

# FEATURE SELECTION USING MODIFIED PARTICLE SWARM OPTIMISATION FOR FACE RECOGNITION

Hemalatha Gayatri L<sup>1</sup>, Govindan V.K<sup>2</sup>

<sup>1</sup>PG scholar, Department of Computer science & engineering, National Institute of Technology Calicut, Kerala, India.

<sup>2</sup>Professor, Department of Computer science & engineering, National Institute of Technology Calicut, Kerala, India.

## Abstract

One of the major influential factors which affects the accuracy of classification rate is the selection of right features. Not all features have vital role in classification. Many of the features in the dataset may be redundant and irrelevant, which increase the computational cost and may reduce classification rate. In this paper, we used DCT(Discrete cosine transform) coefficients as features for face recognition application. The coefficients are optimally selected based on a modified PSO algorithm. In this, the choice of coefficients is done by incorporating the average of the mean normalized standard deviations of various classes and giving more weightage to the lower indexed DCT coefficients. The algorithm is tested on ORL database. A recognition rate of 97% is obtained. Average number of features selected is about 40 percent for a  $10 \times 10$  input. The modified PSO took about 50 iterations for convergence. These performance figures are found to be better than some of the work reported in literature.

**Keywords:** Particle swarm optimization, Discrete cosine transform, feature extraction, feature selection, face recognition, classification rate.

-----\*\*\*-----

## 1. INTRODUCTION

Humans have the ability to recognize faces easily and effortlessly but in the area of image analysis and computer vision it remained as a difficult problem on which many years of research is going on. A complete review of techniques for face recognition can be found in [1]. An unknown digital image is identified from a database of known images in face identification whereas in face recognition the person's identity is confirmed by the system from a stored database of faces. To verify or identify a person in the digital image, features extracted from the digital image are compared with features of the images in the facial database. Face recognition has applications in legacy systems such as voter registration, passports and driver's licenses, crowd surveillance, human-computer interaction, multimedia management, smart cards, access control and authentication. Occlusion, pose variation and illumination problems are still challenging for face recognition. Face recognition as a biometric has an advantage that it requires no co-operation of the person unlike other biometrics.

A problem called as curse of dimensionality exists because of large number of features. The search space is explored by the feature selection algorithm to find the optimal feature subset. Feature selection algorithms can be broadly classified into wrappers and filters [2], which are distinguished based on the choice of evaluation metric. A learning algorithm is used as a part of evaluation function in the search process of wrapper approaches whereas filters methods are independent of learning algorithm. Wrapper methods tend to provide better results and are computationally intensive than filter methods.

Optimal feature selection (FS) is to find the minimal set of features from original feature set that can represent the whole set. Redundant and irrelevant features which have insignificant effect on the output are removed. Redundant features are those which doesn't provide any extra information than the present and irrelevant features doesn't provide any useful information. Feature Selection has significant role in many applications like biometrics, machine learning, computer vision, and medical imaging. Exhaustive search can be done on all combinations possible with input features to obtain the optimal feature subset but it is computationally very expensive. With increase in the dimension of feature space, cost associated for search of optimal feature subset becomes exponentially high. PSO and GA are some of the evolutionary computation techniques which are applicable to the feature selection problems because of their global search ability.

### 1.1 Feature Extraction

In face recognition, the first step is feature extraction from the images. In this paper, feature extraction technique used is DCT. Information about feature extraction by DCT for face recognition can be found in [3], [4], [5]. DCT has strong energy compaction property [6]. DCT can represent effective features of input image with few coefficients in the upper left corner of the matrix. Image can be roughly constructed using these few DCT coefficients.

### 1.2 Particle Swarm Optimization

Swarm intelligence is a discipline that deals with a system that is composed of many individuals, that are homogeneous, self-organized and are coordinated using decentralized control. The behaviour of the system depends upon interactions among the individuals and with their environment.

PSO is inspired by collective behaviour of social models such as flock of birds and school of fishes, designed by Kennedy and Eberhart in 1995. Implementation of PSO is easier, it has few parameters, it is computationally inexpensive, and the convergence rate is quicker when compared with genetic algorithm and genetic programming [7]. PSO possesses the search ability for the best solutions for optimization problems because of its intelligent properties like adaptation and self-organizing [8]. Only primitive mathematical operators are used in PSO. PSO has an advantage because it is computationally inexpensive and has low memory requirements [9].

### 1.2.1 PSO Algorithm

In PSO, a group of particles explores the solution space to find the best solution. In particle swarm optimization, each solution is treated as an individual bird in the swarm, i.e., a particle in the search space. Position of  $i^{\text{th}}$  particle is denoted by  $X_i = (x_1, x_2, \dots, x_d)$  and velocity  $V_i$  be  $(v_1, v_2, \dots, v_d)$  in the  $D$  dimensional search space. Each particle uses its individual experience and its neighbouring particles experience to move in the search space, that is, every particle has knowledge about its best position it has found so far and the best position found by swarm. We denote,  $PB_i$  as the best position (pbest) of  $i^{\text{th}}$  particle, and  $gbest$  as the best position found by the swarm so far. Every particle is given random initial position and random initial velocity. Fitness function evaluates every particle in the search space. It determines the quality of the particle. Particles are attracted towards best particles in their search space which have best fitness values. The velocity and position of particles are updated according to the equation.

$$V_i^{t+1} = \omega * V_i^t + c1 * rand1 * (PB_i - X_i^t) + c2 * rand2 * (gbest - X_i^t) \quad (1)$$

where,  $i=1,2, \dots, N$ ,  $N$  represents the swarm size,  $rand1$  and  $rand2$  are two random numbers uniformly distributed in the range 0 and 1, and  $c1$  and  $c2$  are constants. Velocity is updated based on the inertia of previous velocity, experience of the particle itself and the experience of the neighboring particles. Particle position is updated by:

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad (2)$$

In every iteration, fitness of every particle is evaluated. Velocity is updated based on  $pbest$  and  $gbest$  values and parameters like  $c1$ ,  $c2$  and  $w$ . Position is updated based on velocity. This process is repeated until maximum iterations or error criterion is reached.

### 1.2.2 Fitness Function

Fitness is a measure that determines the quality or goodness of a particle. In each iteration every particle is evaluated for fitness value. The particles quality is evaluated by its ability to increase class separation. Position vector is represented by binary values 0's and 1's. Let  $c_1, c_2, \dots, c_L$  denote the classes and  $N_1, N_2, \dots, N_L$  denotes the number of images in different classes.  $L$  be the total number of classes.  $N$  is the

total number of images. Let,  $M_1, M_2, \dots, M_L$  be the mean of each class and  $M_0$  be the global mean, i.e., mean of all classes. They are calculated as below.

$$M_i = \frac{1}{N_i} \sum_{j=1}^i C_j^i \quad i = 1, 2, \dots, L \quad (3)$$

$$M_0 = \frac{1}{N} \sum_{i=1}^L N_i M_i \quad i = 1, 2, \dots, L \quad (4)$$

Fitness function is calculated as:

$$F = \sqrt{\sum_{i=1}^L (M_i - M_0)(M_i - M_0)^t} \quad (5)$$

Pseudo code of PSO algorithm

```

Initialise Parameters w, c1, c2;
For every particle i
Initialise Velocity Vector Vi;
Initialise Position Vector Xi;
end

Do until maximum number of iterations
For every particle i
f=fitness(Xi);
If fitness(PBi) < f
    PBi=Xi;
end
If f > fitness(gbest);
    gbest= Xi;
end
Update Vi;
Update Xi;
end
end

```

## 2. RELATED WORK

Some of the recent publications in this topic of research are briefly reviewed in the following.

Ramadan and Kader [10] has presented a feature selection algorithm using PSO for face recognition. PSO is applied on the coefficients (features) of Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) to select the optimum features of image. The optimum subset of features are selected by PSO based on the fitness function. ORL database is used for testing the performance. Face recognition is carried out based on Euclidean distance between features of unknown face and the features of faces in the database. The proposed algorithm has better performance when compared to GA-based feature selection algorithms in terms of number of features selected and classification accuracy rate.

Yun et al. [11] proposed feature selection methods using algorithms like Genetic algorithm and Particle Swarm Optimisation. An algorithm which makes use of relevance of features is also proposed. A measure known as minimum redundancy and maximum relevance (MRMR) is used to select features, using which redundancy of features is

minimised and relevance of features is maximised. This method, when used with genetic algorithm and PSO, provided quality feature set and achieved better accuracy rate.

Bing Xue et al. [12] has proposed a binary particle swarm optimization (BPSO) with two new fitness functions to minimise the number of features selected and to maximise the classification accuracy rate. First one linearly changes weights of fitness function and the other one uses two-stage fitness function. Experiment results have shown that the BPSO with the two new fitness functions provides better results compared to classification performance used as fitness function.

Unlerand Murat [13] has proposed a modified discrete PSO algorithm for feature selection. In this approach, the relevance and dependence of features to be included in the feature sub-set is dynamically decided by using an adaptive feature selection algorithm proposed by them. Experimental results show that it is computationally less expensive and provides improved recognition rate .

Yuanning Liu et al. [14] has proposed a modified multiswarm PSO (MSPSO). The particles are split into set of subsets called sub swarms. Sub swarms are controlled and monitored by a scheduler. They introduced a mechanism to decide the survival of sub swarm called as survival of the fittest. An improved feature selection method (IFS) has also been designed. The proposed algorithm gave better results when compared to standard PSO, GA and grid search in terms of classification accuracy rate.

Liam Cervante et al. [15] has developed two new feature selection approaches based BPSO. The relevance and redundancy in the feature subsets is measure by mutual information and entropy. The two new algorithms are applied on training data from the dataset and the feature subset is selected. The subsets' classification performance is evaluated by a learning algorithm. The resulting feature set is found to have less number of features and achieved same or higher classification rate.

Bing Xue et al.[16] has proposed a new approach for feature selection based on PSO. In this approach, an external archive is introduced which stores solutions which are later used to determine gbest of particles. Particles gbest is chosen from the solutions in archive using selection methods such as random selection and roulette wheel selection. The proposed approaches perform better than the standard PSO. They select less number of features and have same or higher classification rate.

From the review given above, it can be concluded that there is scope for further reduction in feature dimensionality using new rules for updating the position in PSO algorithm. The use of features selected by such a modified algorithm is expected to provide faster computation and improved recognition performances.

### 3. MODIFIED PSO ALGORITHM

A modified PSO algorithm has been proposed for feature selection. Feature selection becomes complex with the increase in number of features in dataset. Dataset contains all kinds of features that may be relevant or irrelevant and redundant. Redundant features decrease the quality of feature set. With increase in feature number in the dataset, computational complexity also increases. Optimal features set is the minimal set of features that provides the optimal performance. Optimal feature selection using PSO is proposed to improve the quality of the feature set by eliminating insignificant features which have very less or no contribution to outcome, thereby, reducing the number of features and decreasing computational cost and/or increasing accuracy of classification rate.

The update mechanism for the position of particle in binary PSO algorithm is modified to minimize the number of features and maximize the recognition rate. In the modified approach, the choice of coefficients is done by incorporating the average of the mean normalized standard deviations of various classes and giving more weightage to the lower indexed DCT coefficients.

Standard deviation for each class is computed as below.

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (6)$$

where n is the number of training images in each class and  $\mu$  denotes the mean value of DCT coefficients of training images in each class and  $x_i$  denotes DCT coefficients of each of the training images.

Average of the mean normalized standard deviations of various classes is calculated as:

$$S = \sum_{i=1}^L SD_i / \text{classmean}\{i\} \quad (7)$$

$$S = S / (\text{totalmean}) \quad (8)$$

$$S = \text{Absolute}(S) \quad (9)$$

$$\text{median} = \text{median}(S) \quad (10)$$

where  $i=1,2,\dots,L$ . L is the number of classes, total mean is the sum of all classmeans. Classmean is computed as in equation 3.

The position of the particle in Modified Binary PSO is updated by the following equation.

$$\text{and}(\text{rand}_3 < \left( \frac{1}{1+e^{-v_i^{t+1}}} + \frac{1}{1+g+h} \right), S < \text{median}) \quad (11)$$

where g,h are matrix indexes, then  $X_i^{t+1} = 1$  else  $X_i^{t+1} = 0$

The procedure is briefly described in the following:

1. Feature Extraction: DCT coefficients are used as features.. DCT coefficients on the left upper corner have significant information. So, only few coefficients are used in this approach. A  $10 \times 10$  matrix is given as input to BPSO algorithm.
2. Apply Binary PSO algorithm to the subsets of DCT coefficients obtained from the images in database.
3. Features of the input images which are more effective are selected. Features which have no significant contribution to classification accuracy are removed.
4. Euclidean distance is used for identifying the images . Face image corresponding to lowest Euclidean distance is considered as result in this approach.

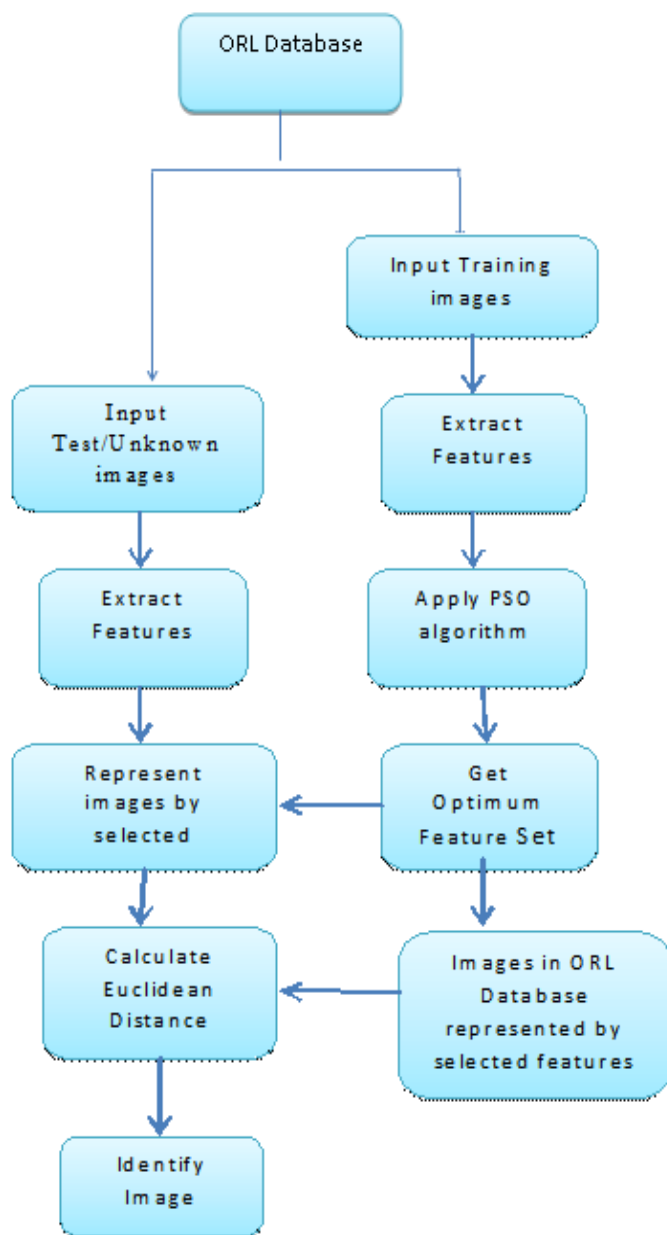


Fig-1: Face recognition using PSO

#### 4. EXPERIMENTAL RESULTS

The 2D DCT is applied on input images. DCT coefficients are used as features. Modified Binary PSO algorithm is applied to the  $10 \times 10$  matrix of DCT coefficients obtained from the images in database. Optimal Feature Set is obtained based on the fitness function after specified number of iterations.

Table-1: PSO Parameters

Swarm size	30
w	0.6
c1	2
c2	2

For performance analysis, ORL database is used which contains 400 images. It contains 10 images for 40 distinct persons. 4 out of 10 images for each person are used for training and the rest are used for testing. Each image resolution is  $92 \times 112$ , with 256 grey levels per pixel.

Table-2: comparison of modified algorithm with [10]

	Modified PSO algorithm	[10]
Number of training images	240	240
Number of test images	260	260
Recognition rate	97%	94.8%
Number of Iterations	50	100
Input Matrix size	$10 \times 10$	$50 \times 50$
Average features selected	40%	50%

#### 5. CONCLUSION

Optimal feature selection (FS) is to find the minimal subset of features from original feature set that can represent the whole set. PSO algorithm explores entire feature set and finds optimal feature subset. There are many important factors in PSO, such as the initialization strategy and the updating mechanisms which can be further analyzed to improve the performance and increase classification accuracy rate. In the modified algorithm, while updating the position vector, the average of the mean normalized standard deviations of various classes is considered and more weightage is given to the lower indexed DCT coefficients. The modified algorithm when tested on ORL database has classification accuracy rate of 97% and the average number of features selected are 40 for a  $10 \times 10$  input matrix.

## REFERENCES

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, 2003, pp. 399–458.
- [2] Isabelle Guyon, Andre Elisseeff, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research* 3 (2003) 1157-1182.
- [3] A. S. Samra, S. E. Gad Allah, R. M. Ibrahim, "Face Recognition Using Wavelet Transform, Fast Fourier Transform and Discrete Cosine Transform," *Proc. 46th IEEE International Midwest Symp. Circuits and Systems (MWSCAS'03)*, vol. 1, pp. 272-275, 2003.
- [4] C. Podilchuk and X. Zhang, "Face Recognition Using DCT-Based Feature Vectors," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, vol. 4, pp. 2144-2147, May 1996.
- [5] Z. Yankun and L. Chongqing, "Efficient Face Recognition Method based on DCT and LDA," *Journal of Systems Engineering and Electronics*, vol. 15, no. 2, pp. 211-216, 2004.
- [6] Ahmed, N.; Natarajan, T.; Rao, K. R. (January 1974), "Discrete Cosine Transform", *IEEE Transactions on Computers* C-23 (1): 90–93.
- [7] J. Kennedy and W. Spears, "Matching algorithms to problems: an experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator," in *IEEE Congress on Evolutionary Computation (CEC'98)*, 1998, pp. 78–83.
- [8] R. C. Eberhart and J. Kennedy, "A New Optimizer Using Particles Swarm Theory," *Proc. Roc. 6<sup>th</sup> International Symp. Micro Machine and Human Science*, pp. 39-43, Oct. 1995.
- [9] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proc. IEEE International Conference on Neural Networks*, pp. 1942-1948, 1995.
- [10] Rabab M. Ramadan And 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, No. 2, June 2009.
- [11] Chulmin Yun, Byonghwa Oh, Jihoon Yang And JonghoNang , "Feature Subset Selection Based On Bio-Inspired Algorithms," *Journal Of Information Science And Engineering* 27, 1667-1686 (2011).
- [12] Bing Xue, Mengjie Zhang, Will N. Browne, "New Fitness Functions in Binary Particle Swarm Optimisation for Feature Selection", *WCCI 2012 IEEE World Congress on Computational Intelligence* June, 10-15, 2012 - Brisbane, Australia.
- [13] A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *European Journal of Operational Research*, vol. 206, no. 3, pp. 528–539, 2010.
- [14] Bing Xue, Mengjie Zhang, Will N. Browne, "Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach", *IEEE TRANSACTIONS ON CYBERNETICS*, VOL. 43, NO. 6, DECEMBER 2013
- [15] Liam Cervante, Bing Xue, Mengjie Zhang, "Binary Particle Swarm Optimisation for Feature Selection: A Filter Based Approach", *WCCI 2012 IEEE World Congress on Computational Intelligence* June, 10-15, 2012 - Brisbane, Australia
- [16] Bing Xue, A. K. Qin, Mengjie Zhang, "An Archive Based Particle Swarm Optimisation for Feature Selection in Classification", *IEEE Congress on Evolutionary Computation (CEC)* July 6-11, 2014, Beijing, China.