FEATURE SELECTION FOR FACE RECOGNITION AND EMOTION RECOGNITION PROBLEMS USING GRAVITATIONAL SEARCH ALGORITHMS

Thesis submitted in partial fulfillment of the requirements for the degree of

MASTER of ELECTRICAL ENGINEERING

by

TAPABRATA CHAKRABORTI

Registration No. - 92679 of 2005-06

Roll No. – M4ELE13-01

Under the guidance of

Dr. AMITAVA CHATTERJEE

DEPARTMENT OF ELECTRICAL ENGINEERING

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

KOLKATA-700032, INDIA

2013

JADAVPUR UNIVERSITY
FACULTY OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING
KOLKATA-700032, INDIA

CERTIFICATE of RECOMMENDATION

I hereby recommend that the thesis prepared under my supervision and guidance by **Tapabrata Chakraborti** entitled "**Feature Selection for Face Recognition and Emotion Recognition Problems using Gravitational Search Algorithms**" be accepted in partial fulfillment of the requirements for award of the degree of "**Master of Electrical Engineering**" at **Jadavpur University**. The project, in my opinion, is worthy of acceptance.

_____

Dr. AMITAVA CHATTERJEE
(In-charge of Thesis)
Associate Professor
Department of Electrical Engineering
Jadavpur University
Kolkata- 700032, India

Forwarded by

_____

Professsor Nirmal Kumar Deb
Professor & Head of the Department
Department of Electrical Engineering
Jadavpur University
Kolkata- 700032, India

JADAVPUR UNIVERSITY
FACULTY OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING
KOLKATA-700032, INDIA

Certificate of Approval*

The foregoing thesis is hereby approved as a creditable study of an engineering subject carried out and presented in a satisfactory manner to warrant its acceptance as a pre- requisite to the degree for which it has been submitted. It is notified to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed and conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

Final Examination for Evaluation of the Thesis Board of Examiners

_____

_____

*Only in case the thesis is approved (Signature of Examiners)

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of his **Master of Electrical Engineering** studies.

All information in this document have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name : **Tapabrata Chakraborti**

Roll Number : M4ELE13-01

Thesis Title : Feature Selection for Face Recognition and Emotion Recognition Problems using Gravitational Search Algorithms

Signature & Date:

Acknowledgements

It has been a great privilege for me to have had the chance to pursue post graduate studies at the Department of Electrical engineering, Jadavpur University. I would like to convey my deepest gratitude to my supervisor, Dr. Amitava Chatterjee for his guidance and critical inputs which has helped me immensely throughout the duration of this work. I am especially thankful to him for introducing me to structured research work and mentoring me about research best practices and methodologies. I would also like to thank him for giving me the opportunity to work in this highly interesting and state of the art topic.

I am grateful to Professor Amit Konar and Dr. Anisha Halder for allowing me to use the Human Facial Emotion Recognition Database developed by their research group in the ETCE Department of Jadavpur University.

I would also like to take this opportunity to express my sincere thanks to Dr. Kaushik Das Sharma (Dept. of Applied Physics, University of Calcutta) for kindly making time to initiate me in the standards of technical writing which has helped me to write this thesis and also the research papers (currently under review) that have resulted from this work.

The works presented in Chapter 2 and Chapter 3 was supported by University Grants Commission (UGC) India under University with Potential for Excellence (UPE) - Phase II Scheme awarded to Jadavpur University, Kolkata, India.

I express my heartfelt thanks to all the faculty members of the Electrical Measurement and Instrumentation Section (Electrical Engineering Department, Jadavpur University) for

always helping me about even the smallest of doubts throughout my current studies. Professor Sugata Munshi, Professor Mita Datta, Professor Palash Kr Kundu, Professor Biswajit Bhattacharyya, Professor Gautam Sarkar and Professor Debangshu Dey have always kindly extended a helping hand to me whenever I needed guidance and motivation.

I would like to express my sincere gratitude to Professor Nirmal Kumar Deb (Head of the Department, Electrical Engineering Department, Jadavpur University) for allowing me to access freely and use the rich infrastructure of the Department. Also I thank the Laboratory-in-charge and all members of the Electrical Measurement and Instrumentation Laboratory (Electrical Engineering Department, Jadavpur University) for allowing me and guiding me with the use of the equipments in the laboratory.

I am also grateful to my friends as the close friendship that has developed with my batch mates created an exciting environment which made my current studies a very enjoyable and memorable one. Lastly I gratefully acknowledge the constant support of my parents who have been there for me always.

Date: _____

Place: Jadavpur University, Kolkata (TAPABRATA CHAKRABORTI)

CONTENTS

CHAPTER 1

INTRODUCTION

In recent years automated identification of human facial traits has emerged as a popular area of research especially for the applications in face recognition and emotion recognition.

Face recognition [1], [2] has gained popularity due to its wide applicability in biometric traits based secure access systems because it is more robust than traditional protection passwords and also data acquisition is much simpler than some other biometric authentication methodologies. Face recognition is also relevant in forensic science where multi class classification of different subjects from facial images is highly relevant.

Emotion recognition [3], [4] on the other hand finds most of its use in HCI systems where appropriate actions are initiated by the computer depending on the facial expression of the subject and hence a robust classification of basic emotions is of primary importance.

Since both face recognition and emotion recognition are necessarily classification problems, supervised learning based systems are almost universally used. Hence choices and modifications in the feature extraction, feature selection and classification phases generate different novel techniques.

If mathematically sophisticated features with problem specific formulation are to be replaced by generalized feature extraction schemes with much simpler implementation and lower computation cost, a proper feature selection strategy becomes very pertinent in order to generate a subset of features with high discriminative characteristics. Some meta

heuristic non gradient optimization based feature selection schemes have been proposed in the existing literature. In this thesis a comparatively new method in this genre, the gravitational search algorithm [5] has been employed and several modifications have been proposed and applied for both the face recognition and emotion recognition problems. The proposed methodologies have been experimentally verified using several databases and over varying feature extraction schemes.

One traditional drawback of supervised learning based classification problems is that when the training size is not sufficiently large, a high recognition rate is difficult to achieve. In fact for those methods which use within class and between class distances, the method itself becomes invalid when the training set size per class is unity. This problem has also been addressed in this thesis in relation to face recognition.

The thesis is organized as follows.

- Chapter 2 presents a novel binary GSA, with adaptive weight, which has been employed for feature selection from the original feature set (extracted by using LBP, MCT, and LGP). The proposed method has been applied to face recognition.
- Chapter 3 proposes a local extrema based variation of GSA and addresses the problem of face recognition using one training image per person by using SVD.
- Chapter 4 presents a novel methodology for emotion recognition by feature selection, using the proposed binary quantum based modification of GSA with differential mutation, from an original feature set generated by 2-D DCT.
- Chapter 5 concludes the thesis along with a discussion for future scope of work.

References

[1] R. Jafri and H. R. Arabnia, "A Survey of Face Recognition Techniques," *Journal of Information Processing Systems*, vol. 5, no. 2, pp. 41-68, June 2009.

[2] 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, Dec. 2003.

[3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. G. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, Jan. 2001.

[4] M. Milanova and N. Sirakov, "Recognition of emotional states in natural human-computer interaction," *IEEE International Symposium on Signal Processing and Information Technology*, pp. 186-191, Dec. 2008.

[5] E. Rashedi, H. Nezamabadi-pour and S. Saryazdi, "GSA: A Gravitational Search Algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232-2248, June 2009.

CHAPTER 2

Face recognition with local patterns and adaptive weight based GSA

2.1 Introduction

Authentication (for access to secure physical or digital locations) and identification (in security, forensics, etc) using human biometric traits like fingerprint, iris data, vocal or facial features, have become very popular in recent times [1]. There are two major reasons behind this growing popularity of biometric recognition (authentication and/or identification) in both research and commercial use. The biometric recognition systems are much less susceptible to impersonation as the biometric traits of a human being are unique to that person. Furthermore, the problems of remembering or securely storing/transmitting the authentication code do not arise in case of biometric recognition.

In particular, there have been intensive research activities in face recognition as a viable means of achieving biometric recognition [2]. This is because the discriminating significance of facial features is more intuitive to understand and also facial images are more convenient to acquire. A facial image acquisition system can comprise only an ordinary digital camera where as a more sophisticated, and hence more costly, set up is required to acquire, e.g., an iris dataset.

Face recognition, whose roots belong to the highly related fields of image processing, machine learning and pattern recognition [3], can be broadly classified into two categories [4], depending on the philosophy of the method of feature extraction used. The first approach utilizes one or

more physical characteristics or geometrical features present in each face to discriminate between faces of different persons, like distance between the eyes, length of the nose, shape of eye and nose, contour of lips, etc. This approach may also seek to utilize differences in expression [5], illumination [6], or pose [7]. The second approach broadly utilizes, in different ways, the inherent statistical properties in the pixel intensities. Thus the second approach generally takes help of a transform which accepts the intensities of the pixels of the image as input and produces the discriminating features as output. The proposed method in this Chapter broadly falls into the second category. Several transformation methods have so far been utilized to solve such face recognition problems e.g. feature extraction using Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) [8]. Another possible transform based method for these application domains is the utilization of Log Polar Transform (LPT), a hybrid version of which has been employed in [9]. Moreover, a successful method of employing transformation can be that of multiplying or convolving the original image matrix with a transformation matrix or kernel as used in the Volterrafaces method in [10]. Some other popular approaches belonging to the second category that have so far been implemented with varied degree of accuracy include Eigenfaces [11], Fisherfaces [12], Laplacianfaces [13], Orthogonal Laplacianfaces [14], KRR [15], MLASSO [16], Spectral Regression [17], RDA [17], ORO [18], KLPPSI [19], , TANMM [20], Smooth LDA [21], CTA [22] etc. The work in [10] has presented, in a comprehensive manner, the relative successes (in terms of recognition rates) of many of these methods in performing face recognition, for several benchmark image databases, that have been extensively used for face recognition purposes over the years.

The feature extraction process in a supervised learning based method usually generates a large set of features leading to a considerable computation cost during classification. Feature

selection/reduction generates a reduced set from the original set of features and is thus hoped to decrease the computational cost of classification, the processing stage that follows the stages of feature extraction and feature selection. The reduced set of features may even have higher discriminatory power as a higher number of features do not necessarily result in a better recognition rate [23]. Classically Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [24] have been popular in performing such feature selection purposes. However, in recent times, it is also believed that the reduced feature set is expected to perform better especially when an optimization algorithm is used to select the suitable features by minimizing a meaningful cost function. In this context fruitful utilization of suitable non-gradient based stochastic optimization algorithms remains a challenging and encouraging domain. Some optimization techniques like particle swarm optimization (PSO) [8] and ant colony optimization [25] have already been utilized in solving these problems. It should be mentioned here that the features in the reduced set may be of two types: i) they may be generated as a function of the original feature set in which case the reduced feature set is not a subset of the original feature set [24] or ii) they may be selected from the original feature set in which case the reduced set is a subset of the original set [25], [8].

The present Chapter shows how a metaheuristic, stochastic optimization algorithm, called gravitational search algorithm (GSA), based suitable feature selection methodology can be developed for face recognition problems utilizing feature extraction methodologies based on contemporary, popular transform based face representation methods known as local binary patterns (LBP) [26], modified census transform (MCT) [27], and local gradient patterns (LGP) [28]. The LBP method was originally proposed mainly for the texture classification purpose and its strength lies in its performance invariance to global intensity variations. MCT was later

proposed as an enlarged variation of LBP where MCT utilizes nine bits to represent a pixel in an image, as opposed to eight bit based representation of a pixel in LBP. However, both LBP and MCT are sensitive to local intensity variations that become prominent along edge components and LGP was recently proposed to overcome these difficulties [28]. While LBP, MCT, and LGP based feature extraction methods are already justified to provide sufficient discriminating power, it is sincerely felt that hybridization of a suitable feature selection methodology with each of these powerful algorithms may further improve the face recognition or authentication performance, with reduced computational burden.

The Gravitational Search Algorithm (GSA), proposed in [29], is a comparatively recent but highly popular meta-heuristic optimization algorithm that is based on Newton's Laws of Gravity and Motion. In this Chapter the feature selection problem is configured as a binary decision making problem. Keeping that factor in mind a binary version of GSA is developed in this Chapter, inspired by similar earlier works carried out in binarization of particle swarm optimization (PSO) [30], [31], [32]. The present Chapter also proposes an improved version of the binary GSA that introduces a novel dynamic adaptation of weight features and this algorithm is named as Binary Adaptive Weight GSA (BAW-GSA) strategy. Then, in this Chapter, both binary version of traditional GSA and also the proposed BAW-GSA are separately utilized to select a reduced set of features from the features extracted separately utilizing LBP, MCT, and LGP. Finally backpropagation neural network (BPNN) employing Levenberg-Marquardt learning algorithm is employed for the classification purpose, utilizing each reduced set of features as the input. Each such hybrid combination of feature extraction-feature selection methodology is employed for three well known, benchmark face databases, namely, Yale A, Yale extended B and ORL face databases and these extensively detail performance comparisons

are tabulated and compared with several competing, contemporary methods popularly employed for similar face detection purposes. These performance evaluations conclusively justify the implementation of the proposed feature selection scheme i.e. BAW-GSA and prove its superiority in providing overall, enhanced face recognition performance.

The rest of the Chapter is organized as follows. In Section 2.2 an overview of the feature extraction schemes utilized in this Chapter is presented. In Section 2.3, a binary version of the traditional GSA algorithm has been introduced to make the method suitable for the desired feature selection. The proposed BAW-GSA algorithm is also introduced and described in detail in this section. A brief description of the classification strategy employed is given in Section 2.4. Section 2.5 contains the detailed experimental results and comparisons with competing methods. The Chapter is concluded in Section 2.6.

2.2 Local Pattern Based Feature Extraction Schemes

As mentioned before, the present Chapter considers three well known, contemporary methods of face representation, called LBP, MCT, and LGP, for the purpose of feature extraction. These methods are described in a nutshell now.

2.2.1 Local Binary Pattern (LBP)

LBP proposed by Ojala et al. [26] is an effective feature extraction tool which captures the local intensity variation trend of an image and has good discrimination characteristics.

Let i_c be the intensity of an image I at pixel (x_c, y_c) and i_n (n=0,1,…,7) be the intensity of a pixel in the 3×3 neighborhood of (x_c, y_c) excluding the center pixel i_c.

Then the LBP value for the pixel (x_c, y_c) is given by [28]

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c).2^n \quad (2.1)$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \tag{2.2}$$

For easy understanding, the computation of LBP is compactly provided in fig. 2.1 with the help of an illustrative example.

2.2.2 Modified Census Transform (MCT)

Modified Census Transform (MCT) proposed by Froba et al. [27] is a variation of LBP where the measure incorporates the mean intensity of the neighborhood. Moreover this is an extended version of the Census Transform taking all the nine bits of a 3x3 neighborhood including the center pixel. Thus the neighboring intensities in MCT, unlike LBP, vary over n=0,…,8. Let the mean intensity be denoted as $\bar{i}_c$. Hence it can be computed as:

$$\bar{i}_c = \frac{1}{9}\left(\sum_{n=0}^{8} i_n\right) \tag{2.3}$$

Then the MCT value for the pixel (x_c, y_c) is given by

$$MCT(x_c, y_c) = \sum_{n=0}^{8} s(i_n - \bar{i}_c).2^n \tag{2.4}$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \tag{2.5}$$

Again, for easy understanding, the computation of MCT is compactly provided in fig. 2.2 with the help of the same illustrative example, considering an identical 3×3 neighborhood.

2.2.3 Local Gradient Pattern (LGP)

LGP is an improvement over the traditional MCT and LBP, proposed by Jun et al. [28], which emphasizes the local variation in the neighborhood by incorporating the intensity gradient profile of the neighborhood in the measure.

Like in the case of LBP, let i_c be the intensity of an image I at pixel (x_c, y_c) and i_n (n=0,1,…, 7) be the intensity of a pixel in the 3×3 neighborhood of (x_c, y_c) excluding the center pixel i_c.

Then define the gradient of the neighboring pixel n as

$$g_n = |i_n - i_c| \tag{2.6}$$

Next, the average gradient $\bar{g}$ of the neighboring n pixels (n=0,1,…,7) is computed. This average gradient is given by

$$\bar{g} = (\tfrac{1}{8}) \sum_{n=0}^{7} g_n \tag{2.7}$$

Then the LGP value for the pixel (x_c, y_c) is given by

$$LGP(x_c, y_c) = \sum_{n=0}^{7} s(g_n - \bar{g}) 2^n \tag{2.8}$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \tag{2.9}$$

Figure 2.3 shows the result of LGP for the same illustrative example considered in fig. 2.1 and fig. 2.2.

2.3 Feature Selection Schemes employing modified GSA

The next stage involves selecting a subset of features from the original set. The present Chapter proposes to solve this problem using a metaheuristic optimization algorithm which minimizes a suitable cost function. The cost function used here is the ratio between the intra class Euclidean distance of features and the inter class Euclidean distance of features. As mentioned before, the gravitational search algorithm has been used in this Chapter to solve this optimization problem However, since GSA was originally introduced in the continuous domain, it has been suitably modified for our binary decision making problem of selecting or not selecting each extracted feature. For this purpose a binary version of the traditional GSA is proposed in this Chapter,

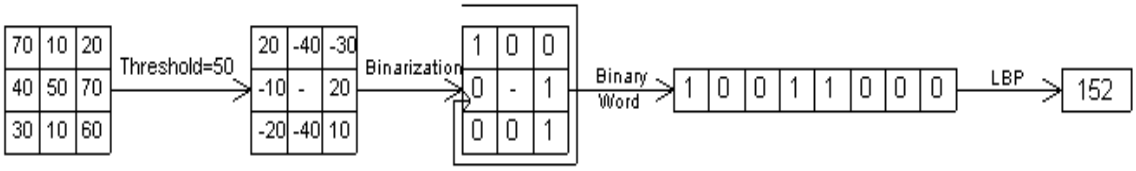

Fig. 2.1. Illustration of computation of LBP.

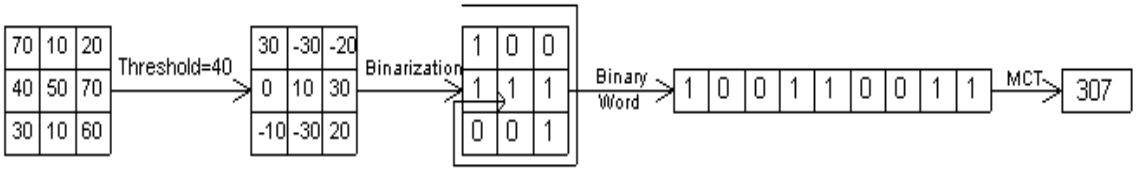

Fig. 2.2. Illustration of computation of MCT.

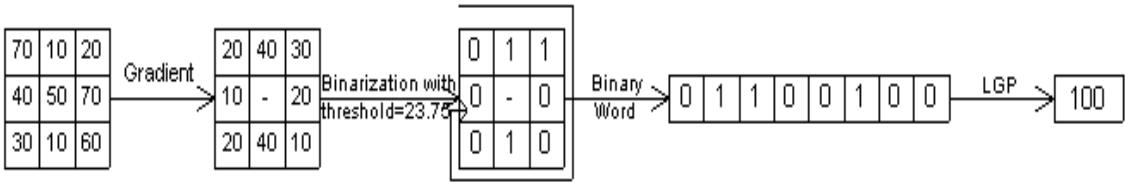

Fig. 2.3. Illustration of computation of LGP.

following similar existing works devoted to the development of binary PSO [8], [30], [31], [32]. In Section 2.3.3 the novel variant of GSA proposed in this Chapter, named as BAW-GSA, is described in detail. The proposed variation introduces a dynamic adaptive inertia weight in the traditional GSA algorithm to achieve a more intelligent iterative update strategy.

2.3.1 Gravitational Search Algorithm (GSA)

Consider an isolated universe of p particles obeying Newton's Laws of motion and gravity. Let the position of the ith particle (i=1, 2, ..., p) in n dimensional space at iteration t be given by

$$X_i(t) = \{x_i^1(t), x_i^2(t), \dots \dots \dots \dots, x_i^n(t)\} \tag{2.10}$$

The gravitational force experienced by a particle i due to the particle j at iteration t in the dimension d is given by the Newton's Law of Gravity as follows [29]:

$$F_{ij}^d(t) = G(t)\frac{M_i(t)\times M_j(t)}{R_{ij}(t)+\varepsilon}\left(x_j^d(t) - x_i^d(t)\right) \tag{2.11}$$

Where M_j is the mass of the particle j, M_i is the mass of the particle i, ε is a small positive constant ($\varepsilon > 0$), $R_{ij}(t)$ is the Euclidean distance between two agents i and j. Here it is assumed that, for each particle i, active gravitational mass (M_{ai}) = passive gravitational mass (M_{pi}) = inertial mass (M_{ii}) = M_i.

$$R_{ij}(t) = \left\|X_i(t), X_j(t)\right\| \tag{2.12}$$

$$G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta}, \beta < 1 \tag{2.13}$$

t_0 is the initial iteration. Hence $G(t)$ gradually decreases over time, as it is desired that the changes should be lesser and lesser as the system converges toward a solution [29, 38]. The total gravitational force on a particle i in the dth dimension is [29]:

$$F_i^d(t) = \sum_{j=1, j\neq i}^{p} rand_j F_{ij}^d(t) \tag{2.14}$$

where $rand_j$ is an uniformly distributed random number in the interval [0,1] and is responsible for the stochastic nature of the algorithm. Next, by Newton's Law of Motion, the acceleration of the agent i at time t in dimension d is calculated as:

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} \tag{2.15}$$

The velocity, position and mass update equations are [29]:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{2.16}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{2.17}$$

$$m_i(t+1) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{2.18}$$

$$M_i(t+1) = \frac{m_i(t+1)}{\sum_{j=1}^{p} m_j(t+1)} \tag{2.19}$$

where $fit_i(t)$ is the fitness value of the agent i at iteration t.

So the $best(t)$ and the $worst(t)$, the best and worst fitness among the p particles at iteration t respectively, are given as

$$best(t) = \min_{j \in \{1,2,\ldots p\}} fit_j(t) \tag{2.20}$$

$$worst(t) = \max_{j \in \{1,2,\ldots p\}} fit_j(t) \tag{2.21}$$

After the stopping criterion is met the position of the agent with the largest mass gives the solution of the search. The larger the mass of an agent, more slowly will it change its position and this is what is desired as it is nearer to the solution. Also it can be noted here that unlike the classical PSO, the basic GSA is a memory less algorithm since unlike PSO there is no need to keep record of the best position of each particle and also the globally best position.

The Algorithm 2.1 describes the implementation of the traditional GSA. Figure 2.4 represents the same algorithm in flow chart form.

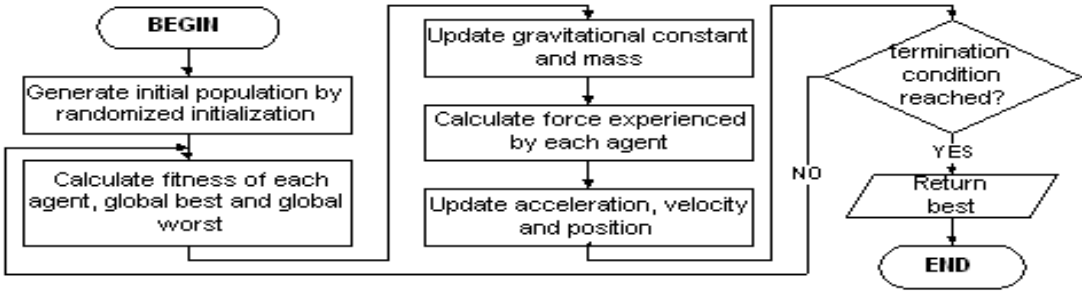

Fig. 2.4. Flow chart for the basic GSA.

Algorithm 2.1: Traditional GSA

BEGIN

Create p particles and make randomized initialization of their n dimensional positions X.

Initialize iteration number $t = 1$

REPEAT:

FOR $i = 1$ to p

Calculate fitness $fit_i(t)$

END FOR

Calculate global best fitness: $best(t) = \min_{i\in\{1,2,\dots p\}} fit_i(t)$

Calculate global worst fitness: $worst(t) = \max_{i\in\{1,2,\dots p\}} fit_i(t)$

FOR $i = 1$ to p

Calculate mass: $m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)}$, $M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{p} m_j(t)}$

Calculate gravitational constant: $G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta}, \beta < 1, t_0 = 1$

FOR $j = 1$ to p, $j \neq i$

Calculate distance between two particles: $R_{ij}(t) = \left\| X_i(t), X_j(t) \right\|$

Calculate force between two particles:

FOR $d = 1$ to n

$$F_{ij}^{d}(t) = G(t) \frac{M_i(t) \times M_j(t)}{R_{ij}(t) + \varepsilon} \left(x_j^d(t) - x_i^d(t) \right)$$

END FOR

END FOR

FOR $d = 1$ to n

Calculate total force on ith particle: $F_i^d(t) = \sum_{j=1, j\neq i}^{p} rand_j F_{ij}^d(t)$

Calculate acceleration: $a_i^d(t) = \frac{F_i^d(t)}{M_i(t)}$

Calculate velocity: $v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t)$

Calculate position: $x_i^d(t+1) = x_i^d(t) + v_i^d(t+1)$

END FOR

END FOR

UNTIL termination criterion is satisfied

END

2.3.2 Binary GSA (BGSA)

Assume that for each image, *n* features were extracted in the first stage and let the *n* dimensions of the position vector of each particle represent the selection variables corresponding to those features. Hence to take the decision of selecting or not selecting a feature based on the value of the corresponding selection variable, a proper binarization of the GSA algorithm is necessary.

So after calculating the velocity vectors $v_i(t+1)$ using equation (2.16), the position vectors are binarized as follows:

$$x_i^d(t+1) = \begin{cases} 1, & \frac{1}{1+e^{-v_i^d(t+1)}} > rand \\ 0, & otherwise \end{cases} \tag{2.22}$$

Here *rand* is an uniformly distributed random number in the interval [0, 1]. At the end of each iteration, the position vector for each particle *i* is carefully scanned for each of its dimension. Then, for particle *i*, a feature is chosen from the complete list of extracted features, if the corresponding dimension of its position vector possesses a value of 1 or rejected, if the corresponding value is 0. Hence the number of features selected for each particle *i* corresponds to the number of entries of 1 in the corresponding position vector.

Obviously at the start of the algorithm the position vectors of the particles need to be randomly initialized with values in {0, 1}.

Suppose at iteration *t,* for the particle *i*, out of the original *n* features, *m* features have been selected.

Fitness or goodness function of the particle *i* at iteration *t* is defined as the ratio of the intra class distance or within class distance ($Dintra_i^t$) to the interclass distance or between class distance ($Dinter_i^t$).

$$fit_i^t = \frac{Dintra_i^t}{Dinter_i^t} \tag{2.23}$$

where $Dintra_i^t$ is the sum of the Euclidean distances between the selected feature sets (for the *i*th particle at iteration *t*) of the images belonging to same class.

$$Dintra_i^t = \sum_c \sum_{u,r \in c, u \neq r} \left\| F_{u,i}^t - F_{r,i}^t \right\| \tag{2.24}$$

$F_{u,i}^t$ is the feature set of the *u*th image in the *c*th class corresponding to the position (selection) vector of the *i*th particle at iteration *t*. Similarly $Dinter_i^t$ is the sum of the Euclidean distances between the selected feature sets (for the *i*th particle at iteration *t*) of the images belonging to different classes.

$$Dinter_i^t = \sum_{c,c'} \sum_{u \in c, r \in c', c \neq c'} \left\| F_{u,i}^t - F_{r,i}^t \right\| \tag{2.25}$$

2.3.3 Binary Adaptive Weight GSA (BAW-GSA)

In this subsection Binary Adaptive Weight GSA is proposed.

From equation (2.16), the velocity update equation for particle *i* in dimension *d* is given as:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{2.26}$$

Now an adaptive dynamic weight w_i^t in the velocity update equation is introduced, which facilitates the convergence of the algorithm.

Thus the velocity update equation is modified as:

$$v_i^d(t+1) = w_i^t \times v_i^d(t) + a_i^d(t) \tag{2.27}$$

where

$$w_i^t = \left| \frac{\min\left(fit_i^t, \overline{fit^t}\right)}{\max\left(fit_i^t, \overline{fit^t}\right)} \right| \tag{2.28}$$

Here $\overline{fit^t}$ is the average fitness of all particles at iteration t such that

$$\overline{fit^t} = \frac{1}{p}\sum_{i=1}^{p} fit_i^t \tag{2.29}$$

The rationale behind introducing the adaptive dynamic weight w_i^t in this form is as follows. From previous equations it is known that the acceleration is minimum for the best particle (with maximum mass) and maximum for the worst particle (with minimum mass). Thus the position (which is the output variable) of the best particle, being nearest to the solution, undergoes the least amount of change in each iteration. Hence there is a specific theoretical insight which guides the update equations which may sometimes be compromised by the random term ($rand_i$) in equation 2.26.

Furthermore instead of providing the worst particle with the maximum change in velocity (and hence position), a more intuitive option would be to design the weight of the velocity $v_i^d(t)$ in such a way as to boost the change in velocity in particles of comparatively higher fitness than the worst so as to accelerate convergence. This is achieved by designing the adaptive weight w_i^t as in equation 2.26 such that it is higher for the particles with fitness nearer to the average value and lower for the particles with fitness far away from the average value.

This modification in conjunction with the discrimination by the acceleration term is introduced in a bid to achieve several important objectives. Firstly, the minimum change applicable for the best particle is preserved in the acceleration term by removing the effect of the random term in the velocity update equation. Secondly, instead of only preferring maximum change in velocity of the worst particle, the change in velocity of particles with better fitness is deliberately boosted for better convergence (through introduction of the adaptive weight term). The Algorithm 2.2 describes the implementation of the proposed BAW-GSA. Figure 2.5 presents the same algorithm in flow chart form.

Algorithm 2.2: BAW-GSA

BEGIN

Create p particles and make randomized initialization of their n dimensional positions X in $\{0, 1\}$.

Initialize iteration number $t = 1$

REPEAT:

FOR $i = 1$ to p

Calculate fitness $fit_i^t = \frac{Dintra_i^t}{Dinter_i^t}$

where $Dintra_i^t = \sum_c \sum_{u,r \in c, u \neq r} \left\| F_{u,i}^t - F_{r,i}^t \right\|$

and $Dinter_i^t = \sum_{c,c'} \sum_{u \in c, r \in c', c \neq c'} \left\| F_{u,i}^t - F_{r,i}^t \right\|$

END FOR

Calculate global best fitness: $best(t) = \min_{i \in \{1,2,\ldots p\}} fit_i(t)$

Calculate global worst fitness: $worst(t) = \max_{i \in \{1,2,\ldots p\}} fit_i(t)$

FOR $i = 1$ to p

Calculate mass: $m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)}$, $M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{p} m_j(t)}$

Calculate gravitational constant: $G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta}, \beta < 1, t_0 = 1$

FOR $j = 1$ to p, $j \neq i$

Calculate distance between two particles: $R_{ij}(t) = \left\| X_i(t), X_j(t) \right\|$

Calculate force between two particles:

FOR $d = 1$ to n

$$F_{ij}^d(t) = G(t) \frac{M_i(t) \times M_j(t)}{R_{ij}(t) + \varepsilon} \left(x_j^d(t) - x_i^d(t) \right)$$

END FOR

END FOR

Calculate average fitness of all particles: $\overline{fit^t} = \frac{1}{p} \sum_{i=1}^{p} fit_i^t$

Calculate dynamic adaptive weight $w_i^t = \left| \frac{\min (fit_i^t, \overline{fit^t})}{\max (fit_i^t, \overline{fit^t})} \right|$

FOR $d = 1$ to n

Calculate total force on ith particle: $F_i^d(t) = \sum_{j=1, j \neq i}^{p} rand_j F_{ij}^d(t)$

Calculate acceleration: $a_i^d(t) = \frac{F_i^d(t)}{M_i(t)}$

Calculate velocity: $v_i^d(t+1) = w_i^t \times v_i^d(t) + a_i^d(t)$

Calculate position: $x_i^d(t+1) = \begin{cases} 1, & \frac{1}{1 + e^{-V_i^d(t+1)}} > rand \\ 0, & otherwise \end{cases}$

END FOR

END FOR

UNTIL termination criterion is satisfied

END

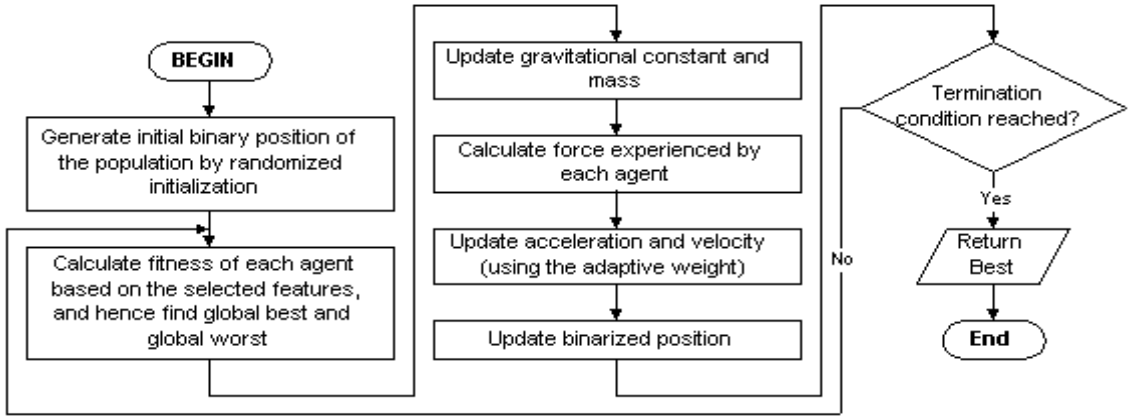

Fig. 2.5. Flow chart for the proposed BAW-GSA.

2.4 Neural Network based Classification

Once the final feature set is selected, the training and testing is performed using neural network based classifiers. Back propagation neural networks (BPNN), employing multi-layered feedforward architecture, still remain one the most popular variants of supervised neural networks which utilize the traditional method of gradient descent and other first order and second order optimization algorithms to suitably determine the weights and biases. Some of these very popular learning algorithms include Levenberg-Marquardt (LM) learning based BPNN, resilient backpropagation, conjugate gradient-Fletcher Powell based BPNN algorithm etc. In this Chapter, a three-layered architecture with 20 hidden layer neurons and the LM-learning based BPNN is employed for training exemplars from the training set.

Fig. 2.6. A few sample images of one person from Yale A database.

Fig. 2.7. A few sample images of one person from Yale B Extended database.

Fig. 2.8. A few sample images of one person from the ORL database

2.5 Experiments and Results

The binary version of the traditional GSA for feature selection proposed in Section 2.3.2 and the Binary Adaptive Weight based variation (BAW-GSA) proposed in Section 2.3.3 have been separately evaluated for each separate set of extracted features employing LBP, LGP and MCT based feature extraction algorithms for three popular standard, benchmark databases, namely, Yale A, Yale B Extended [33], [34], [35] and ORL [36], [37] face databases. The results obtained are compared with several competing, state-of-the-art algorithms in existence for solving identical face recognition problems.

The Yale A database published by Yale University has 11 images each from 15 individuals with lighting, occlusion and facial emotion variations. Figure 2.6 shows 10 sample images for one

person in the Yale A database. The Yale B Extended database published by Yale University has 64 images each from 32 individuals with similar variations of emotion, lighting, etc. Figure 2.7 shows 10 sample images for one person in the Yale B Extended database. The ORL database published by Cambridge University has 10 images each from 40 individuals. Here there are also variations of occlusion and facial expressions along with changes in scale and tilting. Figure 2.8 shows the 10 different images for one person in the ORL database.

Extensive performance evaluation has been carried out for images from all three face databases and the experimental results obtained for Yale A, Yale B Extended and ORL database are presented in table 2.1, table 2.2 and table 2.3 respectively. For each database, the size of the training dataset is varied in conformation with the similar variations utilized to report other competing algorithms in existing literatures, in a bid to provide performance comparisons in similar experimental platforms. Several competing algorithms, whose performances, in terms of percentage errors in face recognition, are comprehensively reported in [10], are considered for comparing the performances of our proposed algorithms. In these tables, some entries for competing algorithms are kept blank because those performances were not reported in their original works. For each training dataset size considered, the winning algorithm, i.e. the algorithm that showed the lowest percentage error rate, is shown in bold.

A closer inspection of these tabular results will reveal that, like other competing algorithms, each of the six algorithms proposed in this Chapter, namely LBP+BGSA, MCT+BGSA, LGP+BGSA, LBP+BAW-GSA, MCT+BAW-GSA, LGP+BAW-GSA, produced enhance recognition rates with increase in the size of the training dataset, which is understandable. Overall, the performances of LBP based algorithms were bettered by corresponding MCT based algorithms and the performances of MCT based algorithms were bettered by corresponding LGP based

algorithms. Also one can easily conclude that the LGP+BAW-GSA algorithm has emerged as the best performing algorithm which produced best performance in seven situations out of eight possible situations for Yale A database, in six situations out of eight possible situations for Yale B extended database and in eight situations out of eight possible situations for ORL database. In fact, in table 2.1, the LGP+BAW-GSA scheme could produce at least 100% improvement in recognition performance compared to the competing, previously published algorithms considered and this improvement in performance undergoes steep rise with increase in the training dataset size. A similar trend in improvement in system performance can also be observed in table 2.2.

On the other hand, for the ORL database, the most improved performance, in comparison with the competing, previously published algorithms, is observed for the smallest training dataset size. This conclusively proves that when the proposed BAW-GSA feature selection algorithm is hybridized with the LGP feature extraction scheme, then the face recognition performance is significantly enhanced, compared to several similar, contemporary, existing, competing face recognition algorithms.

Table 2.1

Percentage error rates for Yale A database

Method | Training size	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
S-LDA [21, 10]	42.4	27.7	22.2	18.3				
S-LDA (updated) [21, 10]	37.5	25.5	19.3	14.7	12.3	10.3	8.7	
TANMM [20, 10]	44.69	29.57	18.44					
OLAP [14, 10]	44.3	29.9	22.7	17.9				
Eigenfaces [11, 10]	56.5	51.1	47.8	45.2				
Fisherfaces [12, 10]	54.3	35.5	27.3	22.5				
Laplacianfaces [13, 10]	43.5	31.5	25.4	21.7				
Volterrafaces(linear) [10]	15.7	12.33	9.47	6.11	5.78	3.96	2.61	1.43
Volterrafaces(quadratic) [10]	22.15	13.36	15.78	10.19	10.04	9.66	9.49	8.74
LBP + BGSA	10.0	8.7	6.6	5.9	4.4	2.7	1.5	1.2
MCT + BGSA	9.3	8.2	6.0	5.3	3.6	2.4	1.3	0.9
LGP + BGSA	8.7	7.1	5.9	4.2	2.4	1.3	1.0	0.7
LBP + BAW-GSA	9.3	7.2	6.4	5.7	4.0	2.5	1.7	1.1
MCT + BAW-GSA	8.6	**7.0**	5.5	5.1	3.7	2.3	1.4	0.8
LGP + BAW-GSA	**8.0**	7.4	**4.6**	**3.5**	**1.9**	**1.2**	**1.0**	**0.4**

Table 2.2
Percentage error rates for Yale B extended database

Method | Training size	**2**	**3**	**4**	**5**	**10**	**20**	**30**	**40**
ORO [18, 10]							9.0	
SR [17, 10]					12.0	4.7	2.0	1.0
RDA [17, 10]					11.6	4.2	1.8	0.9
KLPPSI [19, 10]				24.74	9.93	3.15	1.39	
KRR [15, 10]				23.9	11.04	3.67	1.43	
CTA [22, 10]				16.99	7.60	4.96	2.94	
Eigenfaces [11, 10]				54.73	36.06	31.22	27.71	
Fisherfaces [12, 10]				37.56	18.91	16.87	14.94	
Laplacianfaces [13, 10]				34.08	18.03	30.26	20.2	
MLASSO [16, 10]	58	54	50					
Volterrafaces(linear) [10]	26.23	18.23	9.33	6.35	2.67	0.9	0.42	0.34
Volterrafaces (quadratic) [10]	40.81	20.47	14.42	13.00	3.98	1.27	0.58	0.43
LBP + BGSA	16.5	12.6	8.9	6.1	2.1	1.1	0.5	0.3
MCT + BGSA	16.3	12.2	8.6	6.0	1.8	1.0	0.4	0.2
LGP + BGSA	16.1	**11.5**	8.6	5.6	1.7	**1.0**	0.3	0.2
LBP + BAW-GSA	16.2	12.4	8.3	5.8	1.6	1.1	0.4	0.2
MCT + BAW-GSA	16.0	12.1	8.0	5.7	1.4	1.1	0.3	0.2
LGP + BAW-GSA	**15.7**	11.8	**7.8**	**5.3**	**1.3**	1.1	**0.3**	**0.1**

Table 2.3
Percentage error rates for ORL Face database

Method | Training size	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Volterrafaces (linear) [10]	29.89	21.39	16.2	11.98	9.59	6.96	6.77	4.37
Volterrafaces (quadratic) [10]	31.77	21.91	16.52	13	9.44	7.65	5.69	4.69
LBP + BGSA	16.8	16.0	13.5	11.3	9.5	6.0	5.2	3.2
MCT + BGSA	16.5	15.1	12.6	10.6	8.7	5.9	4.4	3.2
LGP + BGSA	15.9	14.5	12.1	9.9	8.2	5.6	3.9	2.8
LBP + BAW-GSA	16.2	15.2	12.4	10.1	8.4	5.9	4.7	2.3
MCT + BAW-GSA	15.7	14.1	12.1	9.8	8.0	5.5	4.2	2.1
LGP + BAW-GSA	**15.0**	**13.7**	**11.5**	**9.3**	**7.5**	**5.0**	**3.2**	**1.8**

2.6 Conclusion

The present Chapter demonstrated how an effective feature selection scheme can be employed for face recognition problems, employing metaheuristic optimization algorithms. A recently proposed algorithm in this genre, called GSA, has been chosen to demonstrate its effectiveness in this regard. At first, the formulation of the fitness function for the purpose of feature selection is carried out using a ratio of the within class distance to the between class distance. As the traditional GSA cannot be used for solving this specific problem, a binary version of the traditional GSA is first developed. Then a novel variation of this BGSA is developed with dynamic adaptive inertia weight, named BAW-GSA, to further enhance its search capability. These feature selection schemes are hybridized with local binary pattern (LBP), modified census transform (MCT), and local gradient pattern (LGP) based well known feature extraction algorithms for face recognition and they were extensively tested for three benchmark face databases. Extensive experimentations carried out established that the proposed feature selection methodologies can yield improved face recognition performance compared to several other competing face recognition algorithms, and, among all algorithms considered, the hybrid LGP + BAW-GSA methodology produced the most superior performance.

References

[1] A. K. Jain and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. Circuits and Systems for video Technology*, vol. 14, no. 1, pp. 4-20, Jan. 2004.

[2] R. Jafri and H. R. Arabnia, "A Survey of Face Recognition Techniques," *Journal of Information Processing Systems*, vol. 5, no. 2, pp. 41-68, June 2009.

[3] 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, Dec. 2003.

[4] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, Oct. 1993.

[5] X. Li, G. Mori, and H. Zhang, "Expression-Invariant Face Recognition with Expression Classification," *Proc. IEEE Third Canadian Conference on Computer and Robot Vision*, pp. 77, June 2006.

[6] A. Barmpoutis, R. Kumar, B. C. Vemuri, and A. Banerjee, "Beyond the Lambertian assumption: A generative model for Apparent BRDF fields of faces using anti-symmetric tensor splines," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[7] V. Blanz and T. Vetter, "Face Recognition based on fitting a 3d Morphable Model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, Sept. 2003.

[8] R. M. Ramadan and 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, pp. 51-66, June 2009.

[9] R. F. Abdel-Kader, R. M. Ramadan, and R. Y. Rizk, "Rotation invariant face recognition based on hybrid LPT/ DCT features," *International Journal of Electrical and Computer Engineering*, vol. 3, no. 7, pp. 488-493, 2008.

[10] R. Kumar, A. Bannerjee, and B. C. Vemuri, "Volterrafaces: Discriminant Analysis using Volterra Kernels," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 150-155, June 2009.

[11] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, June 1994.

[12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.

[13] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face Recognition using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, March 2005.

[14] D. Cai, X. He, J. Han, and H. J.Zhang, "Orthogonal Laplacianfaces for Face Recognition," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3608-3614, Nov. 2006.

[15] S. An, W. Liu, and S. Venkatesh, "Face recognition using kernel ridge regression," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2007.

[16] D. Pham and S. Venkatesh, "Robust Learning of Discriminative Projection for Multicategory Classification on the Stiefel Manifold," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[17] D. Cai, X. He, and J. Han, "Spectral Regression for Efficient Regularized Subspace Learning," *Proc. IEEE International Conference on Computer Vision*, Oct. 2007.

[18] G. Hua, P. A. Viola, and S. M. Drucker, "Face Recognition using Discriminatively Trained Orthogonal Rank One Tensor Projections," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2007.

[19] S. An, W. Liu, and S. Venkatesh, "Exploiting side information in locality preserving projection," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[20] F. Wang and C. Zhang, "Feature Extraction by Maximizing the Average Neighborhood Margin," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[21] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2007.

[22] Y. Fu and T. S. Huang, "Image Classification using Correlation Tensor Analysis," *IEEE Trans. Image Processing*, vol. 17, no. 2, pp. 226-234, Feb. 2008.

[23] C.-J. Tu, L.-Y. Chuang, *J.-Y. Chang,* and C.-H. Yang, "Feature Selection using PSO-SVM," *International Journal of Computer Science* (*IAENG)*, vol. 33, no. 1, Feb. 2007.

[24] Z. Cataltepe, H. M. Genc, and T. Pearson, "A PCA/ICA based feature selection method and its application for corn fungi detection," *15th European Signal Processing Conference*, Poznan, Poland, 2007.

[25] H. R. Kanan, K. Faez, and M. Hosseinzadeh, "Face Recognition System using Ant Colony Optimization based Selected Features," *IEEE Symposium Computational Intelligence in Security and Defense Applications*, pp. 57-62, April 2007.

[26] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[27] B. Froba and A. Ernst, “Face Detection with the Modified Census Transform,” *Proc. IEEE Sixth International Conference on Face and Gesture Recognition*, pp. 91–96, May 2004.

[28] B. Jun and D. Kim, “Robust face detection using local gradient patterns and evidence accumulation,” *Pattern Recognition*, vol. 45, no. 9, pp. 3304-3316, Sep. 2012.

[29] E. Rashedi, H. Nezamabadi-pour and S. Saryazdi, “GSA: A Gravitational Search Algorithm,” *Information Sciences*, vol. 179, no. 13, pp. 2232-2248, June 2009.

[30] J. Kennedy and R. C. Eberhart, “A discrete binary version of the particle swarm algorithm,” *Proc. IEEE International Conference on Systems, Man, and Cybernetics, Computer Cybernetics and Simulation,* vol. 5, pp. 4104-4108, Oct. 1997

[31] M. A. Khanesar, M. Teshnehlab, and M. A. Shoorehdeli, “A Novel Binary Particle Swarm Optimization,” *Proc. IEEE Mediterranean Conference on Control and Automation*, Tehran, June 2007.

[32] X. Jun and H. Chang, “The Discrete Binary Version of the Improved Particle Swarm Optimization Algorithm,” *Proc. IEEE International Conference on Management and Service Science,* Sept. 2009

[33] Yale Univ. A Face Database <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>, 2002.

[34] Yale Univ. B Extended Database <http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html >, 2001.

[35] A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, June 2001.

[36] ORL database <http://www.cam-orl.co.uk/facedatabase.html>, 2002.

[37] F. Samaria and A. Harter, "Parameterization of a Stochastic Model for Human Face Identification," *Proc. Second IEEE workshop on Applications of Computer Vision*, pp. 138-142, Dec. 1994.

[38] R. Mansouri, F. Nasseri, M. Khorrami, "Effective time variation of G in a model universe with variable space dimension," *Physics Letters A*, vol. 259, no. 3-4, pp. 194-200, August 1999.

CHAPTER 3

Face Recognition with single image per person using SVD and GSA

3.1 Introduction

In recent years face recognition has become a widely researched topic since it has numerous real world applications like authentication, identification, advanced human computer interaction and many other emerging fields of research. Face recognition spans the subjects of pattern recognition, image processing, computer vision, machine learning, etc. With the growing importance of biometric recognition systems [1], due to low susceptibility to security loss, face recognition based biometrics has gained much popularity in recent times.

Many approaches to face recognition have been proposed over the last two decades ([2], [3]) most of which are based on supervised learning. Hence they follow a common sequence of steps. There is a feature extraction step in which a set of discriminating features are extracted from a set of training images [4]. Then if the set of features extracted is dimensionally large, there may be a feature selection/reduction procedure where a reduced set of highly discriminating features is selected employing a suitable algorithm which may attempt to optimize a suitable cost function. For the face recognition problem, since images of different persons are, after all, human faces they have some common characteristics which indicates that some features in a large set of extracted features will not have enough discriminating power. This makes feature selection/reduction an important step as a large feature set might not necessarily result in a

higher recognition rate [5]. This step is followed by the classification step where the final conclusion regarding recognition or authentication is actually performed. Several variations of methods proposed in each of these steps generates new approaches in solving the problem.

A general drawback of the supervised learning method is that for a good classification accuracy rate, the number of training samples needs to be sufficiently large (depending on the number of test images and the number of classes). In those particular methods where inter class and intra class distances are used, the methods do not work at all when there is a single training image of each subject because in this case, the intra class distances are not defined [11]. This drawback is prominent in several approaches which include the popular methodology of Fisher Linear Discriminant Analysis (FLDA). A few methods have been proposed in recent years to solve this problem of FLDA based face recognition where there is only one training image per person e.g. generalized inverse method [6], perturbation based method [7], direct FLDA method [8], null space method [9], 2-D FLDA method [10], singular value decomposition (SVD) based method [11] etc.

The present Chapter is inspired by the method proposed by Gao et al. [11] in which the single training image of a particular class is decomposed into two component images using SVD and then the intra class distance can be conveniently determined using these two resulting images. However, the SVD based FLDA approach in [11] uses the general eigen value theory to solve the cost function. In this Chapter a novel method of solving the FLDA cost function using an intelligent iterative stochastic optimization algorithm is proposed which can simultaneously solve the feature selection/reduction phase along with the feature extraction phase, thus effectively merging the operations required in two steps. The iterative stochastic optimization problem is solved using a recently proposed method, called gravitational search algorithm

(GSA). GSA is a powerful iterative optimization algorithm based on Newton's laws of gravity and motion [12]. Three modifications of the GSA have been proposed in this Chapter with the objective of solving this problem. The first variation proposes a 2-D GSA to adopt the GSA in processing of 2-D images. The second variation introduces a novel random local extrema based GSA (RLEGSA). To the best of my knowledge and belief, although some local best methods have been proposed earlier for a similar swarm intelligence based method called particle swarm optimization (PSO) [13], [14], this is the first such variation developed in the genre of GSA. The third variation incorporates the automated selection of projection vectors within the GSA based cost function optimization framework.

The rest of this Chapter is organized as follows. Section 3.2 provides a description of the SVD and FLDA based feature extraction schemes for the single training image per person scenario. Section 3.3 presents detailed descriptions of the novel variants of GSA proposed in this Chapter. Section 3.4 presents the experiments and simulation results. Section 3.5 concludes the Chapter.

3.2 SVD based feature extraction scheme

Consider there are C classes with each having a single image $I_k \in \Re^{m\times n}$ ($k=1,\ldots,C$). If $m \geq n$, then let $U_k \in \Re^{m\times m}$ and $V_k \in \Re^{n\times n}$ be the eigenvector matrices of $I_k I_k^T$ and $I_k^T I_k$ respectively. Let u_i^k and v_i^k be the ith column of U_k and V_k respectively. Let σ_i^k be the ith singular value of I_k such that σ_i^k is in descending order of magnitude as i increases i.e. $\sigma_1^k \geq \sigma_2^k \geq \cdots \sigma_{i-1}^k \geq \sigma_i^k \geq \sigma_{i+1}^k \cdots \sigma_n^k$. Then the image can be described as [11, 20]:

$$I_k = \sum_{i=1}^{n} \sigma_i^k . u_i^k . \left(v_i^k\right)^T \tag{3.1}$$

Hence each image can be thought of being constituted as a summation of *n* basis images where each basis image corresponds to a particular singular value and the energy content of a basis image is higher if its associated singular value is higher in magnitude. Following the philosophy described in [11], an image $\widehat{I_k}$ is constructed taking the three most significant SVD basis images as

$$\widehat{I_k} = \sum_{i=1}^{3} \sigma_i^k \,.\, u_i^k \,.\, (v_i^k)^T \tag{3.2}$$

Thus after obtaining $\widehat{I_k}$ one has two image matrices I_k and $\widehat{I_k}$ in each class *k*, and one can also compute a difference image, $\Delta I_k = I_k - \hat{I}_k = \sum_{i=4}^{n} \sigma_i^k u_i^k . \left(v_i^k\right)^T$. The creation of $\widehat{I_k}$ and ΔI_k facilitates the calculation of an approximate within-class scatter matrix, which is, otherwise, not possible when there is only one training image per person/class. Let the within class scatter matrix be denoted as S_w and the between class scatter matrix be denoted as S_b. To compute S_w and S_b we need to compute the global mean image $\bar{I}$ and mean image of the *k*th class $\overline{I_k}$, which are defined as [11]:

$$\overline{I_k} = \frac{1}{2}\left(I_k + \widehat{I_k}\right) \tag{3.3}$$

$$\bar{I} = \frac{1}{2C}\sum_{k=1}^{C}(I_k + \widehat{I_k}) \tag{3.4}$$

Then S_w and S_b can be computed as:

$$S_b = \frac{1}{C}\sum_{k=1}^{C}(\overline{I_k} - \bar{I})^T(\overline{I_k} - \bar{I}) \tag{3.5}$$

$$S_w = \frac{1}{C}\sum_{k=1}^{C}[(I_k - \overline{I_k})^T(I_k - \overline{I_k}) + \left(\widehat{I_k} - \overline{I_k}\right)^T\left(\widehat{I_k} - \overline{I_k}\right)] \tag{3.6}$$

Using (3.3) and (3.4), one can obtain [11]:

$$S_w = \frac{1}{2C}\sum_{k=1}^{C}(I_k - \widehat{I_k})^T\,(I_k - \widehat{I_k}) \tag{3.7}$$

From the theory of two-dimensional FLDA [10], then the goal will be to seek a set of d optimal discriminating column vectors w_j (j = 1, 2, …, d) constituting $m \times d$ optimal projection matrix W so as to minimize the cost function

$$J(W) = \frac{trace(W^T S_w W)}{trace(W^T S_b W)} \tag{3.8}$$

Once W is determined, all the training images are projected on to W to obtain the feature matrices Z_k

$$Z_k = I_k \times W, \quad k = 1,2,\cdots,C \tag{3.9}$$

Then, for each class k there is a feature matrix Z_k. If there is an input test image I then the corresponding feature matrix is $Z = I \times W$

Then one can utilize the nearest neighbor classifier method, where the Euclidean distance D_k of feature matrix Z from each feature matrix Z_k pertaining to class k is calculated as:

$$D_k = \|Z_k - Z\|, \; k = 1, \ldots, c \tag{3.10}$$

Then, I is identified to belong to that class k for which D_k is minimum.

Now in [11] the optimal value of W has been determined by using the general eigen value theorem and for that the cost function considered is the inverse of (3.8), which is the traditional cost function considered in [10], as they solved a maximization problem. In this Chapter, to solve a minimization function by utilizing the gravitational search algorithm and its proposed variants, a form of $J(W)$ has been utilized which is inverse of that considered in [10], [11].

3.3 Proposed Variants of GSA for Feature Selection

The gravitational search algorithm (GSA) is a relatively recently proposed optimization algorithm based on the Newtonian law of gravity [12]. The traditional GSA has been described previously in Chapter 2 in subsection 2.3.1 along with the relevant flow chart and algorithm.

3.3.1 2-D Gravitational Search Algorithm (2-D GSA)

The first modification of GSA proposed in this Chapter is developed in order to account for a two dimensional solution space. In this case the position variable for the *i*th particle (*i* = 1, 2, ..., *p*) becomes an $n \times n$ matrix (corresponding to the size of the *W* matrix) as follows:

$$X_i(t) = \left\{x_i^{kl}(t)\right\} \quad i = 1, \dots, p; k = 1, \dots, n; l = 1, \dots, n \tag{3.11}$$

Thus the modified force calculation equation between two particles is developed as:

$$F_{ij}^{kl}(t) = G(t)\frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \varepsilon}\left(x_j^{kl}(t) - x_i^{kl}(t)\right) \tag{3.12}$$

where $R_{ij}(t)$ is the Euclidean distance between two agents *i* and *j*, given as:

$$R_{ij}(t) = \left\|X_i(t), X_j(t)\right\| \tag{3.13}$$

The total gravitational force on a particle *i* in the dimension *k*, *l* is

$$F_i^{kl}(t) = \sum_{j=1, j \neq i}^{p} rand_j F_{ij}^{kl}(t) \tag{3.14}$$

Accordingly, the modified acceleration, velocity and position update equations are given as:

$$a_i^{kl}(t) = \frac{F_i^{kl}(t)}{M_{ii}(t)} \tag{3.15}$$

$$v_i^{kl}(t+1) = rand_i \times v_i^{kl}(t) + a_i^{kl}(t) \tag{3.16}$$

$$x_i^{kl}(t+1) = x_i^{kl}(t) + v_i^{kl}(t+1) \tag{3.17}$$

The mass update equations remain the same as in the traditional GSA.

3.3.2 2-D Random Local Extrema Gravitational Search Algorithm (2-D RLEGSA)

This is the second proposed variation of GSA in this Chapter, where instead of using the global best and global worst values in the mass update equation, the local best and local worst values are used. For each particle, these are the best and worst fitness values in the 'locality' or neighborhood of that particle. The process of determination of this neighborhood is the main contribution of this method. The method proposes incorporation of a random element in the determination of neighborhood to boost the rate of convergence.

Like before let there be p particles with n dimensional position vectors $X_i(t)$, i=1, 2, …, p. Now a normalized Euclidian distance of ith particle from any other particle is calculated as:

$$d_{ij} = \frac{\|X_i - X_j\|}{\max_j(\|X_i - X_j\|)} \quad where\ j = 1\ to\ p\ and\ j \neq i \tag{3.18}$$

A corresponding distance vector for particle i is created as:

$$D_i = \{d_{ij}\}, j = 1,2, \dots, p \tag{3.19}$$

Now, let the vector D_i be rearranged such that its individual elements appear in ascending order of magnitude and let this new vector be denoted as D_i^{asc}. At iteration number t, those particles in the neighborhood of particle i are considered which have distances corresponding to the first $q_i(t)$ number of elements of D_i^{asc}, where $q_i(t)$ is given as:

$$q_i(t) = round(\left[K \times \frac{t}{iter_{max}}\right] \times p \times rand[0,1]) \tag{3.20}$$

If $q_i(t) > p$ then $q_i(t) = p$, and K=scaling constant.

Once the neighborhood of the ith particle is thus found out, the best and worst fitness values are calculated over the neighborhood only, hence the name of the method. As the number of iterations increases, the neighborhood of each particle grows in size and it finally approaches the traditional global extrema case, where the entire search space becomes the neighborhood of each particle. The Algorithm 3.2 describes the implementation of the proposed Random Local

Extrema based GSA (RLEGSA). Figure 3.2 presents the same RLEGSA algorithm in flow chart form. The entire algorithm has been developed as a modification of the 2-D GSA described in section 3.3.1 and hence it is called 2-D RLEGSA.

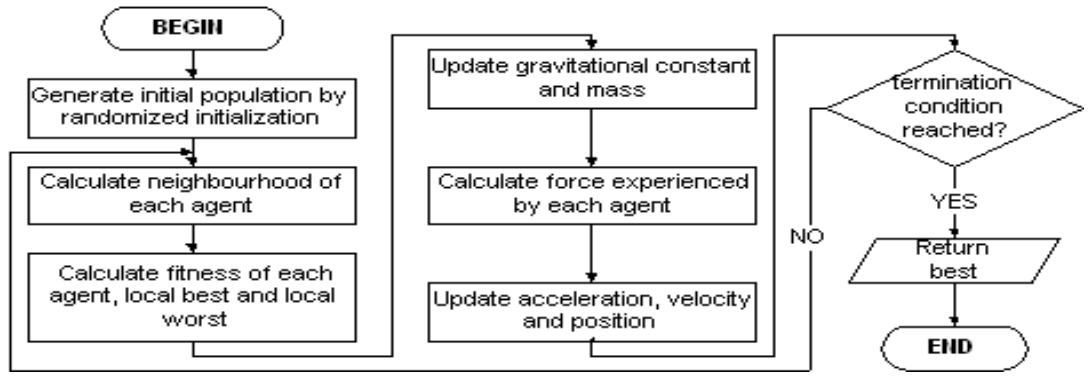

Fig. 3.2. Flow chart for the proposed RLEGSA.

Algorithm 3.2: RLEGSA

BEGIN

Create p particles and make randomized initialization of their positions X. $X_i(t) = \{x_i^{kl}(t)\}$ $i = 1, \dots, p; k = 1, \dots, n; l = 1, \dots, n$. Initialize iteration number $t = 1$.

REPEAT:

FOR $i = 1$ to p

FOR $j = 1$ to $p, j \neq i$

Calculate normalized distance between two particles:

$$d_{ij} = \frac{\|X_i - X_j\|}{\max_j(\|X_i - X_j\|)}$$

END FOR

Form distance vector: $D_i = \{d_{ij}\}, j = 1, \dots, p$ $\quad D_i^{asc} = sort_{ascend}(D_i)$

Calculate neighborhood:

$$N_i^{asc} = particles\ corresponding\ to\ first\ q_i^t\ elements\ of\ D_i^{asc}$$

$$q_i^t = \left[K \times \frac{t}{iter_{max}}\right] \times p \times rand[0,1]$$

$$if\ q_i^t > p\ then\ q_i^t = p$$

Calculate fitness $fit_i(t)$

Calculate local best fitness: $lbest_i(t) = \min_{l \in \{N_i^{asc}\}} fit_l(t)$

Calculate global best fitness: $lworst_i(t) = \max_{l \in \{N_i^{asc}\}} fit_l(t)$

Calculate mass: $m_i(t) = \frac{fit_i(t) - lworst_i(t)}{lbest_i(t) - lworst_i(t)}$, $M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{p} m_j(t)}$

Calculate gravitational constant: $G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta}, \beta < 1, t_0 = 1$

FOR $j = 1$ to $p, j \neq i$

Calculate distance between two particles: $R_{ij}(t) = \|X_i(t), X_j(t)\|$

Calculate force between two particles:

FOR $k = 1$ to n

FOR $l = 1$ to n

$$F_{ij}^{kl}(t) = G(t) \frac{M_i(t) \times M_j(t)}{R_{ij}(t) + \varepsilon} \left(x_j^{kl}(t) - x_i^{kl}(t)\right)$$

END FOR

END FOR

END FOR

FOR $k = 1$ to n

FOR $l = 1$ to n

Calculate total force on ith particle: $F_i^{kl}(t) = \sum_{j=1, j \neq i}^{p} rand_j F_{ij}^{kl}(t)$

Calculate acceleration: $a_i^{kl}(t) = \frac{F_i^{kl}(t)}{M_i(t)}$

Calculate velocity: $v_i^{kl}(t+1) = rand_i \times v_i^{kl}(t) + a_i^{kl}(t)$

Calculate position: $x_i^{kl}(t+1) = x_i^{kl}(t) + v_i^{kl}(t+1)$

END FOR

END FOR

END FOR

UNTIL termination criterion is satisfied

END

3.3.3 Modified Random Local Extrema Gravitational Search Algorithm (MRLEGSA) with automated selection of number of projection vectors

In [11], the scheme actually utilized different subsets of projection vectors from the optimal transform matrix solution W obtained by solving the generalized eigenvalue problem. The selection of number of f such projection vectors from the n vectors in W matrix ($f \leq n$) and which such f vectors should be chosen were both carried out manually. Their extensive experimentations showed that, for each separate face database, the performance varies individually with variation in number and choices of f vectors and no correlation can be obtained in determining their optimal choices which means they had to make an extensive manual search for the best combination of f vectors for each face database. This provided the inspiration to propose an improved algorithm which can perform an additional task of an automated selection of the number of optimal projection vectors to be chosen from the W matrix and which candidate f vectors should be chosen that will lead to better performance accuracy. This has been referred to here as the third proposed modification of GSA. It actually augments the RLEGSA proposed in the previous subsection by incorporating the automated selection of number of projection vectors and also which projection vectors to be chosen from the W matrix.

Here for the ith agent, the solution space contains, besides the $n \times n$ matrix $X_i(t)$, an $n \times 1$ vector $S_i(t)$, given as:

$$X_i(t) = \left\{x_i^{kl}(t)\right\}_{n \times n} \quad i = 1, \dots, p; k = 1, \dots, n; l = 1, \dots, n \tag{3.21}$$

$$S_i(t) = \left\{s_i^{l}(t)\right\} \qquad i = 1, \dots, p; l = 1, \dots, n \tag{3.22}$$

The vector $S_i(t)$ is named the selection vector since it will be used to select the optimal number of projection vectors. Each element in the selection vector $S_i(t)$ is initialized by a number pulled from a uniform random distribution in [0, 1] and then it undergoes modifications in its velocity and position in each subsequent iteration. At any iteration t, if the position of the lth element of the vector $S_i(t)$ i.e. $s_i^l(t)$ has a value greater than 0.5 then the lth column of the position matrix $X_i(t)$, i.e., the lth projection vector, is selected, otherwise, if this value is less than 0.5, then the lth projection vector is not selected. This process is repeated for each lth dimension of the vector $S_i(t)$ in each iteration and, at the completion of the implementation of the modified GSA algorithm, from the best $X_i(t)$ solution i.e. the final W solution, those projection vectors are finally selected that are chosen according to the best solution of the $S_i(t)$ vector.

As the Modified RLEGSA algorithm now has to evolve both the candidate $X_i(t)$ matrices and the candidate $S_i(t)$ vectors in each iteration, the corresponding equations for calculating force between two particles are modified as:

$$Fx_{ij}^{kl}(t) = G(t)\frac{M_{pi}(t)\times M_{aj}(t)}{Rx_{ij}(t)+\varepsilon}\left(x_j^{kl}(t) - x_i^{kl}(t)\right) \tag{3.23}$$

$$Fs_{ij}^{l}(t) = G(t)\frac{M_{pi}(t)\times M_{aj}(t)}{Rs_{ij}(t)+\varepsilon}\left(s_j^{l}(t) - s_i^{l}(t)\right) \tag{3.24}$$

$$Rx_{ij}(t) = \left\|X_i(t), X_j(t)\right\| \tag{3.25}$$

$$Rs_{ij}(t) = \left\|S_i(t), S_j(t)\right\| \tag{3.26}$$

Now, the total gravitational force on a particle i is calculated as:

$$Fx_i^{kl}(t) = \sum_{j=1, j\neq i}^{p} rand_j Fx_{ij}^{kl}(t) \tag{3.27}$$

$$Fs_i^{l}(t) = \sum_{j=1, j\neq i}^{p} rand_j Fs_{ij}^{l}(t) \tag{3.28}$$

Accordingly, new acceleration update equations are developed as:

$$ax_i^{kl}(t) = \frac{Fx_i^{kl}(t)}{M_{ii}(t)} \tag{3.29}$$

$$as_i^{l}(t) = \frac{Fs_i^{l}(t)}{M_{ii}(t)} \tag{3.30}$$

This gives the new velocity update equations as:

$$vx_i^{kl}(t+1) = rand_i \times vx_i^{kl}(t) + ax_i^{kl}(t) \tag{3.31}$$

$$vs_i^{l}(t+1) = rand_i \times vs_i^{l}(t) + as_i^{l}(t) \tag{3.32}$$

Hence the new position and selection update equations are given as:

$$x_i^{kl}(t+1) = x_i^{kl}(t) + vx_i^{kl}(t+1) \tag{3.33}$$

$$s_i^{l}(t+1) = s_i^{l}(t) + vs_i^{l}(t+1) \tag{3.34}$$

$$smax_i(t+1) = \max\{S_i(t+1)\} \tag{3.35}$$

$$smin_i(t+1) = \min\{S_i(t+1)\} \tag{3.36}$$

$$\widehat{s_i^{l}}(t+1) = \frac{\mathrm{s_i^l(t+1) - smin_i(t+1)}}{\mathrm{smax_i(t+1) - smin_i(t+1)}} \tag{3.37}$$

$$sbin_i^{l}(t+1) = \begin{cases} 0, & \widehat{s_i^{l}}(t+1) < 0.5 \\ 1, & \widehat{s_i^{l}}(t+1) \geq 0.5 \end{cases} \tag{3.38}$$

Now normalized Euclidian distance of the *i*th particle from any other particle is given as:

$$d_{ij} = \frac{\|X_i - X_j\|}{\max_j(\|X_i - X_j\|)} \;\; where\; j = 1\; to\; p\; and\; j \neq i \tag{3.39}$$

The corresponding distance vector for particle *i* is given as:

$$D_i = \{d_{ij}\}, j = 1,2,\dots,p \tag{3.40}$$

$$D_i^{asc} = sort_{ascend}(D_i) \tag{3.41}$$

Thus the neighborhood of the *i*th particle is given as:

$$N_i^{asc} = particles\; corresponding\; to\; first\; q_i^t\; elements\; of\; D_i^{asc}$$

where

$$q_i(t) = \left[K \times \frac{t}{iter_{max}}\right] \times p \times rand[0,1] \tag{3.42}$$

If $q_i(t) > p$ then $q_i(t) = p$, and K= scaling constant.

$fit_i(t)$ is the fitness value of the agent i at time t.

Now, the selection variables introduced will become operative while calculating the fitness $fit_i(t)$. The feature matrix $W_i(t)$ used for evaluation of the fitness function corresponding to the candidate solution $X_i(t)$ is extracted as:

$$W_i(t) = \left\{x_i^{kl}(t)\right\}_{n \times r_i^t} \quad i = 1, \dots, p; k = 1, \dots, n; l \in \{l; sbin_i^l(t+1) = 1\} \tag{3.43}$$

where r_i^t is the number of elements in $sbin_i\,(t + 1)$ (over $l = 1,\ldots, n$) which are equal to 1.

The formulations of $fit_i(t)$, $lbest_i(t)$, $lworst_i(t)$, $m_i(t+1)$ and $M_i(t+1)$ remain unchanged as given in subsection 3.3.2. Algorithm 3.3 describes the implementation of the Modified RLEGSA. Figure 3.3 presents the same algorithm in flow chart form.

Algorithm 3.3: MRLEGSA

BEGIN

Create p particles and make randomized initialization of position matrices X and selection vectors S.

Initialize iteration number $t = 1$

REPEAT:

FOR $i = 1$ to p

FOR $j = 1$ to $p, j \neq i$

Calculate normalized distance between two particles:

$$d_{ij} = \frac{\|X_i - X_j\|}{\max_j(\|X_i - X_j\|)}$$

END FOR

Form distance vector: $D_i = \{d_{ij}\}, j = 1, \dots, p \qquad D_i^{asc} = sort_{ascend}(D_i)$

Calculate neighborhood:

$$N_i^{asc} = particles\ corresponding\ to\ first\ q_i^t\ elements\ of\ D_i^{asc}$$

$$q_i^t = \left[K \times \frac{t}{iter_{max}}\right] \times p \times rand[0,1]$$

$$if\ q_i^t > p\ then\ q_i^t = p$$

$smax_i(t) = \max\{S_i(t)\}$

$smin_i(t) = \min\{S_i(t)\}$

FOR $l = 1$ to dim $\{S_i(t)\}$

$$\widehat{s_i^l}(t+1) = \frac{s_i^l(t+1) - \min_i(t+1)}{\max_i(t+1) - \min_i(t+1)}$$

$$sbin_i^l(t+1) = \begin{cases} 0, & \widehat{s_i^l}(t+1) < 0.5 \\ 1, & \widehat{s_i^l}(t+1) \geq 0.5 \end{cases}$$

END FOR

Compose the feature matrix $W_i(t)$:

$$W_i(t) = \{x_i^{kl}(t);\ x_i^{kl}(t) \in X_i(t)\}$$

$$i = 1, \dots, p; k = 1, \dots, n; l \in \{l; sbin_i^l(t+1) = 1\}$$

$$fit_i(t) = \frac{trace(W_i(t)^T S_w W_i(t))}{trace(W_i(t)^T S_b W_i(t))}$$

END FOR

Calculate local best fitness: $lbest_i(t) = \min_{l \in \{N_i^{asc}\}} fit_l(t)$

Calculate global best fitness: $lworst_i(t) = \max_{l \in \{N_i^{asc}\}} fit_l(t)$

FOR $i = 1$ to p

Calculate mass:

$$m_i(t) = \frac{fit_i(t) - lworst_i(t)}{lbest_i(t) - lworst_i(t)}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{p} m_j(t)}$$

Calculate gravitational constant: $G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta}, \beta < 1, t_0 = 1$

FOR $j = 1$ to $p, j \neq i$

Calculate distance:

$Rx_{ij}(t) = \left\|X_i(t), X_j(t)\right\|$

$Rs_{ij}(t) = \left\|S_i(t), S_j(t)\right\|$

Calculate force:

FOR l = 1 to n

FOR k = 1 to n

$Fx_{ij}^{kl}(t) = G(t)\frac{M_i(t)\times M_j(t)}{Rx_{ij}(t)+\varepsilon}\left(x_j^{kl}(t) - x_i^{kl}(t)\right)$

END FOR

$Fs_{ij}^{l}(t) = G(t)\frac{M_i(t)\times M_j(t)}{Rs_{ij}(t)+\varepsilon}\left(s_j^{l}(t) - s_i^{l}(t)\right)$

END FOR

END FOR

FOR k= 1 to n

FOR l = 1 to n

Calculate total force:

$Fx_i^{kl}(t) = \sum_{j=1, j\neq i}^{p} rand_j Fx_{ij}^{kl}(t)$

$Fs_i^{l}(t) = \sum_{j=1, j\neq i}^{p} rand_j Fs_{ij}^{l}(t)$

Calculate acceleration:

$ax_i^{kl}(t) = \frac{Fx_i^{kl}(t)}{M_i(t)}$

$as_i^{l}(t) = \frac{Fs_i^{l}(t)}{M_i(t)}$

Calculate velocity:

$vx_i^{kl}(t+1) = rand_i \times vx_i^{kl}(t) + ax_i^{kl}(t)$

$vs_i^{l}(t+1) = rand_i \times vs_i^{l}(t) + as_i^{l}(t)$

Calculate position and selection variables:

$x_i^{kl}(t+1) = x_i^{kl}(t) + vx_i^{kl}(t+1)$

$s_i^{l}(t+1) = s_i^{l}(t) + vs_i^{l}(t+1)$

END FOR

END FOR

END FOR

UNTIL termination criterion is satisfied

END

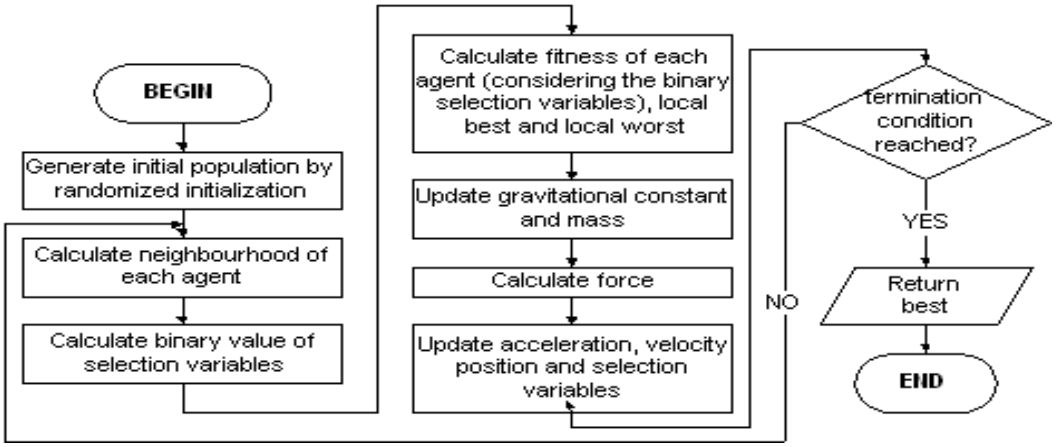

Fig. 3.3. Flow chart for the proposed RLEGSA.

3.4 Experiments and Results

Extensive experimentations have been carried out to determine the effectiveness of the proposed algorithms. At first the 2-D version of the traditional GSA proposed in subsection 3.3.2 and the local extrema based variation (RLEGSA) proposed in Section 3.3.3 have been separately evaluated on two well known benchmark face databases, namely Yale A [18] and ORL [19] databases and the recognition accuracies have been compared with four existing comparable methods, as shown in Table 3.1. The Yale A database published by Yale University has 11 images each from 15 individuals with lighting, occlusion and facial emotion variations. Figure 3.4 shows the 11 different images for one person in the Yale A database. The ORL database published by Cambridge University has 10 images each from 40 individuals. Here there are also variations of occlusion and facial expressions along with changes in scale and tilting. Figure 3.5 shows the 10 different images for one person in the ORL database.

Fig. 3.4. A few sample images of one person from Yale A database.

Fig. 3.5. A few sample images of one person from the ORL database.

During training, the first image of each individual is taken as the single training image. All the remaining images have been taken as testing samples. For classification purpose a simple nearest neighbor classifier is used [11, 22]. These conditions are chosen to be identical with the experimental conditions considered in other methods chosen here for comparison in order to perform a fair comparison in an identical platform. As mentioned, the performance of the proposed method is compared with four state-of-the-art, competing, existing methods, all of whom use different variations of 2-D FLDA for a similar face recognition problem using single training image per person. The first method, proposed by Chen et al. [15], is based on the projection approach. It reported a top recognition accuracy of 18.67% for Yale A database and 44.17% for ORL database. The second method, proposed by Zhang et al. [16], is based on the singular value perturbation approach. It reported a top recognition accuracy of 23.33% for Yale A database and 46.39% for ORL database. The third method, proposed by Chen et al. [17], is

based on the non-overlapping image block approach. It reported a top recognition accuracy of 32.00% for Yale A database and 70.83% for ORL database. The fourth method, proposed by Gao et al. [11], reported a top recognition accuracy of 34.67% for Yale A database and 75.56% for ORL database.

In comparison, by using the 2-D version of the traditional GSA, the top recognition rates achieved are 32.73% for Yale A database (comparable to method 3 [17]) and 75.56% for ORL database (comparable to method 4 [11]). However using the proposed 2-D RLEGSA top recognition rates achieved are 38.57% for Yale A database and 76.92% for ORL database which could comfortably outperform all its five competing algorithms. This shows that the proposed local extrema based version of the 2-D GSA could prove to be indeed effective in achieving better recognition accuracies for a variety of face recognition problems using single training images per person.

To make a performance comparison of the recognition rates for different competing methods with variations in the manually chosen number of projection vectors, for different face databases, the detailed performance evaluation results are plotted in fig. 3.6 and fig. 3.7, for Yale A database and ORL database respectively, where the results of the other four comparing methods are taken from [11]. The curves clearly show that the proposed RLEGSA outperforms the other competing methods in almost all choices of number of projection vectors, for both databases. This results in an ample justification of the fact that RLEGSA can be claimed as an overall superior method in comparison with the four other competing methods considered. Also, these plots signify the variability of the recognition rate with number of projection vectors for each algorithm considered and no trend from these curves can be determined which can facilitate any easy determination of choice of the optimal number of projection vectors. This fact further

strengthens the logic of development of improved methods, using stochastic optimization techniques, which can make an additional search for automated selection of number of projection vectors in order to maximize the recognition rate, as proposed in MRLEGSA.

The concept of top recognition rates, as mentioned in table 3.1, arises for all those methods as experiments are conducted by varying the dimension of the discriminating feature matrix that is by choosing different numbers of projection vectors. However, as mentioned before, the possibility of automated selection of the dimension of the discriminating feature matrix to obtain higher recognition rate had not been considered in the earlier four methods. The proposed MRLEGSA has already been described in subsection 3.3.3 which is armed with this additional feature of selection/automation. Table 3.2 shows the recognition performance comparisons for RLEGSA and MRLEGSA for the two benchmark databases considered. It is seen that the top recognition rates thus achieved with MRLEGSA are almost similar to the results obtained with RLEGSA except for minor deviations which may be attributed to the stochastic nature of activation of each variant of GSA. However, it should be kept in mind that the performance obtained with RLEGSA was obtained after a tedious, manual procedure of selecting different number of projection vectors by turn, activating the algorithm for each such chosen number *f*, noting down their performances, and then determining for which *f* the best performance was achieved. In MRLEGSA this entire procedure is automated involving significant reduction in computational time and effort and yet it could produce almost identical performance as RLEGSA which means it was successful, in each case, to automatically determine the best performing *f* optimal projection vectors, because of which the top performance of RLEGSA could also be approached using MRLEGSA. Hence the objective of

Table 3.1

Top recognition accuracies for six competing methods on two benchmark face databases

Database	Method 1 [15]	Method 2 [16]	Method 3 [17]	Method 4 [11]	Proposed 2-D GSA	Proposed 2-D RLEGSA
Yale A	18.67	23.33	32.00	34.67	32.73	38.57
ORL	44.17	46.39	70.83	75.56	75.56	76.92

Table 3.2

Top recognition accuracies for two variants of GSA for two benchmark face databases

Database	Proposed 2-D RLEGSA	Proposed 2-D MRLEGSA
Yale A	38.57	38.52
ORL	76.92	76.90

introducing the innovation in MRLEGSA to automate a manual, tedious procedure and yet achieve the desired, best performance was fully achieved.

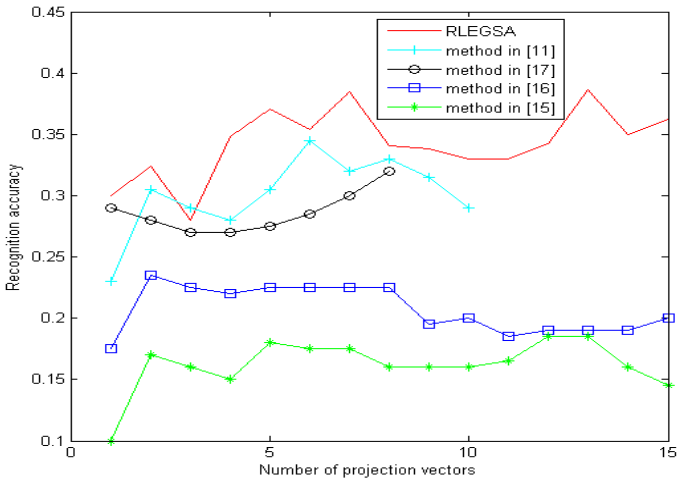

Fig. 3.6. A comparison of recognition accuracy vs. number of projection vectors for five competing methods, for Yale A dataset.

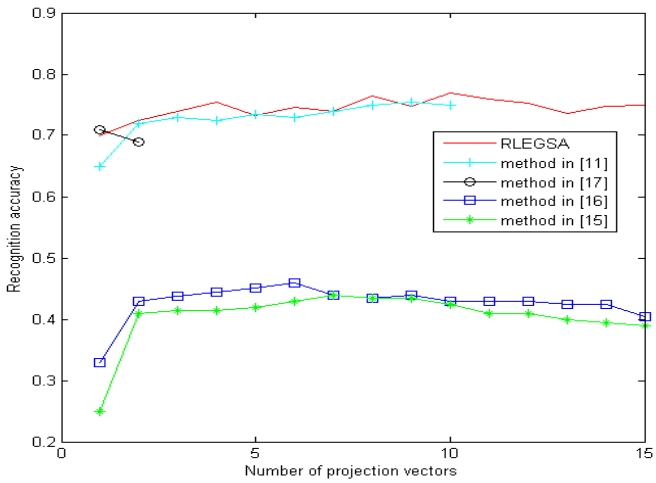

Fig. 3.7. A comparison of recognition accuracy vs. number of projection vectors for five competing methods, for ORL dataset.

3.5 Conclusions

In this Chapter, a new approach to solve SVD based face recognition problems involving single training image per person is proposed using stochastic optimization approaches. The problem is solved using GSA, a contemporary algorithm recently proposed under the category of heuristic optimization methods, which attempts to determine an optimal transform matrix W such that a cost functional $J(W)$ is minimized. In this context two new variants of GSA, called the 2-D version of GSA, and a 2-D randomized local extrema based GSA (RLEGSA) have also been proposed. Then further sophistication is incorporated in this approach, by automating the process of selection of projection vectors to propose another new variant called modified RLEGSA (MRLEGSA). Several experiments carried out for two benchmark databases i.e. Yale A and ORL databases, demonstrated that both RLEGSA and MRLEGSA produced comparable recognition accuracies and could outperform several state-of-the-art algorithms already produced for the identical problem genre. In one hand this proved the utility of proposing the random local extrema based version of GSA for this class of face recognition problems. At the same time it was aptly demonstrated that the process of automation proposed in MRLEGSA could significantly reduce the computational effort involved in RLEGSA to determine the top accuracy and yet it was able to approach the top accuracy obtained with RLEGSA.

References

[1] A. K. Jain and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. Circuits and Systems for video Technology*, vol. 14, no. 1, pp. 4-20, Jan 2004.

[2] 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.

[3] R. Jafri and H. R. Arabnia, "A Survey of Face Recognition Techniques," *Journal of Information Processing Systems*, vol. 5, no. 2, pp. 41-68, June 2009.

[4] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[5] C.J. Tu, L.Y. Chuang, J.Y. Chang, and C.H. Yang, "Feature Selection using PSO-SVM," *International Journal of Computer Science* (*IAENG)*, vol. 33, no. 1, IJCS_33_1_18, 2007.

[6] Q. Tian, Y. Fainman, and Sing H. Lee, "Comparison of statistical pattern-recognition algorithms for hybrid processing II. Eigenvector-based algorithm," *Journal of the Optical Society of America A.*, vol. 5, no. 10, pp. 1670–1682, October 1988.

[7] Z.Q. Hong, and J.Y. Yang, "Optimal discriminant plane for a small number of samples and design method of classifier on the plane," *Pattern Recognition*, vol. 24, no. 4, pp. 317–324, 1991.

[8] H. Yu, and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34, pp 2067–2070, 2001.

[9] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. Neural Networks*, vol. 14, no. 1, pp. 195–200, Jan 2003.

[10] J. Ye, R. Janardan, and Q. Li, "Two-dimensional Linear Discriminant Analysis," *Proceedings of Neural Information Processing Systems (NIPS)*, Vancouver, BC, Canada, 2004.

[11] Q. Gao, L. Zhang and D. Zhang, "Face Recognition using FLDA with single training image per person", *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 726-734, November 2008.

[12] E. Rashedi, H. Nezamabadi-pour and S. Saryazdi, "GSA: A Gravitational Search Algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232-2248, June 2009.

[13] P. N. Suganthan, "Particle swarm optimizer with neighborhood operator," in *Proc. Congr. Evoutionary Computation*, Washington, DC, Jul. 1999, pp. 1958–1962.

[14] K. Das Sharma, A. Chatterjee and A. Rakshit, "A Random Spatial *lbest* PSO-Based Hybrid Strategy for Designing Adaptive Fuzzy Controllers for a Class of Nonlinear Systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1605-1612, June 2012.

[15] S.C. Chen, D.Q. Zhang, and Z.H. Zhou, "Enhanced $(PC)^2A$ for face recognition with One training image per person," *Pattern Recognition Letters*, vol. 25, no. 10, pp. 1173–1181, 2004.

[16] D. Zhang, S.C. Chen, and Z.H. Zhou, "A new face recognition method based on SVD perturbation for single example image per person," *Applied Mathematics and Computation*, vol.163, no. 2, pp. 895–907, April 2005.

[17] S. Chen, J. Liu, and Z.H. Zhou, "Making FLDA applicable to face recognition with one sample per person," *Pattern Recognition*, vol. 37, no. 1, pp. 553–1555, 2004.

[18] Yale Univ. Face Database <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>, 2002.

[19] ORL database <http://www.cam-orl.co.uk/facedatabase.html>, 2002.

[20] G. Golub, C. Loan, Matrix Computations, John Hopkins University Press, Baltimore, 1983.

[21] R. Mansouri, F. Nasseri, M. Khorrami, "Effective time variation of G in a model universe with variable space dimension," Physics Letters A, vol. 259, no. 3-4, pp. 194-200, August 1999.

[22] Code for [11] < http://www4.comp.polyu.edu.hk/~cslzhang/code.htm >, 2008.

CHAPTER 4

Emotion Recognition with 2-D DCT and Quantum-behaved GSA

4.1 Introduction

Emotion Recognition refers to the problem of inferring the significance of human facial expressions of different emotions [24], [25]. This inference is natural for human observers but is a non trivial problem for machines. Proper classifications of different human emotions that can be clearly expressed on the human face serves two principal critical purposes [26]. Firstly emotion recognition helps the human or machine observer to understand the current mental state of the subject. In this case the observer actively seeks to extract information from the facial emotions of the subject induced naturally by the subject's immediate environment. Secondly, the subject may deliberately use his facial expressions to actively send information to the observer and thus trigger the observer to initiate a response.

These applications, when using a machine based automated, observer system, require intelligent robust classification of emotions and take appropriate actions based on the inferences drawn. Thus this problem is basically associated with HCI (Human Computer Interaction) systems which have gained immense popularity and applicability in recent times. One of the numerous applications of emotion recognition in HCI systems can be minimum disturbance fatigue monitoring system of drivers [1], [2]. Facial emotions are only one category of human interactive inputs that an HCI system may acquire and process. Other human inputs may be touch, voice signals, bioelectrical signals, biometric parameters, etc [27]. Human Computer Interactions, if

properly implemented, make the process more intuitive and dynamic and thus possess immense potential in designing more natural and user friendly systems.

The present Chapter is focused in proposing an efficient system that can be used for automated, intelligent emotion recognition purpose. It falls mainly in the highly interdisciplinary fields of image processing, computer vision, machine learning and pattern recognition. Since emotion recognition is essentially a classification problem, supervised learning based methods are traditionally applied to solve such a problem. A comprehensive survey of different established emotion recognition techniques is provided in [3], [4]. However, depending on the nature of the features extracted, the processes can be of mainly two types. The first approach deals with local features concentrating on those facial regions (usually the lips and the eyes), which are more expressive to emotion variation [5], [6]. Thus the features extracted in this approach generally have higher discriminating characteristics but require involved feature extraction procedures (like contour detections of lips and/or eyes, etc). The second approach concerns global features (like DCT or DWT features, etc) [7] or simple local patterns (like LBP, LGP, MCT, etc) [8] extracted from the entire face and thus can, in most cases, be extracted with lesser computational burden. However the main drawback of this approach is that some regions of the human face are practically insensitive to variations of emotion (like the upper region of the forehead, the chin, etc). Hence many of the features extracted may not have sufficient discriminatory characteristics and may inhibit robust classification. Moreover, generally the size of the feature set generated by global methods is very large and may be of the order of the image size itself. The effects of these drawbacks can be effectively minimized by using a proper feature selection/reduction phase by minimizing a suitable cost function in order to generate a highly discriminating subset from the original feature set. To achieve this, a viable option can be to use a metaheuristic non gradient

based iterative optimization algorithm. Binary versions of PSO have been earlier employed for feature selection purpose in different problems [9].

In this Chapter, the 2-D Discrete Cosine Transform (DCT) [10] is used as the global feature extracting technique and a novel modified binary Quantum-behaved GSA algorithm with Differential Mutation is proposed for the feature selection phase. The DCT is a feature extracting scheme that exploits the energy spectrum of the image corresponding to different frequencies but, unlike the Fourier transform, operates in the real domain. The low frequency components contain the bulk of the energy and information of the image whereas the high frequency components are responsible mainly for the variations and hence are more useful for discriminating different images for classification purposes. The gravitational search algorithm (GSA) [11] utilizes Newton's Laws (gravity and motion) and has gained wide popularity in recent times in different applications. Recently a quantum behaved modification of GSA [12] has been proposed, as an improvement of GSA for multidimensional optimization purposes. The present Chapter proposes to further enhance the quantum behaved GSA by incorporating the differential mutation methodology [13] in the quantum behaved GSA and has utilized this modified form of GSA for feature selection purpose. A proper binarization strategy is also suitably incorporated within this algorithm as the feature selection problem here has been configured essentially as a binary decision problem. The proposed algorithm is named the Modified Binary Quantum GSA with Differential Mutation (MBQGSA-DM). The algorithm finally performs the classification task, based on the extracted features, by utilizing the popular back propagation based ANN algorithm. The proposed method is evaluated on the standard JAFFE database [14] and also on facial expression database developed at the ETCE Department of Jadavpur University [6]. A detailed comparison with several state-of-the-art, competing

algorithms that exist for similar emotion recognition problems comprehensively prove that our algorithm could significantly outperform these algorithms.

The rest of the Chapter is organized as follows. In Section 4.2 an overview of the discrete cosine transform is presented. In Section 4.3, the proposed MBQGSA-DM is described in detail along with a brief description of the traditional GSA and the quantum behaved GSA. Section 4.4 contains a brief description of the classification strategy employed. The detailed experimental results and comparisons with competing methods are provided in Section 4.5. The Chapter is concluded in Section 4.6.

4.2 Feature Extraction using 2-D Discrete Cosine Transform (2-D DCT)

As mentioned before, in this Chapter the 2-D version of the DCT has been employed as the feature extraction scheme to generate features in a transformed domain from each input image. DCT is known to exhibit good discriminatory properties and low computation cost. Also DCT offers more energy compaction than Discrete Fourier Transform (DFT) and it operates in real domain [10]. These advantages have contributed to the popularity of DCT as a preferred feature extraction scheme.

For processing images 2-D version of DCT is employed. For the rest of this Chapter, the term DCT will refer necessarily to the 2-D DCT. The DCT of an NxM image I with intensity $f(x, y)$ at co-ordinates (x, y) is given by [10]:

$$F(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}\{\cos\left[\frac{\pi.u}{2.N}(2x+1)\right].\cos\left[\frac{\pi.v}{2.M}(2y+1)\right].f(x,y)\} \quad (4.1)$$

u =1, 2,…,N-1, v=1, 2,…,M-1, x =1, 2,…,N-1, y=1, 2,…,M-1

where

$$\alpha(u), \alpha(v) = \begin{matrix} \sqrt{\frac{1}{N}} & for\ u, v = 1 \\ \sqrt{\frac{2}{N}} & for\ u, v \neq 1 \end{matrix} \tag{4.2}$$

The DCT feature matrix hence generated will have the same size as the original image which may induce a considerable computational cost. Moreover, all the features may not have strong discriminatory characteristics. Thus a robust feature selection scheme is employed next to generate a highly discriminatory subset of features. To achieve this, the proposed modification of GSA is utilized as explained next.

4.3 Feature Selection using MBQGSA-DM

The Gravitational Search Algorithm [11] is a recent popular metaheuristic non gradient based optimization method that has been suitably modified and incorporated in the present Chapter to achieve proper feature selection based on minimizing a suitable cost function. The traditional GSA has been described previously in Chapter 2 in subsection 2.3.1 along with the relevant flow chart and algorithm. The original quantum behaved GSA [12] is presented in Section 4.3.1. Section 4.3.2 contains a detailed description of the modification proposed in this Chapter that is named as modified binary quantum GSA with differential mutation (MBQGSA-DM).

4.3.1 Quantum-behaved Gravitational Search Algorithm (QGSA)

The quantum-behaved GSA (QGSA) introduced by Moghadam et al. [12] incorporated the basic concepts of quantum mechanics into the traditional GSA. The uncertainty principle dictates that the position and velocity of a particle cannot be accurately measured simultaneously and lower the mass of the particle higher is this inaccuracy. This is the basis of quantum-behaved GSA and the main governing equation is the general time-dependent Schrödinger equation.

Inspired by this concept provided in [12], the final position update equation can be derived as:

$$\begin{cases} X_i(t+1) = Xbest_i(t) + \lambda.|Xbest_i(t) - X_i(t)|.\ln\left(\frac{1}{rand}\right) \; if \; R \geq 0.5 \\ X_i(t+1) = Xbest_i(t) - \lambda.|Xbest_i(t) - X_i(t)|.\ln\left(\frac{1}{rand}\right) \; if \; R < 0.5 \end{cases} \tag{4.3}$$

where *rand* and *R* are uniformly distributed random numbers in [0, 1] and λ is the expansion-contraction coefficient. Here $Xbest_i(t)$ is computed using the following equation:

$$Xbest_i(t) = \frac{c_1 Xmbest_i(t) + c_2 Xpbest_i(t)}{c_1 + c_2} \tag{4.4}$$

where

$$Xmbest_i(t) = \frac{\sum_{j=1}^{K} \frac{1}{\left\| X_i(t), Xkbest_j(t) \right\|} Xkbest_j(t)}{\sum_{j=1}^{K} \frac{1}{\left\| X_i(t), Xkbest_j(t) \right\|}} \tag{4.5}$$

Here $Xpbest_i(t)$ is the position with best fitness of *i*th particle up to iteration *t* and $Xkbest_j(t)$ is the position of the *j*th particle belonging to the set of *K* particles having the best fitness at iteration *t*.

4.3.2 Modified Binary Quantum-behaved GSA with Differential Mutation (MBQGSA-DM)

The quantum behaved PSO with differential mutation was proposed by Jamalipour et al. in [13] and it adopts the mutation operation (defined by Lu et al. in [15]) within the basic Quantum-behaved PSO framework as follows:

$$X_i(t+1) = X_j(t) + (1 - F) \times [X_l(t) - X_m(t)] + F \times [Xgbest(t) - X_j(t)] \tag{4.6}$$

where *i*, *j*, *l*, *m* are random integers uniformly selected from the set {1, 2, …, *N*} and $i \neq j \neq l \neq m$.

The Differential Mutation factor *F* is defined as

$$F = \frac{t}{T} \tag{4.7}$$

where t is the current iteration and T is the total number of iterations. Here $Xgbest(t)$ is the position of the particle with the best fitness at iteration t.

A further modification is proposed in the algorithm in the computation of $Xmbest_i(t)$ in equation (4.8). This was earlier computed as in equation (4.5) as shown in Section 4.3.1. The intuition behind the earlier formulation of $Xmbest_i(t)$ is to use it as a measure of the mean finess while incorporating the concept of best position by taking the best K particles. An alternative formulation is proposed to calculate the position of $Xmbest_i(t)$ as an average of positions of all particles with respect to the ith particle as defined below:

$$Xmbest_i(t) = \frac{\sum_{j=1}^{p} \frac{1}{\left\| X_i(t), X_j(t) \right\|} X_j(t)}{\sum_{j=1}^{p} \frac{1}{\left\| X_i(t), X_j(t) \right\|}} \ , \ j \neq i \tag{4.8}$$

In addition, the effect of best position is retained by incorporating $Xgbest\ (t)$ in equation (4.4) as follows:

$$Xbest_i(t) = \frac{c_1 Xmbest_i(t) + c_2 Xpbest_i(t) + c_3 Xgbest(t)}{c_1 + c_2 + c_3} \tag{4.9}$$

A random number r is chosen uniformly from [0, 1]. The equation (4.3) is used for position update if $r < 0.5$, and, otherwise, the equation (4.6) is used for position update.

Assume that for each image, n features were extracted in the first stage and let the n dimensions of the position vector of each particle represent the selection variables corresponding to those features. Hence to take the necessarily binary decision of selecting or not selecting a feature based on the value of the corresponding selection variable, a binarization of the GSA algorithm is necessary. Here a simple binarization as expressed below suffices.

$$x_i^d(t+1) = \begin{cases} 1, & \frac{1}{1+e^{-x_i^d(t+1)}} > rand \\ 0, & otherwise \end{cases} \tag{4.10}$$

Here *rand* is an uniformly distributed random number in the interval [0, 1]. At the end of each iteration, the position vector for each particle *i* is carefully scanned for each of its dimension. Then, for particle *i*, a feature is chosen from the complete list of extracted features, if the corresponding dimension of its position vector possesses a value of 1 or rejected, if the corresponding value is 0. Hence the number of features selected for each particle *i* corresponds to the number of entries of 1 in the corresponding position vector.

Obviously at the start of the algorithm the position vectors of the particles need to be randomly initialized with values in {0, 1}. Suppose at iteration *t,* for the particle *i*, out of the original *n* features, *m* features have been selected. Fitness or goodness function of the particle *i* at iteration *t* is defined as the ratio of the intra class distance or within class distance ($Dintra_i^t$) to the interclass distance or between class distance ($Dinter_i^t$).

$$fit_i^t = \frac{Dintra_i^t}{Dinter_i^t} \tag{4.11}$$

where $Dintra_i^t$ is the sum of the Euclidean distances between the selected feature sets (for the *i*th particle at iteration *t*) of the images belonging to same class.

$$Dintra_i^t = \sum_c \sum_{u,r \in c, u \neq r} \left\| F_{u,i}^t - F_{r,i}^t \right\| \tag{4.12}$$

$F_{u,i}^t$ is the feature set of the *u*th image in the *c*th class corresponding to the position (selection) vector of the *i*th particle at iteration *t*. Similarly $Dinter_i^t$ is the sum of the Euclidean distances between the selected feature sets (for the *i*th particle at iteration *t*) of the images belonging to different classes.

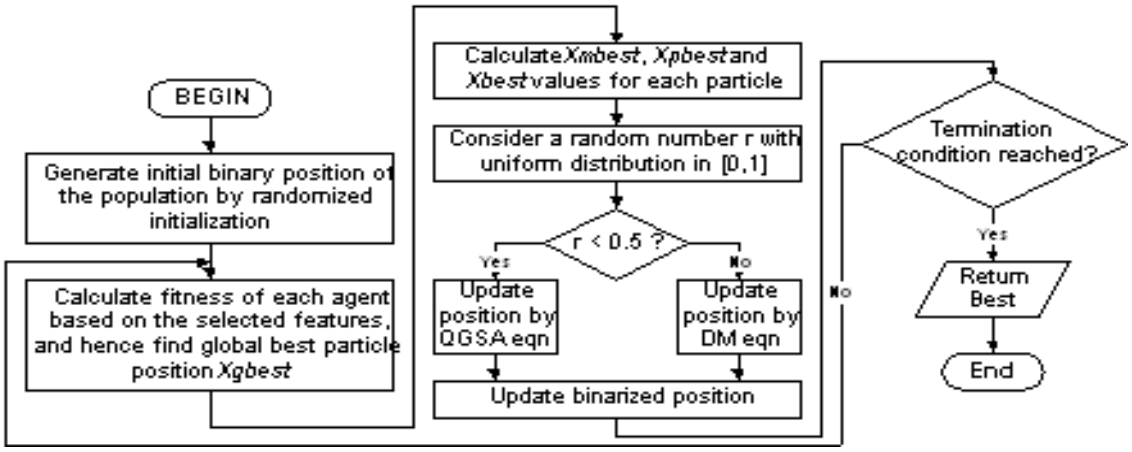

Fig. 4.1. Flow chart for the proposed MBQGSA-DM.

$$Dinter_i^t = \sum_{c,c'} \sum_{u \in c, r \in c', c \neq c'} \left\| F_{u,i}^t - F_{r,i}^t \right\| \tag{4.13}$$

The Algorithm 4.1 describes the implementation of the proposed MBQGSA-DM. Figure 4.1 presents the same algorithm in flow chart form.

4.4 Neural Network based Classification

For the final classification phase, the same methodology as adopted in Chapter 2 has been successfully utilized in the present Chapter also. The details have been previously discussed in Section 2.4.

Algorithm 4.1: MBQGSA-DM

BEGIN

Create p particles and make randomized initialization of their n dimensional positions X in $\{0, 1\}$. Total number of iterations = T.

Initialize iteration number $t = 1$

REPEAT:

FOR $i = 1$ to p

Calculate fitness $fit_i^t = \frac{Dintra_i^t}{Dinter_i^t}$

where $Dintra_i^t = \sum_c \sum_{u,r \in c, u \neq r} \left\| F_{u,i}^t - F_{r,i}^t \right\|$

and $Dinter_i^t = \sum_{c,c'} \sum_{u \in c, r \in c', c \neq c'} \left\| F_{u,i}^t - F_{r,i}^t \right\|$

END FOR

Calculate $Xgbest(t)$

FOR $i = 1$ to p

Calculate $Xpbest_i(t)$

Calculate $Xmbest_i(t) = \frac{\sum_{j=1}^{p} \frac{1}{\left\| X_i(t), X_j(t) \right\|} X_j(t)}{\sum_{j=1}^{p} \frac{1}{\left\| X_i(t), X_j(t) \right\|}}$

Calculate $Xbest_i(t) = \frac{c_1 Xmbest_i(t) + c_2 Xpbest_i(t) + c_3 Xgbest(t)}{c_1 + c_2 + c_3}$

Consider uniform random numbers $r \in [0, 1]$ and $R \in [0, 1]$

IF $r < 0.5$

$$\begin{cases} X_i(t+1) = Xbest_i(t) + \lambda . |Xbest_i(t) - X_i(t)| . \ln\left(\frac{1}{rand}\right) & if\ R \geq 0.5 \\ X_i(t+1) = Xbest_i(t) - \lambda . |Xbest_i(t) - X_i(t)| . \ln\left(\frac{1}{rand}\right) & if\ R < 0.5 \end{cases}$$

ELSE

$X_i(t+1) = X_j(t) + (1 - F) \times [X_l(t) - X_m(t)] + F \times [Xgbest(t) - X_j(t)]$

where Differential Mutation constant $F = \frac{t}{T}$

END IF

FOR $d = 1$ to n

$$x_i^d(t+1) = \begin{cases} 1, & \frac{1}{1 + e^{-x_i^d(t+1)}} > rand \\ 0, & otherwise \end{cases}$$

END FOR

END FOR

UNTIL $t = T$

END

4.5 Experiments and Results

The proposed method using the 2-D DCT for feature extraction, the MBQGSA-DM for feature selection, and back propagation based ANN for final classification has been evaluated on the standard JAFFE database [14] and the results indicate that the proposed strategy outperforms several recently proposed, state-of-the-art, competing techniques, utilized for solving similar emotion recognition problems.

The JAFFE (JApanese Female Facial Expression) Database was developed at the Psychology Department of Kyushu University in Japan in 1997. The database consists of 213 images of 10 Japanese female subjects with seven facial expressions each, six different annotated emotions (Anger/Annoyed, Disgust, Surprise, Fear, Happy, and Sad) and one Neutral expression. Figure 4.2 shows the samples of seven different expressions for one subject from the JAFFE Database. For each expression of each subject mostly there are three images and in a few cases there are four images. Hence, in the experiments, three images per expression (with a total of seven expressions) per subject (total 10 subjects) have been used, which constitutes a total of 210 images under consideration from the database.

The experimental results for the emotion recognition problem are provided in table 4.1 along with the results obtained for the same database using recently proposed, state-of-the-art, competing algorithms, as mentioned before. These methods are chosen for comparison because they follow a similar framework of feature extraction, feature selection and classification instead of a regional geometry based approach and hence are more suitable for comparison with the proposed algorithm.

Fig. 4.2. Seven different expressions for one subject from the JAFFE Database [14].

The methods have minor differences in experimental conditions but the results provided give a general idea of the recognition rates they are capable of achieving. All the results of the competing algorithms are obtained from [4]. It can be seen that our proposed method could achieve a recognition rate higher than 97% which all other competing algorithms reported could achieve recognition rate less than 96% only. This aptly demonstrates the utility of the proposed algorithm.

Table 4.1. Comparison of experimental results using JAFFE database

Method	**Average Recognition Rates (%)**	**Features**
Zhang et al. (1998) [16]	90.1	Geometry and Gabor
Bashyal and Venayagamoorthy (2008) [17]	90.2	Gabor and LVQ
Koutlas and Fotiadis (2008) [18]	92.3	Gabor filters
Liu and Wang (2006) [19]	92.5	Gabor filters
Oliveira et al. (2011) [20]	94.0	2D PCA with feature selection and SVM
Liao et al. (2006) [21]	94.5	LBP
Cheng et al. (2010) [22]	95.2	Gaussian Process
Zhi and Ruan (2008) [23]	95.9	2D LPP (Locality Preserving Projections)
Proposed Method	**97.2**	**2-D DCT + MBQGSA-DM + ANN**

Fig. 4.3. Five different expressions for one subject from the Jadavpur University Database (images courtesy Anisha Halder and Amit Konar, Robot Vision Lab, ETCE Department of Jadavpur University, Kolkata, India).

Zhang et al. showed that the generalized feature extraction plus selection strategy can outperform geometry based regional features specially formulated for emotion recognition. For further validation of this observation, experiments have also been performed on the facial emotion database developed by the Robot Vision Lab in the ETCE Department of Jadavpur University, Kolkata, India [6]. This database consists of five different emotions (Anger, Happy, Disgust, Fear, and Sad) and one Neutral expression with 10 instances of each of the 6 expressions per person. The different emotions of one sample subject in this database are shown in fig. 4.3.

The results reported in [6] by the research group which developed this database at Jadavpur University showed an average accuracy of 85.11% whereas the method proposed in this Chapter could achieve an average accuracy of 89.63%. This experimentation further validates the usefulness of the proposed method. Also it should be noted that this database instead of only providing images focused only on the facial region, includes a fair amount of background scenes

as well (as evident in fig. 4.3), thus making the database a challenging one. This may have played a significant role in the outcome of a lower recognition rate compared to the results obtained for the JAFFE database.

4.6 Conclusion

The present Chapter demonstrated how a computationally intensive and mathematically involved local feature extraction strategy for the emotion recognition problem can be effectively replaced by a much simpler global feature extraction strategy like the DCT if a proper feature selection phase is used in conjunction with it. The recently proposed quantum behaved GSA has been suitably modified and adopted in this Chapter to achieve the desired objective. The proposed MBQGSA-DM algorithm minimizes the ratio of within class to between class distances and selects an optimized subset of features with high discriminative characteristics. The experimental results clearly supports the effectiveness of this strategy as the proposed algorithm has been evaluated on two separate databases one of which (JAFFE) is already an established and popular standard used frequently in this field. Also the proposed algorithm is much more generic than extraction of a special feature sensitive to emotion variation (like lip or eye contour). This is because, keeping the basic framework same, other alternative methods can be used in the three stages. For example, the global techniques like the DWT or simple local pattern like LBP, MCT or LGP can be utilized in the feature extraction phase. Other optimization techniques like the binary PSO or some other form of evolutionary computing based technique can be adopted in the feature selection stage. Lastly, other variants of ANN or modern non linear classifiers like support vector machines (SVMs), learning vector quantizers (LVQs) etc. can be utilized for the final classification purposes. The author intends to pursue these potent research directions for the emotion recognition problems in near future.

References

[1] A. Kolli, A. Fasih, F. Al Machot, and K. Kyamakya, "Non-intrusive car driver's emotion recognition using thermal camera," *Proc. IEEE Joint Third International workshop on Nonlinear Dynamics and Synchronization and Sixteenth International Symposium on Theoretical Electrical Engineering*, pp. 1-5, July 2011.

[2] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," *Proc. IEEE Conference on Instrumentation and Measurement Technology*, pp. 1-4, May 2011.

[3] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, Jan. 2003.

[4] S. Shinde and S. Pande, "A Survey on emotion recognition with respect to database and various recognition techniques," *International Journal of Computer Applications*, vol. 58, no. 3, Nov. 2012.

[5] A. Youssif and W. Asker, "Automatic Facial Expression Recognition System based on Geometric and Appearance Features," *Computer and Information Science*, vol. 4, no. 2, March 2011.

[6] A. Halder, S. Shaw, K. Orea, P. Bhowmik, A. Chakraborty and A. Konar, "A computational intelligence approach to emotion recognition from the Lip-Contour of a subject," *book chapter* in A. Chatterjee and P. Siarry (eds.), *Computational Intelligence in Image Processing*, DOI: 10.1007/978-3-642-30621-1_15, pp. 281-298, Springer-Verlag Berlin Heidelberg, 2013.

[7] F. Y. Shih, C-F Chuang, and P. S. P. Wang, “Performance comparisons of facial expression recognition in JAFFE database,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 22, no. 3, pp. 445-459, 2008.

[8] B. Jun and D. Kim, “Robust face detection using local gradient patterns and evidence accumulation,” *Pattern Recognition*, vol. 45, no. 9, pp. 3304-3316, Sep. 2012.

[9] C.-J. Tu, L.-Y. Chuang, *J.-Y. Chang,* and C.-H. Yang, “Feature Selection using PSO-SVM,” *International Journal of Computer Science* (*IAENG)*, vol. 33, no. 1, Feb. 2007.

[10] R. M. Ramadan and 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, pp. 51-66, June 2009.

[11] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, “GSA: A Gravitational Search Algorithm,” *Information Sciences*, vol. 179, no. 13, pp. 2232-2248, June 2009.

[12] M. S. Moghadam, H. Nezamabadi-Pour, and M. M. Farsangi, “A Quantum Behaved Gravitational Search Algorithm,” *Intelligent Information Management*, vol. 4, pp. 390-395, 2012.

[13] M. Jamalipour, R. Sayareh, M. Gharib, F. Khoshahval, and M. R. Karimi, “Quantum behaved Particle Swarm Optimization with Differential Mutation operator applied to WWER-1000 in-core fuel management optimization,” *Annals of Nuclear Energy*, vol. 54, pp. 134-140, April 2013.

[14] *Michael J. Lyons, Miyuki Kamachi, Jiro Gyoba: Japanese Female Facial Expressions (JAFFE), Database of digital images (1997).*

[15] S. Lu, C. Sun, and Z. Lu, “An improved quantum-behaved particle swarm optimization method for short-term combined economic emission hydrothermal scheduling,” Energy Conversion and Management, vol. 51, no. 3, pp. 561-571, March 2010.

[16] Z. Zhang, M. J. Lyons, M. Schuster, and S. Akamatsu, “Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron,” *Proc. Third IEEE international conference on automatic face and gesture recognition*, pp. 454–459, April 1998.

[17] S. Bashyal and G. K. Venayagamoorthy, “Recognition of facial expressions using Gabor wavelets and learning vector quantization,” *Engineering Applications of Artificial Intelligence*, vol. 21, no. 7, pp. 1056-1064, Oct. 2008.

[18] A. Koutlas and D. I. Fotiadis, “An automatic region based methodology for facial expression recognition,” *Proc. IEEE International Conference on Systems, Man and Cybernetics*, pp. 662–666, Oct. 2008.

[19] W. Liu and Z. Wang, “Facial expression recognition based on fusion of multiple Gabor features,” *Proc. IEEE 18th International Conference on Pattern Recognition*, vol. 3, pp. 536-539, 2006.

[20] L. E. S. Oliveira, A. L. Koerich, M. Mansano, and A. S. Jr. Britto, “2-D principal component analysis for face and facial-expression recognition,” *Computing in Science and Engineering*, vol. 13, no. 3, pp. 9–13, 2011.

[21] S. Liao, W. Fan, C. S. Chung, and D.-Y. Yeung, “Facial expression recognition using advanced local binary patterns,” *Proc. IEEE International Conference on Image Processing*, pp. 665–668, Oct. 2006.

[22] F. Cheng, J. Yu, and H. Xiong, "Facial expression recognition in JAFFE dataset based on Gaussian process classification," *IEEE Trans. on Neural Networks*, vol. 21, no. 10, pp. 1685–1690, Oct. 2010.

[23] R. Zhi, and Q. Ruan, "Facial expression recognition based on two-dimensional discriminate locality preserving projections," *Neurocomputing*, vol. 71, no. 7-9, pp. 1730–1734, March 2008.

[24] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. G. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, Jan. 2001.

[25] M. Milanova and N. Sirakov, "Recognition of emotional states in natural human-computer interaction," *IEEE International Symposium on Signal Processing and Information Technology*, pp. 186-191, Dec. 2008.

[26] R. V. Yampolskiy and V. Govindaraju "Direct and indirect human computer interaction based biometrics," *Journal of Computers*, vol. 2, no. 10, pp. 76-88, Dec. 2007.

[27] R. M. Jiang, A. H. Sadka, and D. Crookes, "Multimodal Biometric Human Recognition for Perceptual Human-Computer Interaction," *IEEE Trans. on Systems, Man and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 6, pp. 676- 681, Nov. 2010.

CHAPTER 5

CONCLUSION

In this thesis several novel modifications of the Gravitational Search Algorithm have been explored and have been adopted to solve the feature selection problem in supervised learning based classification applications like face recognition and emotion recognition.

The original feature set has been extracted using a variety of generalized methodologies like DCT (global features) in Section 3 and LBP, MCT, LGP (local features) in Section 2. The feature selection phase generates a highly discriminative subset which is then used for the final classification. Experimental results show that a methodology incorporating proper feature selection from even a traditional generalized feature extractor has the potential to outperform a method with much more mathematically complicated specialized features (with no feature selection) formulated for particular applications. The singularity problem of single image per person for methods utilizing within class distances in the algorithm has also been addressed by employing the SVD technique in Chapter 2.

The possibility of using other alternative techniques in one or more of the three phases, namely, feature extraction, feature selection and classification indicates a wide scope of future research. Another direction of future work can be to develop a synergistic merging of feature extraction and selection in a single step which has been to some extent explored in the present thesis in Section 2. Again further variations are possible in the proposed methods themselves like a combination of local and global extrema based GSA as a possible modification of the algorithm proposed in Section 3.