Optimal Feature Selection based on Image Pre-processing using Accelerated Binary Particle Swarm Optimization for Enhanced Face Recognition

Aneesh M U, Abhishek A K Masand, K Manikantan

Abstract

Feature Selection is an optimization problem in any Face Recognition technology. This paper proposes a novel method of Binary Particle Swarm Optimization called Accelerated Binary Particle Swarm Optimization (ABPSO) by intelligent acceleration of particles. Together with Image Pre-processing techniques such as Resolution Conversion, Histogram Equalization and Edge Detection, ABPSO is used for feature selection to obtain significantly reduced feature subset and improved recognition rate. The performance of ABPSO is established by computing the recognition rate and the number of selected features on ORL database and Cropped Yale B database.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of ICCTSD 2011.

Keywords: Face Recognition; Image Pre-processing; Feature Selection; Feature Extraction; DCT; ABPSO.

1. Introduction

Face Recognition (FR) has evolved drastically over the last decade and has found innumerable applications in various fields. Major advancements in the recent past have propelled FR technology into the spotlight. FR is used for both verification and identification (open-set and closed-set). W. Zhao et al., [1] provides an excellent survey of various FR techniques. A detailed survey of illumination invariant FR techniques can be found in [2]. [3] provides a survey on 2D FR techniques. An FR system basically involves Feature Extraction and Feature Selection. A feature extractor is known to be efficient for an FR system when it selects as large as possible the set of discriminate features which are independent of changes in facial expression and differences in scale. Also, variations in pose, lighting and illuminating conditions should not hamper its efficiency.

*K Manikantan. Tel.: +0-944-978-7043; fax: +0-802-358-7731.
E-mail address: kmanikantan@msrit.edu
For the application of an algorithm, the process employed to select the feature subset from the database is commonly known as feature selection. In this paper we propose Accelerated Binary Particle Swarm Optimization (ABPSO) algorithm based on an intelligently updated velocity equation. We apply ABPSO for feature selection and establish its improved performance over the basic Binary PSO algorithm. The set of selected features are found to be significantly reduced. This causes a reduction in the memory space required for storing face features in the face feature gallery of the proposed FR system (ref: Fig. 3). The experiments are conducted for two databases: Cambridge ORL and Cropped Yale B. From the experimental results it can be verified that by employing both, the pre-processing techniques and ABPSO, the recognition rate is significantly improved along with largely reduced set of selected features.

2. Image Pre-processing, Feature Extraction using DCT and Binary PSO.

2.1. Image Pre-processing

Among the various pre-processing techniques, the three techniques (ref: Fig. 5) of relevance are Bi-Cubic interpolation, Histogram equalization and Edge detection using Laplacian of Gaussian (LoG). Image interpolation provides a technique of producing high-resolution image from its low-resolution counterpart [4]. Interpolation basically, is the process of estimating intermediate values of a continuous event from discrete samples [5]. It is a type of approximating function whose value must coincide with the sample data at the interpolation nodes or sample points. Bi-Cubic interpolation is a resolution conversion method preserving finer details of images with increased sharpness, better than bilinear algorithm. Whenever an image is resampled, there will be a loss of focus within the image, but bi-cubic interpolation, among various methods, provides maximum sharpness. Histogram equalization is a non-linear process aimed to highlight brightness in a way particularly suited to human visual analysis [6]. This tries to transform the distribution of pixel intensity values in the image into a uniform distribution and consequently improves the image’s global contrast. In order to preserve local features, despite the influence of lighting, useful for recognition, LoG [7] is applied after histogram equalization. LoG is applied for Cropped Yale B database to obtain efficient edge extraction and detection of different resolution edges.

2.2. Feature Extraction using Discrete Cosine Transform (DCT)

Ahmed et al.,[8] state that ‘DCT is a real transform that has great advantages in energy compaction’. Higher recognition rates with lower computational complexity are achieved with DCT. Also, [9] describes the use of DCT for feature extraction in FR.

The DCT also helps to separate the images into spectral sub-bands of differing importance, with respect to image’s visual quality and transforms a signal or an image from spatial domain to frequency domain. For an \( N \times M \) image \( f(x,y) \), the general equation for DCT is:
\[ F(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \cos \left( \frac{\pi u}{2N} (2x+1) \right) \cos \left( \frac{\pi v}{2M} (2y+1) \right) f(x,y) \]  

where \( f(x,y) \) is intensity of the pixel in row \( x \) and column \( y \);
\[
\alpha(u), \alpha(v) = \sqrt{\frac{1}{N}} \quad \text{for } u, v = 0 \text{ else } \sqrt{\frac{2}{N}} 
\]

The signal energy for most images lies at low frequencies which relate to large DCT coefficient magnitudes; these are located at the upper-left corner of the DCT array as shown in the 3-D log plot of Fig. 1 (Eg: For 100% resolution, 56x56 features are considered for a 92x112 DCT coefficient matrix). Inversely, the higher frequencies are present at the remaining portions of the DCT array and can easily be removed with little distortion in visibility as and when \( u \) and \( v \) move toward the sub-image width and height respectively. High frequency information by itself is insufficient for good face recognition performance [10].

2.3. Binary Particle Swarm Optimization (Binary PSO)

PSO is an intelligent optimization algorithm inspired by bird flocking and fish-schooling [11]. In the basic PSO, potential solutions called particles fly through the problem space by following the current optimum solutions. Binary PSO was formulated to have a discrete version of PSO and the algorithm has been developed in [12]. The only difference compared to the continuous PSO is that each new positional value is set to 0 or 1 by applying a sigmoid transformation and a probabilistic rule. The Binary PSO requires the position to be coded as a string of 1’s and 0’s. The particle velocity function is used as the probability function for the position update. A potential solution is represented as a particle having positional coordinates \( X_i = [x_{i1}, x_{i2}, \ldots, x_{iD}] \) in a \( D \)-dimensional space where the subscript \( i \) denotes the particle number and superscript \( t \) represents the iteration number. Each \( i \)th particle maintains a record of the position of its previous best performance in a personal best position vector \( P_{best,i} \). An iteration comprises evaluation of each particle, then stochastic adjustment of its velocity \( V_i = [v_{i1}, v_{i2}, \ldots, v_{iD}] \) in the direction of its own previous best and the best previous position of any particle in the neighbourhood. The best position of any individual in the whole swarm is stored in global best position \( G_{best} \). PSO is described by the following velocity and position update equations:

\[
V_{i}^{t+1} = w \cdot V_{i}^{t} + c_1 \cdot \text{rand} \cdot (P_{best,i} - X_{i}^{t}) + c_2 \cdot \text{rand} \cdot (G_{best} - X_{i}^{t}) \tag{3}
\]

where \( w = \) inertia weight, \( c_1 = \) cognitive parameter, \( c_2 = \) social parameter.

\[
X_{i}^{t+1} = X_{i}^{t} + V_{i}^{t+1} \quad \text{for } i = 1 \text{ to } N ; \quad N = \text{number of particles} \tag{4}
\]

If \( r \) is a random number between 0 and 1, the equation that updates the particle position becomes:

\[
X_{i}^{t+1} = 1 \text{ if } r < \frac{1}{1+e^{-V_{i}^{t+1}}} , \quad \text{else } X_{i}^{t+1} = 0 \tag{5}
\]
The Binary PSO algorithm is:
Step 1) Initialize \( w, c_1, c_2 \)
Step 2) Initialize particle positions \( X'_i \) and velocities \( V'_i \)
Step 3) Do steps 4 - 8 for a fixed number of iterations until stopping criterion is not met.
Step 4) For particles from 1 to N do steps 5 - 8
Step 5) If fitness of \( X'_i \) > fitness of \( Pbest_i \) then update personal best positions \( Pbest_i = X'_i \)
Step 6) If fitness of \( X'_i \) > fitness of \( Gbest \) then update the global best position \( Gbest = X'_i \)
Step 7) Update Velocity Vector using equation (3)
Step 8) Update Position Vector using equation (4) and (5)
Step 9) Display \( Gbest \)

3. Accelerated Binary Particle Swarm Optimization and Feature Selection

This paper proposes an improved version of Binary PSO, called Accelerated Binary Particle Swarm Optimization (ABPSO), based on the concept given in [13], to perform well for multimodal problems. The parameter initializations and the velocity and position update equations in ABPSO are:

\[
V'^{i+1} = w \cdot V'_i + c_1 \cdot rand \cdot (Pbest_i - X'_i) + c_2 \cdot rand \cdot (Gbest - X'_i) \tag{6}
\]

\[
u_i = X'_i + V'^{i+1} \tag{7}
\]

\[
u_i = \frac{1}{1 + e^{-\nu_i}}
\]

\[
X'^{i+1} = 1 \text{ if } r < u_i, \text{ else } X'^{i+1} = 0 \tag{8}
\]

ABPSO is based on the modified velocity update Eq. (7). The velocity is updated for every iteration by summing it with the previous positional values for each particle. This velocity is modified by a sigmoidal function, and transformed into a probability which is used to compute the new position of every particle.

The velocity sigmoidal functions for BPSO and ABPSO when applied for images from ORL database are shown in Fig. 2. For simplicity, only 4 dimensions out of 10304 (=92x112) are used for plotting. From the plots, it is clear that in the case of ABPSO, immediately after about 30 iterations, the velocity sigmoidal function converges to either 1 or 0, whereas, in BPSO, convergence to 1 or 0 even after 100 iterations is not certain. Hence the proposed BPSO is termed Accelerated BPSO.
The block diagram of the proposed FR system is shown in Fig. 3. The two stages involved are: Training stage and Recognition stage.

3.1. Scatter Index as Fitness Function

After pre-processing, once DCT is extracted, the objective of the ABPSO algorithm is to search for the most representative feature subset. The scatter index is estimated in each generation where every particle is evaluated and a fitness value is returned to the scatter index function $S$. Let $z_1, z_2, \ldots, z_L$ and $N_1, N_2, \ldots, N_L$ represent the various classes and number of images within each class respectively, and $L$ denotes the total number of classes. If $\Psi_1, \Psi_2, \ldots, \Psi_L$ be the mean values of the corresponding classes and $\Gamma^{(i)}$ represents sample images from class $z_i$, then $\Psi_i$ can be calculated as:
\[ \Psi_j = \frac{1}{N_i} \sum_{j=1}^{N_i} \Gamma_n^{(j)}, \quad i = 1, 2, \ldots, L \text{ and } j = 1, 2, \ldots, N_i \]  

(9)

Also the grand mean \( \Psi_o = \frac{1}{N} \sum_{i=1}^{L} N_i \Psi_i \)  

(10)

where \( N \) is the total number of images for all the classes. If \( (\Psi_i - \Psi_o)^T \) represents the transpose of \( (\Psi_i - \Psi_o) \), then the between class scatter index function \( S \) is computed as:

\[ S = \sqrt{\sum_{i=1}^{L} (\Psi_i - \Psi_o)(\Psi_i - \Psi_o)^T} \]  

(11)

Euclidean distance is employed to measure the similarity between the test and the reference vectors in the image gallery. In an \( N \)-dimensional space, if \( p_i \) and \( q_i \) are the coordinates of \( p \) and \( q \) in the \( i^{th} \) dimension then, the Euclidean distance between any two points \( p \) and \( q \) is given as:

\[ D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_i - q_i)^2 + \ldots + (p_N - q_N)^2} = \sqrt{\sum_{i=1}^{N} (p_i - q_i)^2} \]  

(12)

In the proposed FR system (ref: Fig 3), distances in the feature space from a query image to every other image in the database are calculated using equation (12).

4. Experimental Results

To measure the performance of the FR system employing feature selection based on ABPSO, two parameters called Recognition Rate and Number of selected features are defined. Recognition rate is defined to reflect the quality of the algorithm.

\[ \text{Recognition rate(\%)} = \frac{\text{Number of test images matched}}{\text{Total number of test images}} \times 100 \% \]  

(13)

Number of selected features is the number of DCT coefficients obtained after applying feature selection algorithm. The ABPSO parameters are: Swarm size =30, \( c_1 = 2 \), \( c_2 = 2 \), \( w = 1 \), Number of Iterations =100. The performance of ABPSO based FR system is evaluated on two databases: ORL and Cropped Yale B.

4.1. Experiment 1: ORL Database

The database comprises of 400 images corresponding to 40 distinct persons taken at the AT&T Laboratories [14]. The original size of each image is 92x112 pixels, with 256 gray levels per pixel. For each class and for every replication, the images used for training and testing are generated randomly. Hence for every replication, a distinct set of testing images and training images are used. For experimentation, 4 images per person are chosen randomly for the training set and the remaining 6 images are used for testing.

ABPSO based Feature Selection algorithm is tested with feature vectors based on various sizes of DCT coefficients. The 2-dimensional DCT is applied to the input image and only a subset of the DCT coefficients corresponding to the upper left corner of the DCT array is retained. Subset sizes of 50x50, 40x40, 30x30 and 20x20 of the original 92x112 DCT array are used as input to the subsequent feature
selection phase. Fig. 4 shows the recognition rates and number of selected features obtained for different feature vector dimensions.

Fig. 4: Plot of Recognition rates and Number of selected features with DCT subset size for ORL for 100% resolution

For each problem instance, 100 replications are conducted. The average recognition rate is measured along with the average number of selected features. The results obtained are compared with those obtained by [15] and this comparison indicates nearly 50% reduction in the number of selected features and an improved recognition rate for all varying sizes of the DCT array (ref: Fig. 4).

Further, experiment is carried out by varying the resolution of the training and test image set using Bi-Cubic interpolation method. A size little more than 1/4th of the upper left of the DCT coefficient matrix is considered so as to cover most of the low frequency components (Eg: For 100% resolution, 56x56 features are considered for a 92x112 DCT coefficient matrix). In this case also, 100 replications are conducted with different resolutions for the complete ORL face database. The recognition rate and the number of selected features are tabulated (ref: Table 1).

Table 1: Recognition results for ORL

| Resolution (%) | DCT subset size | Average Recognition rate (%) | Average Number of selected features |
|----------------|-----------------|-------------------------------|------------------------------------|
| 100            | 56x56           | 95.2208                       | 889                                |
| 90             | 51x51           | 95.0958                       | 647                                |
| 80             | 45x45           | 95.2666                       | 489                                |
| 70             | 40x40           | 95.1083                       | 453                                |
| 60             | 34x34           | 95.6833                       | 342                                |
| 50             | 28x28           | 95.3041                       | 230                                |

Table 2: Recognition results for Cropped Yale B

| Resolution (%) | DCT subset size | Average Recognition rate (%) | Average Number of selected features |
|----------------|-----------------|-------------------------------|------------------------------------|
| 100            | 96x96           | 94.5482                       | 2459                               |
| 90             | 87x87           | 95.5014                       | 2528                               |
| 80             | 77x77           | 96.4985                       | 1639                               |
| 70             | 68x68           | 97.0789                       | 1416                               |
| 60             | 58x58           | 97.6988                       | 1193                               |
| 50             | 48x48           | 97.1885                       | 769                                |
4.2. Experiment 2: Cropped Yale B database

This database comprises of faces that are manually aligned, cropped and resized to 168x192, with varying illumination conditions containing single light source images of 38 individuals [16]. A total of 2204 images (58 images per individual for 38 individuals) are used for experimentation. 22 images per person are chosen randomly for the training set and the remaining 36 images are used for testing. Fig. 5 shows results of histogram equalization and LoG edge detection applied for an image. In the original image, it is evident that the pixels are not balanced showing darker pixels. The quality of the image is improved and feature parts such as eyes, nose and mouth become prominent after histogram equalization. The histograms obtained before and after equalization are shown. Following this, the LoG is applied.

![Image pre-processing on an image from Cropped Yale B database](image)

Fig. 5: Image pre-processing on an image from Cropped Yale B database

A total of 50 replications are done for different resolutions for the Cropped Yale B database (Eg: For 100% resolution, 96x96 features are considered for a 168x192 DCT coefficient matrix). The recognition rate and the number of selected features are tabulated (ref: Table 2).

5. Conclusion

This paper proposes ABPSO based feature selection algorithm for FR. The algorithm is applied to feature vectors, extracted using DCT and resolution conversion through Bi-cubic interpolation. It is also utilized to search the feature space for the optimal feature subset. Experiments on ORL and Cropped Yale B databases show that the algorithm generates improved results depicting high recognition rate with minimalistic set of selected features. The results obtained, when compared with those obtained by [15], are found to yield higher recognition rate with nearly 50% reduction in the number of selected features.

6. Scope for Further Research

The use of other Fitness functions and other image pre-processing techniques like Block processing, Gaussian pyramid etc. needs to be investigated. Further, ABPSO needs to be applied on pose-variant databases (such as FERET), human detection dataset (INRIA person dataset), and PASCAL object recognition database.

7. References
[1] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, “Face Recognition” A Literature Survey, ACM Computing Surveys, vol. 35, p. 399-458, 2003.

[2] Xuan Zou, J. Kittler, K. Messer, “Illumination Invariant Face Recognition” : A Survey, First IEEE International Conference on Biometrics: Theory, Applications, and Systems, p. 1-8, September 2007.

[3] A.M. Patil, S.R. Kolhe, P.M. Patil, “2D Face Recognition Techniques” A Survey, International Journal of Machine Intelligence, vol. 2, p. 74-83, 2010.

[4] Zhou Dengwen, “An Edge-Directed Bicubic Interpolation Algorithm”, Third International Congress on Image and Signal Processing, vol. 3, p. 1186 – 1189, October 2010.

[5] Robert G. Keys, “Cubic Convolution Interpolation for Digital Image Processing”, IEEE Transactions on acoustics, speech, and signal processing, vol. assp-29, December 1981.

[6] Mark S. Nixon, Alberto S. Aguado, “Feature Extraction & Image Processing”, Second edition, Elsevier Ltd, 2008.

[7] John Canny, “A Computational Approach to Edge Detection”, IEEE Transactions on pattern analysis and machine intelligence, vol. PAMI-8, November 1986.

[8] N. Ahmed, T. Natarajan, K. R. Rao, “Discrete Cosine Transform”, IEEE Transactions on Computers, vol. C-23, p. 90–93, 1974.

[9] F. M. Matos, L. V. Batista, and J. Poel, “Face Recognition Using DCT Coefficients Selection”, Proceedings of the 2008 ACM Symposium on Applied Computing, p. 1753-1757, March 2008.

[10] Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, Richard Russel, “Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About”, Proceedings of the IEEE, vol. 94, November 2006.

[11] J. Kennedy, R. Eberhart, “Particle Swarm Optimization”, IEEE International Conference on Neural Networks, 1995.

[12] J. Kennedy and R. C. Eberhart, “A Discrete Binary Version of the Particle Swarm Algorithm”, Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, p. 4104-4108, October 1997.

[13] Maurice Clerc, “Binary Particle Swarm Optimisers: toolbox, derivations, and mathematical insight’s, version 1, January 2007.

[14] AT&T Laboratories, Cambridge, UK. “The ORL Database of Faces” (now AT&T “The Database of Faces”). Available Online: http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.zip.

[15] Rabab M. Ramadan, Rehab F. Abdel - Kader, Face Recognition Using Particle Swarm Optimization-Based Selected Features, International Journal of Signal Processing, Image processing and Pattern Recognition, vol. 2, June 2009.

[16] Athinodoros S. Georghiades, Peter N. Belhumeur, David J. Kriegman, From Few to Many :Illumination Cone Models for Face Recognition under Variable Lighting and Pose, IEEE Transactions on pattern analysis and machine intelligence, vol. 23, June 2001.