Particle Swarm Optimization- Best Feature Selection method for Face Images

Ms. P.V. Shinde
M.E. 2nd Year
Dept. Of Computer Engg.
AVCOE Sangamner
priti25.shinde@gmail.com

Prof. B.L. Gunjal
Assistant Professor
Dept. Of Computer Engg.
AVCOE Sangamner
hello_baisa@yahoo.com

ABSTRACT
Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. It is the most important step that affects the performance of a pattern recognition system. This paper presents a novel feature selection algorithm based on particle swarm optimization (PSO). PSO is a computational paradigm based on the idea of collaborative behavior inspired by the social behavior of bird flocking or fish schooling. The algorithm is applied to coefficients extracted by the discrete wavelet transform (DWT). The proposed PSO-based feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully selected according to a well defined discrimination criterion. Evolution is driven by a fitness function defined in terms of maximizing the class separation (scatter index). The classifier performance and the length of selected feature vector are considered for performance evaluation using the ORL face database. Experimental results show that the PSO-based feature selection algorithm was found to generate excellent recognition results with the minimal set of selected features.

General Terms
Pattern Recognition, Security, PSO Algorithm.

Keywords
Discrete Cosine Transform, Discrete Wavelet Transform, Face Recognition, Feature Selection, Genetic Algorithm, Particle Swarm Optimization.

.

1. INTRODUCTION
In this paper, a face recognition algorithm using a PSO-based feature selection approach is presented. The algorithm utilizes a novel approach effectively explore the solution space for the optimal feature subset. The selection algorithm is applied to feature vectors extracted using the DWT. The search heuristics in PSO is iteratively adjusted guided by a fitness function defined in terms of maximizing class separation. The proposed algorithm was found to generate excellent recognition results with less selected features.

The main contribution of this work is:

- Formulation of a new feature selection algorithm for face recognition based on the binary PSO algorithm. The algorithm is applied DWT feature vectors and is used to search for the optimal feature subset to increase recognition rate and class separation.

- Evaluation of the proposed algorithm using the ORL face database and comparing its performance with a PCA, ICA LDA feature selection algorithm and various FR algorithms found in the literature.

2. FEATURE EXTRACTION
The first step in any face recognition system is the extraction of the feature matrix. A typical feature extraction algorithm tends to build a computational model through some linear or nonlinear transform of the data so that the extracted feature is as representative as possible. In this paper DWT were used for feature extraction as explained in the following Sections.

3. FEATURE EXTRACTION METHODS

3.1 Discrete Wavelet Transform (DWT)
Feature extraction can be carried out by DCT i.e. Discrete Cosine Transform. But there are some disadvantages of DCT. It can't detect sharp edges, continuous spikes in face images. This can be overcome by DWT. In this paper FR using the DWT is based on the facial features extracted from a Haar Wavelet Transforms. The Haar wavelet transform is a widely used technique that has an established name as a simple and powerful technique for the multi-resolution decomposition of time series. Earlier studies concluded that information in low spatial frequency bands play a dominant role in face recognition. In 1986, Sergent shows that the low frequency band and high frequency band play different roles. The low frequency components contribute to the global description, while the high frequency components contribute to the finer details required in the identification task. Sergent has also demonstrated that as human face is a nonrigid object, it has abundant facial expressions; and expressions influence local spatial components of face. The Haar wavelet transform has been proven effective for image analysis and feature extraction. It represents a signal by localizing it in both time and frequency domains .Wavelets can be used to improve the image registration accuracy by considering both spatial and spectral information and by providing multiresolution representation to avoid loosing any global or local information. Additional advantages of using the wavelet-decomposed images include bringing data with

different spatial resolution to a common resolution using the low frequency subbands while providing access to edge features using the high frequency sub-bands.

As shown in Figure 1 at each level of the wavelet decomposition, four new images are created from the original *N x N*-pixel image. The size of these new images is reduced to ¼ of the original size, i.e., the new size is *N/2 x N/2.* The new images are named according to the filter (low-pass or highpass), which is applied to the original image in shorizontal and vertical directions. For example, the LH image is a result of applying the low-pass filter in horizontal direction and high-pass filter in vertical direction. Thus, the four images produced from each decomposition level are LL, LH, HL, and HH. The LL image is considered a reduced version of the original as it retains most details. The LH image contains horizontal edge features, while the HL contains vertical edge features. The HH contains high frequency information only and is typically noisy and is, therefore, not useful for the registration. In wavelet decomposition, only the LL image is used to produce the next level of decomposition

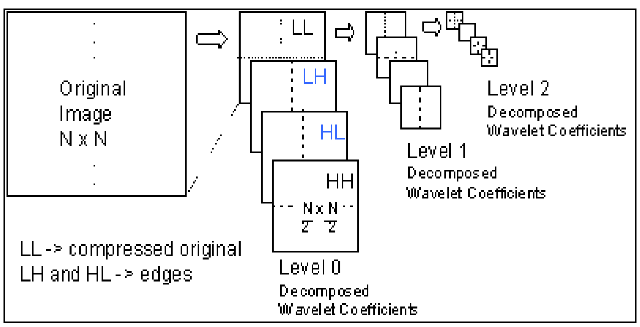
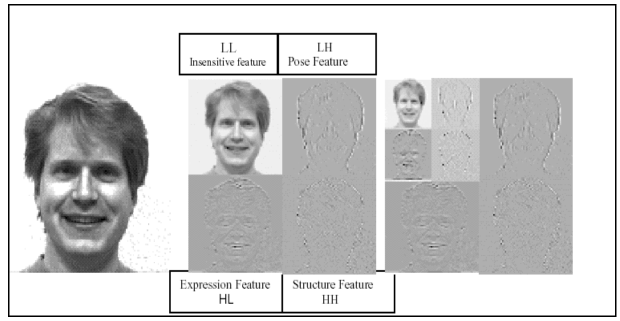


Figure 1: A 3 level Wavelet Decomposition of N x N image

Figure 2 shows the decomposition process by applying the 2D Wavelet Transform on a face image.

The original image (shown in Figure (2a)) is decomposed into four sub band images (shown in Figure (2b)) similarly, 2 levels of the Wavelet decomposition as shown Figure (2c) can be obtained by applying the wavelet transform on the low frequency band sequentially. In Figure (2b),the sub band LL corresponds to the low frequency components in both vertical and horizontal directions of the original image. Therefore, it is the low frequency sub band of the original image. The sub band LH corresponds to the low frequency component in the horizontal direction and high frequency components in vertical direction.Therefore it holds the vertical edge details. Similar interpretation is made on the sub bands HL and HH.



(a) (b) (c)

Figure 2: 2D Wavelet Decomposition of Face Image (a) Original image (b) 1- level wavelet decomposition (c) 2- level wavelet decomposition

As the change of facial expressions mainly varies in eyes, mouth and other face muscles, from the technical point of view, it involves mainly changes of edges. Let's take Figure (2b) as an example, the horizontal features of eyes and mouth are clearer than its vertical features, the sub band HL can therefore depict major facial expression features. The sub band LH, the vertical features of outline and nose are clearer than its horizontal features, depicts face pose features. The sub band HH is therefore the most important for rigid object recognition because it depicts the structure feature of the object. But human faces indeed are nonrigid objects, the sub band HH is the unstable band in all sub bands because it is easily disturbed by noises, expressions and poses. Therefore, if wavelet transform is applied to decompose face images, the sub band LL will be the most stable sub band. (a) (b) (c)

3.2 Particle Swarm Optimization (PSO)

Based on the idea of collaborative behavior and swarming in biological populations inspired by the social behavior of bird flocking or fish..Recently PSO has been applied as an effective optimizer in many domains such as training artificial neural networks, linear constrained function optimization, wireless network optimization, data clustering, and many other areas where GA can be applied . Computation in PSO is based on a population (swarm) of processing elements called particles in which each particle represent a candidate solution. The system is initialized with a population of random solutions and searches for optima by updating generations. The search process utilizes a combination of deterministic and probabilistic rules that depend on information sharing among their population members to enhance their search processes. Information sharing mechanism in PSO is considerably different. In GAs, chromosomes share information with each other, so the whole population moves like one group towards an optimal area. In PSO, the global best particle found among the swarm is the only information shared among particles. It is a one-way information sharing mechanism. Computation time in PSO is significantly less than in GAs because all the particles in PSO tend to converge to the best solution quickly.

3.2.1 PSO ALGORITHM

Initialize parameters

Initialize population
while (number of generations, or the stopping criterion is not met) {
for (i = 1 to number of particles N) {
if the fitness of t
i X is greater than the fitness of i_best p
then update i_best p = t
i X
if the fitness of t
i X is greater than that of gbest then
then update gbest = t
i X
Update velocity vector
Update particle position
Next particle
}
Next generation
}

Figure 3: PSO Algorithm

3.2.2. BINARY PSO AND FEATURE ALGORITHM

In the binary version, the particle position is coded as a binary string that imitates the chromosome in a genetic algorithm. This feature is not selected as a required feature for the next generation.

The feature selection is done by
1. Chromosome Representation
2. Fitness Function
3. Classifier

4. Comparison of PSO with PCA, LDA, and ICA

The PSO is implemented by using feature extraction technique called DWT. If PSO is compared with PCA, ICA and LDA the performance of PSO was found better than PCA, ICA, LDA. On different datasets these algorithms are applied and some results are obtained.

Following are software part implementation part that compares PSO with other techniques.

Running Algorithms: pso, pca, ica, lda

Using Datasets: att, ifd, yale

Separating train/test (60/40%) set images in ./datasets/att/_csu7, run 1...

Training algorithm "pso" on dataset "att"...

Testing algorithm "pso" on dataset "att"...

100.00% accuracy from "pso" on dataset "att" in 0.6 min

Training algorithm "pca" on dataset "att"...

Testing algorithm "pca" on dataset "att"...

91.88% accuracy from "pca" on dataset "att" in 0.0 min

Training algorithm "ica" on dataset "att"...

Testing algorithm "ica" on dataset "att"...

90.63% accuracy from "ica" on dataset "att" in 0.1 min

Training algorithm "lda" on dataset "att"...

Testing algorithm "lda" on dataset "att"...

95.63% accuracy from "lda" on dataset "att" in 0.0 min

Separating train/test (60/40%) set images in ./datasets/ifd/_csu7, run 1...

Training algorithm "pso" on dataset "ifd"...

Testing algorithm "pso" on dataset "ifd"...

100.00% accuracy from "pso" on dataset "ifd" in 1.1 min

Training algorithm "pca" on dataset "ifd"...

Testing algorithm "pca" on dataset "ifd"...

70.42% accuracy from "pca" on dataset "ifd" in 0.0 min

Training algorithm "ica" on dataset "ifd"...

Testing algorithm "ica" on dataset "ifd"...

68.33% accuracy from "ica" on dataset "ifd" in 0.1 min

Training algorithm "lda" on dataset "ifd"...

Testing algorithm "lda" on dataset "ifd"...

85.00% accuracy from "lda" on dataset "ifd" in 0.0 min

Separating train/test (60/40%) set images in ./datasets/yale/_csu7, run 1...

Training algorithm "pso" on dataset "yale"...

Testing algorithm "pso" on dataset "yale"...

100.00% accuracy from "pso" on dataset "yale" in 0.2 min

Training algorithm "pca" on dataset "yale"...

Testing algorithm "pca" on dataset "yale"...

78.33% accuracy from "pca" on dataset "yale" in 0.0 min

Training algorithm "ica" on dataset "yale"...

Testing algorithm "ica" on dataset "yale"...

83.33% accuracy from "ica" on dataset "yale" in 0.1 min

Training algorithm "lda" on dataset "yale"...

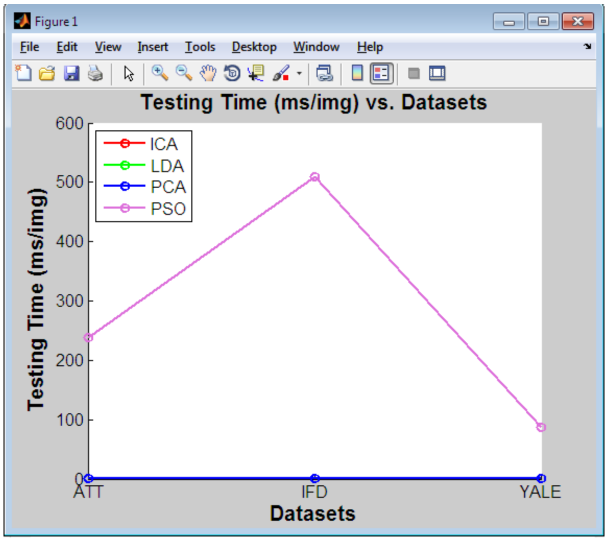
Testing algorithm "lda" on dataset "yale"...

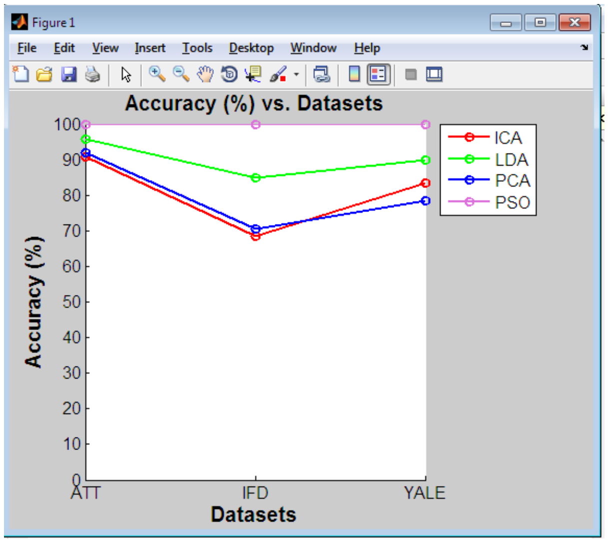
90.00% accuracy from "lda" on dataset "yale" in 0.0 min

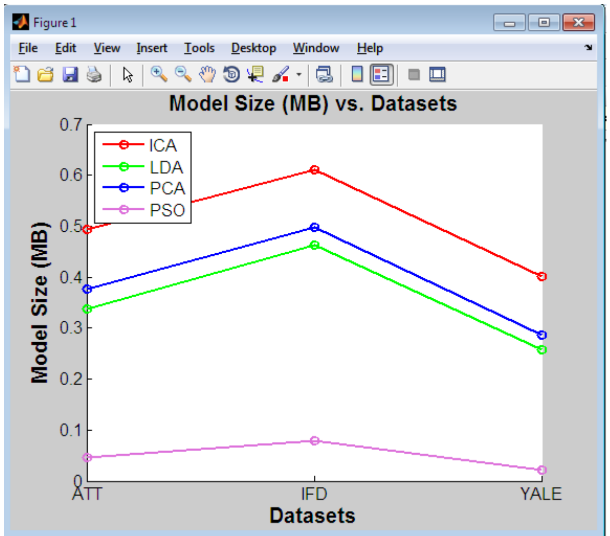
Algorithm runs done, generating report...

Report generated, check stats/results.csv for details

Face Recognition Evaluator done in 2.8 min







Hence from above snapshots we can say that PSO gives better performance as compared to other algorithms.

5. CONCLUSION

In this paper, a novel PSO-based feature selection algorithm for FR is proposed. The algorithm is applied to feature vectors extracted by the DWT. The algorithm is utilized to search the feature space for the optimal feature subset. Evolution is driven by a fitness function defined in terms of class separation. The classifier performance and the length of selected feature vector were consider for performance evaluation using the ORL face database.

6. REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[2]R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042- 052, 1993.

[3] C. Liu and H. Wechsler, "Evolutionary Pursuit and Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000.

[4] M. A. Turk and A. P. Pentland, "Face Recognition using Eigenfaces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.

[5] L. Du, Z. Jia, and L. Xue, "Human Face Recognition Based on Principal Component Analysis and Particle Swarm Optimization -BP Neural Network," *Proc 3rd Conference. on Natural Computation (ICNC 2007)*,vol. 3, pp. 287-291, August 2007.

[6] X. Yi-qiong, L. Bi-cheng and W. Bo, "Face Recognition by Fast Independent Component Analysis and Genetic Algorithm," *Proc. of the 4th International Conference on Computer and Information Technology (CIT'04)*, pp. 194-198, Sept. 2004.

[7] X. Fan and B. Verma, "Face recognition: a new feature selection and classification technique," *Proc. 7th Asia-Pacific Conference on Complex Systems*, December 2004.

.