filters, BPSO algorithm and FSVM classifier. Next, the obtained results are shown and discussed in section 3. Finally, section 4 concludes with some perspectives.

II. PROBLEM FORMULATION

As mentioned above our face recognition system involves the following three stages

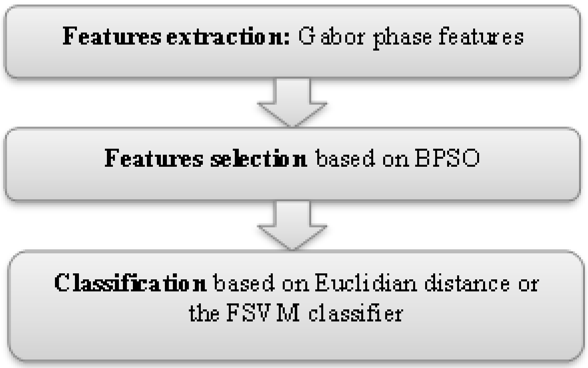

Fig. 1 Face recognition stages

A. Gabor phase features

Gabor wavelets also called Gabor filters represent a powerful feature extractor in facial recognition due to their interesting properties such as representation in different orientations and resolutions. Moreover, they are robust with changes of illumination, distortions and scaling [11]. The works in [12, 13] indicate that Gabor wavelets are invariant to changes caused by variations of illumination or modifications of facial expressions.

A Gabor kernel filter is the product of a complex sinusoidal wave with a Gaussian envelope as defined below:

$$\psi_{u,v}(Z) = \frac{\left\|k_{u,v}\right\|^2}{\sigma^2} e^{\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{u;v} z} - e^{-\frac{\sigma^2}{2}}\right] \quad (1)$$

Where z = (x, y) is an image point in the coordinates (x, y). u and v are two parameters which define the Gabor kernels' orientation and frequency. ‖ ‖ is the norm operator, and δ is the standard deviation of the Gaussian envelope. The Gabor wavelet representation results from the convolution of the input image with a set of Gabor kernels having different frequency and orientations as defined by Equation (1) [14].

The convolution of an image I and a Gabor kernel is defined as:

$$G_{u,v}(z) = I(z) * \psi_{u,v}(Z) \qquad (2)$$

$G_{u,v}(y,x)$ is the convolution result which can be decomposed into real and imaginary parts [15]

$$E_{u,v}(y,x) = Re\left[G_{u,v}(y,x)\right]$$
$$O_{u,v}(y,x) = Im\left[G_{u,v}(y,x)\right] \quad (3)$$

Based on convolution results, the Gabor phase responses $\phi_{u,v}(y,x)$ and its amplitude $A_{u;v}(y,x)A_{u;v}(y,x)$ can be evaluated by the following equation [15]:

$$A_{u,v}(y,x) = \sqrt{E_{u,v}^2(y,x) + O_{u,v}^2(y,x)}$$
$$\phi_{u,v}(y,x) = \arctan\left(O_{u,v}(y,x)/E_{u,v}(y,x)\right) \quad (4)$$

Features extraction: Gabor phase features

In our experiments, we used the Gabor phase with 40 Gabor kernels (5 scales and 8 orientations). The convolution of an image with a bank of 40 Gabor kernels (5 scales and 8 orientations) produces 40 phase cards that have the same size as the original face.

B. Features selection using BPSO

Features selection is a critical step for facial recognition and has many other fields of application such as classification, learning and Data Mining. Feature selection is a research area which extracts from the whole set of features an optimal set of the most appropriate features in order to improve the performance of a given system. [16].

BPSO is an iterative algorithm inspired from PSO to search for an optimal binary solution; it manipulates a population of a predetermined size. Each particle is a potential solution which consists of binary elements 0 or 1 distributed in random bits within the particle.

For each iteration: a new population with the same initial number of particles is produced. The creation of a new population is based on the previous one by using the BPSO rules to displace particles towards the best found solutions.

1) Particle swarm optimization (PSO)

PSO is a population based optimization method. The basic idea is that a group of unintelligent particles may produce global intelligent organization. Thus, thanks to simple rules, particles can gradually converge to optimal solution. However, this method seems to be well suited for continuous problems [17]. For this purpose, BPSO has been introduced to solve binary problems.

PSO begins by placing particles into the search space in a random way and then each particle moves according to three factors [17]:

1. Its current speed: V_{id}
2. Its personal best solution P_{id}
3. The best found solution in its neighborhood: P_{gd}

This is through the following equations:

$$\begin{cases} V_{id} = wV_{id} + c_1 \times r_1 \times (p_{id} - x_{id}) + c_2 \times r_2 \times (p_{gd} - x_{id}) & (5) \\ x_{id} = x_{id} + v_{id} & (6) \end{cases}$$

Where:
X: is the particle position.
V: is its speed.
i: is its index.
d: is the d th dimension into the search space.
Pi and Pg are respectively its personal best position and the best position found in its neighborhood.
C_1 and C_2 are respectively the cognitive and social factors which control the individual and collective behaviors of each particle.
r_1 and r_2 are two uniformly distributed values into the range [0,1] to allow better exploration.
W: is called the inertia weight, it is used to control the balance between exploration and exploitation

2) Binary particle swarm optimization (BPSO)

Kennedy and Eberhart proposed a discrete PSO to solve binary problems. In their model a particle contains values such as: 0 or 1, "yes" or "no", "true" or "false", ... etc.. In the binary PSO, personal solutions and the best ones are updated using PSO equations (5). The main difference between continuous and binary PSO is that particles' velocities are defined as probabilities limited into the range [0,1]. The function used for speed normalization in [0,1] is the sigmoid function (7) [17]:

$$V'_{id}(t) = sig(v_{id}(t)) = \frac{1}{1 + e^{-v_{id}(t)}} \quad (7)$$

Thus equation (5) is used to update the particles velocities, and the following equation (8) is used to compute their new positions:

$$x_{id}(t+1) = \begin{cases} 1 \; if \; r_{id} < sig\left(v_{id}(t+1)\right) \\ 0 \; otherwise \end{cases} \quad (8)$$

Where r_{id} is a uniform random number into the range [0,1]

3) The objective function

In BPSO, particles evaluation is based on a "fitness" function. We used the sum of squared errors as objective function where the selected features are considered as centres of groups to choose only the representative features. Thus, the selected features are considered as centres of the available face classes.

The sum of squared errors (SSE) criterion is often used for clustering.

Its formula is as below:

$$SSE = \sum_{i=1}^{M} \sum_{x \in C_i}^{N} \| x - \overline{C}_i \|^2 \quad (9)$$

Where:
M: is the number of face classes
N: is the number of faces per class
$\overline{C}_i$: Are the centres of classes: C_i

The objective of BPSO algorithm is to minimize this criterion in order to reduce the number of selected features by considering only those that are the most discriminant.

4) Algorithm

After construction of train and test data by evaluation of their Gabor phase features, we proceed to features selection based on BPSO. The purpose of BPSO is to minimize the fitness function (SSE) to find the best particle which represents the subset of the most discriminant features in order to save classification time and its accurateness in terms of recognition rate.

The initial population is randomly created; a particle is a binary vector of size M where M is the number of initial feature vectors. The i^{th} bit of the particle is 1 if the i^{th} feature vector is included in the learning set and 0 otherwise.

First, 20 particles are randomly initialized then for each iteration, the algorithm evaluates the particles fitness through the used SSE function. After evaluation, particles are displaced towards the best found positions using the recursive BPSO equations. The algorithm ends when the number of iterations reaches 30 iterations. The result is the best set of selected features given by the best found particle.

BPSO for features selection

Initialize a number of particles (xi)

For a number of iterations do
For i= 1 to number of particles do
Evaluate particles fitness using equation (9)

If fitness (*xi*) < Pbest then
Update Pbest = *xi* % Pbest is the personal best solution
End %**if**

If fitness(*xi*) <gbest then
Update gbest = *xi* % gbest is the best found solution
End %**if**

For *d* = 1 to number of dimensions do
Update positions and their speed using equations (5) & (6)
Normalize particles speed using equation (7)
Evaluate new positions using equation (8)
End for % dimensions

End for % particles
End for % iterations

C. Classification

We used matching based on Euclidian distance and Fuzzy support vector machines classifier (FSVM). FSVM is an improved pairwise classifier of SVM proposed to deal with unclassifiable Regions; a fuzzy membership function is introduced into the pairwise decision function [18]. Thus a fuzzy membership is assigned to each data point [19]. The fuzzy membership indicates to what degree the corresponding data point belongs to each of the two classes {−1, +1} [20]. FSVM generates better classification results than with SVM for data in the unclassifiable regions [18].

1) The basic SVM

SVMs are based on two main ideas: the notion of margin maximization and the concept of kernel function. Margin maximization is used for linear classification problems. Its objective is to separation between two classes by a hyper-plan so that their distance from the support vectors is maximized [24]. Kernel functions are used in the case of non-linear classification problems to transform the space of input data into a higher space in which it is possible to find linear separators [25].

Suppose the training points are N pairs of (x_i, y_i) i =1:N where x_i IS a data point and $y_i \in \{-1, 1\}$ is its label (y_i =1 for class 1, and -1 for class 2).

The linearly separable case

the optimal separation hyperplan is a margin classifier whose output is given by the following equation [26], [24]:

$$G(x) = sign\ (x^T B + B_0) \qquad (10)$$

Where B is the unit vector i.e $\|B\| = 1$ and the bias B_0 is adjusted by learning through margin maximization 1 /||B|| with the constraint of N points classification must be outside the margin [26]:

$$\begin{cases} \min \|B\| \\ \text{with} \quad y_i(x_i^T B + B_0) \geq 1, \quad i = 1, \dots\dots, N \end{cases} \qquad (11)$$

2) Fuzzy SVM

In this work, we have considered a Gaussian function for weighting data points as described in equation (12) . That is uncertain points should have lower weights [27].

$$W(x_i) = \prod_{j=1}^{P} exp^{\frac{-(x_{ij} - \mu_{jk})^2}{2\sigma_{jK}^2}} \quad \forall x_i \in class\ k \qquad (12)$$

Where μ_{jk} *and* σ_{jk} are respectively the mean and standard deviation of jth feature of all k-class points and x_{ij} indicates the j^{th} feature of i^{th} data point.

The normalized weight so that the total sum of all weights $W_n(xi)$ is equal to N is as follows[27].:

$$W_n\ (x_i) = \frac{N}{\sum_{i=1}^{N} W(x_i)} W(x_i) \qquad (13)$$

Finally, the weights objective function is defined as below [27]:

$$\min \quad \|B\| + \sum_{i=1}^{N} W_n\ (x_i) \qquad (14)$$

3) *Fuzzy Classification*

a fuzzy rule-based classification is used for data classification; for each test point yi placed in the margin: a Gaussian fuzzy membership function A_{ik} is defined as below [27]:

$$A_{ik} = \prod_{j=1}^{P} exp^{\frac{-(y_{ij} - \mu_{jk})^2}{2\sigma_{jK}^2}} \quad \forall k \in 1,2, \qquad (15)$$

Then a "membership probability" is defined for each marginal point to measure its classification certainty in $class\ c_1\ or\ class\ c_2$ [27].

$$P_{i,}c_1 = \frac{A_{i,}c_1}{A_{i,}c_1 + A_{i,}c_2} \quad and\ P_{i,}c_2 = 1 - P_{i,}c_1 \qquad (16)$$

III. EXPERIMENTS:

To keep the most important information we have applied BPSO feature selector after a reduction space algorithm such as PCA, LDA, KFA or KPCA. Our experiments were performed on ORL database. This latter contains grayscale face images which are normalized to 128 x 128 pixels. This database contains 400 images of 40 different persons (10 samples per class). We have subdivided it into 2 parts, one for training and the other for assay. Also we have employed the implemented reduction space algorithms of the Matlab Phd_tool developed by Vitomir Struc [15].

The following table shows the recognition rates of BPSO and the considered combinations PCA then BPSO, KPCA then BPSO, LDA-BPSO and KFA-BPSO when matching based on Euclidean distance is considered and LDA-FSVM & PCA-FSVM when Fuzzy SVM classifier is considered.

TABLE 1
THE RECOGNITION RATE AFTER 3 RUNS FOR EACH ALGORITHM

200 faces for train and other 200 faces were used for test; in all these experiments the Gabor phase features were employed.	
Algorithm	**Recognition rate**
BPSO	84,333 ± 1,666
PCA - BPSO	77,833 ± 3,166
KPCA-BPSO	79,5 ± 1,5
KFA - BPSO	79,666 ± 1,333
LDA - BPSO	84,5 ± 1
LDA + FSVM	96.50
PCA + FSVM	96.75

The obtained results when matching based on Euclidian distance is used show that the recognition rate is almost the same for PCA-BPSO, KPCA-BPSO and KFA-BPSO algorithms and shows the superiority of BPSO and LDA-BPSO algorithms. Specifically, the LDA-BPSO algorithm that gives the highest recognition rate. The recognition rate of BPSO without the use of reduction space algorithm is also a competing approach. We also observed that KFA is able to outperform KPCA in terms of recognition rate.

The obtained results when the Fuzzy SVM classifier is used are very promising especially in combination with PCA reduction space algorithm; this proves the importance of the classification step in face recognition. Also, it should be noted that recognition rate can be higher when testing faces are randomly chosen as published in some works.

IV. CONCLUSION

In this work a hybrid approach to features selection is presented. This method is based on binary PSO algorithm and the reduction space algorithms: PCA, KPCA, LDA and KFA. The used objective function is the sum of squared errors between the selected features and the training data. This method has been tested on ORL faces using the phase of Gabor filters as features descriptor and classification is done by matching based on the minimum Euclidian distance or Fuzzy support vector machines. Results show the robustness of the proposed approach and prove the superiority of BPSO-LDA as features selector. They also prove the effectiveness of FSVM classifier in face recognition. The work presented in this paper is devoted to improve recognition of human faces; it would be interesting to combine outputs of different classifiers in order to get improved recognition rate.

REFERENCES

[1] Y.Sumi, Y.Ohta, 1996 .Human Face Analysis Based on Distributed Two-Dimensional Appearance Models, *Systems and Computers in Japan*, 27, 7, 97-108.
[2] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multimodal 2d+3d face biometrics. *IEEETransactions on Pattern Analysis and Machine Intelligence*, 27(4) :619{624, 2005.
[3] M. J. Jones. Face recognition : Where we are and where we go from here. IEEJ Transactions on Electronic, Information and Systems, 129 :770–777, 2009.
[4] V.Pesce Delfino, M.Colonna, E.Potente, E.Vacca, F. Jr Introna , Computer aided skull-face superimposition. *Am J Forensic Med Pathol*, 7(3): 201-12 (1986).
[5] P.S. Penev and J.J. Atick. Local feature analysis : A general statistical theory for object representation. Network : Computation in Neural Systems, 7(3) :477_500, 1996.
[6] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, pages 84_91, Jun 1994.
[7] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers. "Fisher discriminant analysis with kernels". Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, pages 41_48, Aug 1999.
[8] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K. Müller." Invariant feature extraction and classification in kernel spaces", 2000.
[9] T Ahonen, A. Hadid, and M. Pietikainen. "Face recognition with local binary patterns". In: Proceedings of the European Conference on Computer Vision, Prague, Czech, pp. 469–481 (2004)
[10] T. Ojala, M. Pietikäinen, T. Mäenpää. "Multiresolution grayscale and rotation invariant texture classification with local binary patterns". IEEE Trans. Pattern Anal. Mach. Intell., 24(7) :971_987, 2002.
[11] D. H. Hubel, T. N. Wiesel, "Ferrier lecture. Functional architecture of macaque monkey visual cortex" , Proceedings of the Royal Society of London, Series B 198:1-59. (1977)

[12] B. A. Olshausen, D.J. Field," Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature* 381, 607 - 609 (1996).

[13] B. Schiele, J.L. Crowley. "Recognition without correspondence using multidimensional receptive field histograms". Int. J. Comput. Vision, 36 :31-50, 2000.

[14] Ngoc-Son VU, Contributions à la reconnaissance de visages à partir d'une seule image et dans un contexte non-contrôlé, l'École Doctorale Électronique, Électrotechnique, Automatique, Traitement du Signal, GRENOBLE, 19 novembre 2010.

[15] V .Struc, B. Vesnicer, N. Pavesic, "The Phase-Based Gabor Fisher Classifier and its Application to Face Recognition Under Varying Illumination Conditions", In Proceedings of the international IEEE conference on Signal Processing and Communication Systems (ICSPCS'08) (December 2008), pp. 1-6.

[16] H.Chouaib, S.Tabbone, O.Ramos Terrades, F.Cloppet, N.Vincent, Sélection de caractéristiques à partir d'un algorithme génétique et d'une combinaison de classifieurs Adaboost , in "Colloque International Francophone sur l'Ecrit et le Document - CIFED 08. 2008.

[17] M. Ahmadieh Khanesar, M.Teshnehlab and M.A.Shoorehdeli., "A Novel Binary Particle Swarm Optimization" (2007). Proceedings of the 15th Mediterranean Conference on Control and Automation.. MED '07. /2007.

[18] Y. Mao, X .Zhou, D. Pi, Y. Sun, ST. Wong. "Multiclass Cancer Classification by Using Fuzzy Support Vector Machine and Binary Decision Tree with Gene Selection". Journal of Biomedicine and Biotechnology 2005(2): 160-171.

[19] Chun-fu Lin, Sheng-de Wang ,"Training algorithms for fuzzy support vector machines with noisy data",In proceeding of: Neural Networks for Signal Processing NNSP'03.2003, Pattern Recognition Letters 25 (2004) 1647–1656.

[20] Chun-fu Lin, Sheng-de Wang, "Training algorithms for fuzzy support vector machines with noisy data", Pattern Recognition Letters 25 (2004) 1647–1656.

[21] R. M. Ramadan, R. F. Abdel – Kader, "Face Recognition Using Particle Swarm Optimization-Based Selected Features", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No. 2, June 2009.

[22] R.Jakhar, N. Kaur, R. Singh, "Face Recognition Using Bacteria Foraging Optimization-Based Selected Features", *(IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence.*

[23] K. Manikantan, V. Govindarajan, V ,S Sasi Kiran, S Ramachandran, 3Face Recognition using Block-Based DCT Feature Extraction", Journal of Advanced Computer Science and Technology, 1 (4) (2012) 266-283.

[24] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007) 249-268. 2007.

[25] A.Borji, et M. Hamidi. Support Vector Machines for Persian Font recognition. In: International Conference on Computer, Electrical, Systems Science, and Engineering (CESSE 2007), Prague, Czech Republic July 27-29, 2007.

[26] F. Lauer, C. Y. Suen and G. Bloch, "A trainable feature extractor for handwritten digit recognition," Pattern Recognition, vol. 40, 1816-1824, 2007.

[27] M. Parandehgheibi, Probabilistic Classification using Fuzzy Support Vector Machines, Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011).

