Expert Systems

Article

DOI: 10.1111/exsy.12104

Automated emotion recognition employing a novel modified binary quantum-behaved gravitational search algorithm with differential mutation

Tapabrata Chakraborti,[1]* Amitava Chatterjee,[1] Anisha Halder[2] and Amit Konar[2]

(1) Department of Electrical Engineering, Jadavpur University, Kolkata, 700032, India
E-mail: cha_ami@yahoo.co.in
(2) Department of Electronics & Telecommunication Engineering, Jadavpur University, Kolkata, 700032, India

Abstract: *The present paper proposes a supervised learning based automated human facial emotion recognition strategy with a feature selection scheme employing a novel variation of the gravitational search algorithm (GSA). The initial feature set is generated from the facial images by using the 2-D discrete cosine transform (DCT) and then the proposed modified binary quantum GSA with differential mutation (MBQGSA-DM) is utilized to select a sub-set of features with high discriminative power. This is achieved by minimising the cost function formulated as the ratio of the within class and interclass distances. The overall system performs its final classification task based on selected feature inputs, utilising a back propagation based artificial neural network (ANN). Extensive experimental evaluations are carried out utilising a standard, benchmark emotion database, that is, Japanese Female Facial Expresssion (JAFFE) database and the results clearly indicate that the proposed method outperforms several existing techniques, already known in literature for solving similar problems. Further validation has also been carried out on a facial expression database developed at Jadavpur University, Kolkata, India and the results obtained further strengthen the notion of superiority of the proposed method.*

Keywords: discrete cosine transform (DCT), gravitational search algorithm (GSA), modified binary quantum gravitational search algorithm with differential mutation (MBQGSA-DM), artificial neural network (ANN)

1. Introduction

Emotion recognition refers to the problem of inferring the significance of human facial expressions of different emotions (Cowie *et al*., 2001; Milanova & Sirakov, 2008). 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 (Yampolskiy & Govindaraju, 2007). 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 (Kolli *et al*., 2011; Abtahi *et al*., 2011). 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, and so on. (Jiang *et al*., 2010). 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 work 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. Because 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 (Fasel & Luettin, 2003; Shinde & Pande, 2012). 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 (Youssif & Asker, 2011; Halder *et al.*, 2013). 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 discrete cosine transform (DCT) or discrete wavelet transform (DWT) features, etc.) (Shih *et al.*, 2008) or simple local patterns, such as local binary pattern (LBP), local gradient pattern (LGP), and modified census transform (MCT), etc. (Jun & Kim, 2012) 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 minimising a suitable cost function in order to generate a highly discriminating sub-set from the original feature set. To achieve this, a viable option is to use a metaheuristic non-gradient-based iterative optimization algorithm. Binary versions of particle swarm optimization (PSO) have been earlier employed for feature selection purpose in different problems (Tu *et al.*, 2007).

In this paper, the 2-D DCT (Ramadan & Abdel-Kader, 2009) 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) (Rashedi *et al.*, 2009) 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 (Moghadam *et al.*, 2012) has been proposed as an improvement of GSA for multi-dimensional optimization purposes. The present paper proposes to further enhance the quantum-behaved GSA (QGSA) by incorporating the differential mutation methodology (Jamalipour *et al.*, 2013) in the QGSA 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 utilising the popular back propagation-based artificial neural network (ANN) algorithm. The proposed method is evaluated on the standard Japanese Female Facial Expression (JAFFE) database (Michael, 1997) and also on facial expression database developed at the Electronics and Telecommunication Engineering (ETCE) Department of Jadavpur University (Halder *et al.*, 2013). 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 paper is organized as follows: In Section II we present an overview of DCT. In Section III, the proposed MBQGSA-DM is described in detail, along with a brief description of the traditional GSA and the QGSA. Section IV contains a brief description of the classification strategy employed. The detailed experimental results and comparisons with competing methods are provided in Section V. The paper is concluded in Section VI.

2. Feature extraction using 2-D discrete cosine transform

As mentioned before in this work, 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 (Ramadan & Abdel-Kader, 2009). 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 work, the term DCT will refer necessarily to the 2-D DCT. The DCT of an $N \times M$ image I with intensity $f(x, y)$ at coordinates (x,y) is given as (Ramadan & Abdel-Kader, 2009)

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

where

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

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 sub-set of features. To achieve this, the proposed modification of GSA is utilized as explained next.

3. Feature selection using MBQGSA-DM

The GSA (Rashedi *et al.*, 2009; Chakraborti & Chatterjee, 2014; Chakraborti *et al.*, 2014) is a recent popular metaheuristic non-gradient-based optimization method that has been suitably modified and incorporated in the present work to achieve proper feature selection based on minimising a suitable cost function. A brief overview of the traditional GSA algorithm is presented in Section 3.1. The original QGSA (Moghadam *et al.*, 2012) is presented in Section 3.2. Section 3.3 contains a detailed description of the modification proposed in this work that is named as MBQGSA-DM.

3.1. Gravitational search algorithm

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

$$X_i(t) = \left\{x_i^1(t), x_i^2(t), \ldots, x_i^n(t)\right\} \tag{3}$$

The gravitational force experienced by a particle i because of the particle j at iteration t in the dimension d is given by the Newton's Law of gravity as follows (Rashedi *et al.*, 2009):

$$F_{ij}^d(t) = G(t)\frac{M_i(t)\times M_j(t)}{R_{ij}(t)+\varepsilon}\left(x_j^d(t) - x_i^d(t)\right) \tag{4}$$

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.

$$R_{ij}(t) = X_i(t), X_j(t) \tag{5}$$

$$G(t) = G(t_0)\times\left(\frac{t_0}{t}\right)^{\beta},\ \beta < 1 \tag{6}$$

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 towards a solution. The total gravitational force on a particle i in the dth dimension is (Rashedi *et al.*, 2009)as follows:

$$F_i^d(t) = \sum_{j=1, j\neq i}^{p} rand_j F_{ij}^d(t) \tag{7}$$

Where $rand_j$ is a 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

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} \tag{8}$$

The velocity, position and mass update equations are (Rashedi *et al.*, 2009) as follows:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{9}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{10}$$

$$m_i(t+1) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{11}$$

$$M_i(t+1) = \frac{m_i(t+1)}{\sum_{j=1}^{p} m_j(t+1)} \tag{12}$$

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 amongst the p particles at iteration t respectively, are given as

$$best(t) = \min_{j\in\{1,2,\ldots p\}} fit_j(t) \tag{13}$$

$$worst(t) = \max_{j\in\{1,2,\ldots p\}} fit_j(t) \tag{14}$$

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, the 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 because unlike PSO, there is no need to keep record of the best position of each particle and also the globally best position.

3.2. Quantum-behaved gravitational search algorithm

The QGSA, introduced by Moghadam *et al.* (Moghadam *et al.*, 2012), 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 QGSA and the main governing equation is the general time-dependent Schrödinger equation.

Inspired by this concept provided in (Moghadam *et al.*, 2012), the final position update equation can be derived as computed as in equation (17), as shown in Section 3.2. The intuition behind the earlier formulation of $Xmbest_i(t)$ is to

$$\begin{cases} X_i(t+1) = Xbest_i(t) + \lambda.|Xbest_i(t) - X_i(t)|.\ln\left(\frac{1}{rand}\right) \; if\ R \geq 0.5 \\ X_i(t+1) = Xbest_i(t) - \lambda.|Xbest_i(t) - X_i(t)|.\ln\left(\frac{1}{rand}\right) \; if\ R < 0.5 \end{cases} \tag{15}$$

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:

$$Xbest_i(t) = \frac{c_1 Xmbest_i(t) + c_2 Xpbest_i(t)}{c_1 + c_2} \tag{16}$$

where

$$Xmbest_i(t) = \frac{\sum_{j=1}^{K} \frac{1}{X_i(t), Xkbest_j(t)} Xkbest_j(t)}{\sum_{j=1}^{K} \frac{1}{X_i(t), Xkbest_j(t)}} \tag{17}$$

Here, $Xpbest_i(t)$ is the position with best fitness of ith particle up to iteration t and $Xkbest_j(t)$ is the position of the jth particle belonging to the set of K particles having the best fitness at iteration t.

3.3. *Modified binary quantum gravitational search algorithm with differential mutation*

The quantum-behaved PSO with differential mutation was proposed by Jamalipour *et al.* and in adopts the mutation operation (defined by Lu et al., 2010) within the basic quantum-behaved PSO framework as follows:

$$X_i(t+1) = X_j(t) + (1-F) \times [X_l(t) - X_m(t)] + F \times \left[Xgbest(t) - X_j(t)\right] \tag{18}$$

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

$$F = \frac{t}{T} \tag{19}$$

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 (20). This was earlier computed as in equation (17), as shown in Section 3.2. The intuition behind the earlier formulation of $Xmbest_i(t)$ is to use it as a measure of the mean fitness whilst 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 in the following equation:

$$Xmbest_i(t) = \frac{\sum_{j=1}^{p} \frac{1}{X_i(t), X_j(t)} X_j(t)}{\sum_{j=1}^{p} \frac{1}{X_i(t), X_j(t)}}, \; j \neq i \tag{20}$$

In addition, the effect of best position is retained by incorporating $Xgbest(t)$ in equation (16) as follows:

$$Xbest_i(t) = \frac{c_1 Xmbest_i(t) + c_2 Xpbest_i(t) + c_3 Xgbest(t)}{c_1 + c_2 + c_3} \tag{21}$$

A random number r is chosen uniformly from [0,1]. Equation (15) is used for position update if $r < 0.5$; otherwise, equation (18) is used for position update.

Let us 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 necessary 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 in the following equation, suffices.

$$x_i^d(t+1) = \begin{cases} 1, & \frac{1}{1 + e^{-x_i^d(t+1)}} > rand \\ 0, & otherwise \end{cases} \tag{22}$$

Here, *rand* is a 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 and 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 intraclass distance or within class distance ($Dintra_i^t$) to the interclass distance or between class distance ($Dinter_i^t$).

$$fit_i^t = \frac{Dintra_i^t}{Dinter_i^t} \tag{23}$$

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

Algorithm 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 $Dintra_i^t = \sum_c \sum_{u,r\in c,\ u\neq r} ||F^t_{u,i} - F^t_{r,i}||$
    and $Dinter_i^t = \sum_{c,c`} \sum_{u\in c, r\in c`, c\neq c`} ||F^t_{u,i} - F^t_{r,i}||$
  END FOR
  Calculate $Xgbest(t)$
  FOR i = 1 to p
    Calculate $Xpbest_i$(t)
    Calculate $\frac{Xmbest_i(t) = \sum_{j=1}^{p} \frac{1}{||X_i(t),X_j(t)||} X_j(t)}{\sum_{j=1}^{p} \frac{1}{||X_i(t),X_j(t)||}}$
    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

$$Dintra_i^t = \sum_c \sum_{u,r\in c,\ u\neq r} ||F^t_{u,i} - F^t_{r,i}|| \tag{24}$$

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

$$Dinter_i^t = \sum_{c,c`} \sum_{u\in c, r\in c`, c\neq c`} ||F^t_{u,i} - F^t_{r,i}|| \tag{25}$$

Algorithm 1 describes the implementation of the proposed MBQGSA-DM. Figure 1 presents the same algorithm in flow chart form.

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 feed-forward architecture, still remain as one of 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 back propagation, conjugate gradient Fletcher Powell-based BPNN algorithm and so on. In this work, we have utilized a three-layered architecture with 20 hidden layer neurons and the LM learning-based BPNN is employed for training exemplars from the training set.

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 (Michael, 1997) 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 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 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 we have four images. Hence, in our 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

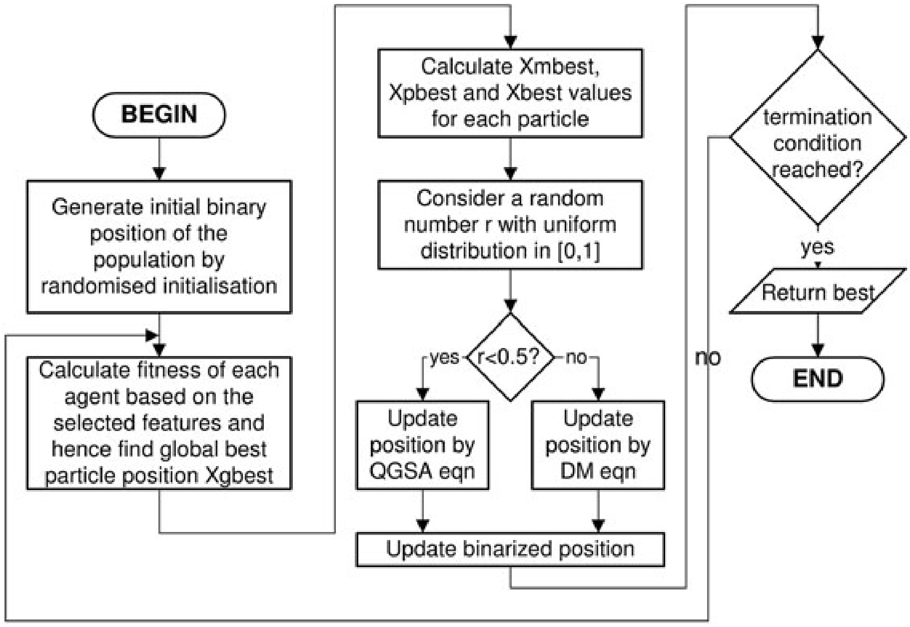

Figure 1: *Flow chart for the proposed modified binary quantum gravitational search algorithm with differential mutation.*

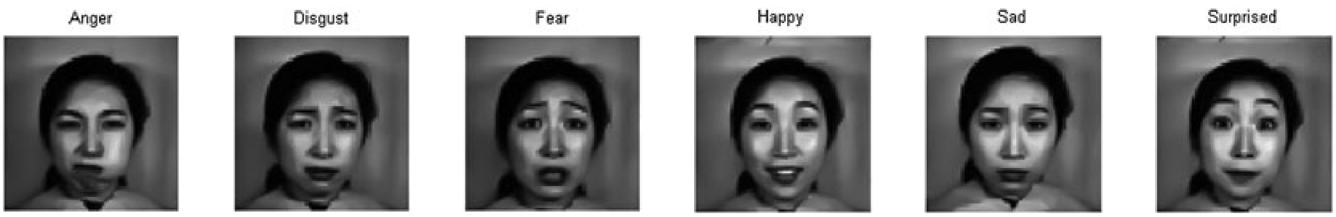

Figure 2: *Seven different expressions for one subject from the Japanese Female Facial Expression Database (Michael, 1997).*

the database. Thus, for each expression per subject, number of images used is three, out of which two are randomly chosen as training images and one for testing. Hence, in the training dataset, there are 140 images for 10 subjects (with seven expressions/emotions per subject). In the testing database, there are 70 images for 10 subjects (seven emotions per subject).

The experimental results for the emotion recognition problem are provided in Table 1, along with the results obtained for the same database using recently proposed state-of-the-art competing algorithms, as mentioned earlier. These methods are chosen for comparison because they follow a similar framework of feature extraction, feature selection and classification instead of a regional geometry-

Table 1: *Comparison of experimental results using Japanese Female Facial Expression database*

Method	Average recognition rates (%)	Features
Zhang *et al*. (1998)	90.1	Geometry and Gabor
Bashyal and Venayagamoorthy (2008)	90.2	Gabor and LVQ
Koutlas and Fotiadis (2008)	92.3	Gabor filters
Liu and Wang (2006)	92.5	Gabor filters
Oliveira *et al*. (2011)	94.0	2-D PCA with feature selection and SVM
Liao *et al*. (2006)	94.5	LBP
Cheng *et al*. (2010)	95.2	Gaussian Process
Zhi and Ruan (2008)	95.9	2-D LPP (Locality Preserving Projections)
Our Method	97.143	2-D DCT + MBQGSA-DM + ANN

LVQ, learning vector quantizer; SVM, support vector machine; DCT, discrete cosine transform; MBQGSA-DM, modified binary quantum gravitational search algorithm with differential mutation; ANN, artificial neural network.

based approach and hence, are more suitable for comparison with our proposed algorithm. The methods have minor differences in experimental conditions but the results provided gives a general idea of the recognition rates they are capable of achieving. All the results of the competing algorithms are obtained from (Shinde & Pande, 2012). 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.

To provide a detailed analysis of our results, the recognition rates per emotion are also reported in Table 2, with detailed data in form of confusion data. The corresponding abbreviations used are as follows: TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative. It should be noted here that in the testing phase, there is a single image per emotion (a total of seven emotions) per person (a total of 10 persons). Thus, there are 10 images per emotion. So there are 70 images total in the testing dataset.

Several interesting conclusions can be drawn from the previously mentioned Table 1. Out of the 10 images, for emotion 'Disgust', two images were falsely identified as 'Sad' and vice versa, thereby signifying certain confusion between the two if these are not expressed clearly enough in the faces. Again in one case, the emotion of 'Fear' was not emphasized enough in the facial image and it was wrongly classified as 'Surprised'. Similarly, the emotion 'Surprised' was also sometimes too subtly registered in faces because it was once misclassified as 'Fear' and was once misclassified as 'Neutral'. However, in overwhelmingly most of the cases, the overall recognition was correct enough to achieve an excellent overall recognition accuracy of 97.143%.

Table 2: *Detailed analysis of our recognition rates per emotion for Japanese Female Facial Expression database*

Emotion	TP	TN	FP	FN	Recognition rate [(TP+TN) / (TP+TN+FP+FN)]*100
Anger	10	60	0	0	100
Disgust	8	58	2	2	94.286
Fear	9	59	1	1	97.143
Happy	10	60	0	0	100
Sad	8	58	2	2	94.286
Surprised	8	59	1	2	95.714
Neutral	10	59	1	0	98.571
Average recognition rate	—	—	—	—	97.143

TP, true positive; TN, true negative; FP, false positive; FN, false negative.

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 (Halder *et al.*, 2013). This database consists of five different emotions (Anger, Happy, Disgust, Fear and Sad) and one Neutral expression with 10 instances of each of the six expressions per person. So for each emotion per subject, there are 10 images, out of which seven have been randomly chosen for training purposes and the rest have been used for testing. The different emotions of one sample subject in this database are shown in Figure 3.

The results reported in (Halder *et al.*, 2013) by the research group that developed this database at Jadavpur University showed an average accuracy of 85.11%, whereas the method proposed in this work 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 that focused only on the facial region, includes a fair amount of background scenes as well (as evident in Figure 3), thus making the database a challenging one. This may have played a significant role in the outcome of a lower recognition rate compared with the results obtained for the JAFFE database.

Figure 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).*

6. Conclusion

The present work 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 QGSA has been suitably modified and adopted in this work to achieve the desired objective. The proposed MBQGSA-DM algorithm minimizes the ratio of within class to between class distances and selects an optimized sub-set 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 such as support vector machines (SVMs), and learning vector quantizers (LVQs) and so on can be utilized for the final classification purposes. The authors intend to pursue these potent research directions for the emotion recognition problems in near future.

References

Kolli, A., A. Fasih, F. Al Machot and K. Kyamakya (2011) 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, 1–5.

Abtahi, S., B. Hariri and S. Shirmohammadi (2011) Driver drowsiness monitoring based on yawning detection. Proc. IEEE Conference on Instrumentation and Measurement Technology, 1–4.

Fasel, B. and J. Luettin (2003) Automatic facial expression analysis: a survey, *Pattern Recognition*, **36**, 259–275.

Shinde, S. and S. Pande (2012) A survey on emotion recognition with respect to database and various recognition techniques, *International Journal of Computer Applications*, **58**, 9–12.

Youssif, A. and W. Asker (2011) Automatic facial expression recognition system based on geometric and appearance features, *Computer and Information Science*, **4**, 115–124.

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

Shih, F.Y., C.F. Chuang and P.S.P. Wang 2008 Performance comparisons of facial expression recognition in JAFFE database, *International Journal of Pattern Recognition and Artificial Intelligence*, **22**, 445–459.

Jun, B. and D. Kim (2012) Robust face detection using local gradient patterns and evidence accumulation, *Pattern Recognition*, **45**, 3304–3316.

Tu, C.J., L.Y. Chuang, J.Y. Chang and C.H. Yang (2007) Feature selection using PSO-SVM, *International Journal of Computer Science (IAENG)*, **33**, 111–116.

Ramadan, R.M. and R.F. Abdel-Kader (2009) Face recognition using particle swarm optimization-based selected features, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **2**, 51–66.

Rashedi, E., H. Nezamabadi-pour and S. Saryazdi (2009) GSA: a gravitational search algorithm, *Information Sciences*, **179**, 2232–2248.

Moghadam, M.S., H. Nezamabadi-Pour and M.M. Farsangi (2012) A quantum behaved gravitational search algorithm, *Intelligent Information Management*, **4**, 390–395.

Jamalipour, M., R. Sayareh, M. Gharib, F. Khoshahval and M.R. Karimi (2013) Quantum behaved particle swarm optimization with differential mutation operator applied to WWER-1000 in-core fuel management optimization, *Annals of Nuclear Energy*, **54**, 134–140.

Lyons M.J., M. Kamachi and J. Gyoba (1997) Japanese Female Facial Expressions (JAFFE), Database of digital images.

Lu, S., C. Sun and Z. Lu (2010) An improved quantum-behaved particle swarm optimization method for short-term combined economic emission hydrothermal scheduling, *Energy Conversion and Management*, **51**, 561–571.

Zhang, Z., M.J. Lyons, M. Schuster and S. Akamatsu (1998) Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron, Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, 454–459.

Bashyal, S. and G.K. Venayagamoorthy (2008) Recognition of facial expressions using Gabor wavelets and learning vector quantization, *Engineering Applications of Artificial Intelligence*, **21**, 1056–1064.

Koutlas A. and D.I. Fotiadis (2008) An automatic region based methodology for facial expression recognition, Proc. IEEE International Conference on Systems, Man and Cybernetics, 662–666.

Liu, W. and Z. Wang (2006) Facial expression recognition based on fusion of multiple Gabor features, Proc. IEEE 18th International Conference on Pattern Recognition, **3**, 536–539.

Oliveira, L.E.S., A.L. Koerich, M. Mansano and A.S. Jr. Britto (2011) 2-D principal component analysis for face and facial-expression recognition, *Computing in Science and Engineering*, **13**, 9–13.

Liao, S., W. Fan, C.S. Chung and D.Y. Yeung (2006) Facial expression recognition using advanced local binary patterns, Proc. IEEE International Conference on Image Processing, 665–668.

Cheng, F., J. Yu, and H. Xiong (2010) Facial expression recognition in JAFFE dataset based on Gaussian process classification, *IEEE Trans. on Neural Networks*, **21**, 1685–1690.

Zhi, R. and Q. Ruan (2008) Facial expression recognition based on two-dimensional discriminate locality preserving projections, *Neurocomputing*, **71**, 1730–1734.

Cowie, R., E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J.G. Taylor (2001) Emotion recognition in human-computer interaction, *Signal Processing Magazine*, **18**, 32–80.

Milanova, M. and N. Sirakov (2008) Recognition of emotional states in natural human-computer interaction, IEEE International Symposium on Signal Processing and Information Technology, 186–191.

Yampolskiy, R.V. and V. Govindaraju (2007) Direct and indirect human computer interaction based biometrics, *Journal of Computers*, **2**, 76–88.

Jiang, R.M., A.H. Sadka and D. Crookes (2010) Multimodal biometric human recognition for perceptual human-computer interaction, *IEEE Trans. on Systems, Man and Cybernetics, Part C: Applications and Reviews*, **40**, 676–681.

Chakraborti, T. and A. Chatterjee (2014) A novel binary adaptive weight GSA based feature selection for face recognition using local gradient patterns, modified census transform, and local binary patterns, Engineering Applications of Artificial Intelligence, Elsevier, **33**, 80–90.

Chakraborti, T., K.D. Sharma and A. Chatterjee (2014) A novel local extrema based gravitational search algorithm and its application in face recognition using one training image per class, Engineering Applications of Artificial Intelligence, Elsevier, **34**, 13–22.

The authors

Tapabrata Chakraborti

has completed BEng and MEng in Electrical Engineering from Jadavpur University, India. His areas of interest are broadly digital image processing and biomedical image analysis. He is currently working in a project on computer-aided hand written text image analysis with the CVPR Unit at the Indian Statistical Institute.

Amitava Chatterjee

currently serves as an associate professor in the Department of Electrical Engineering, Jadavpur University, Kolkata, India. In 2003, he received the Japanese Government (Monbukagakusho) Scholarship and visited Saga University, Saga, Japan. From November 2004 to November 2005, he was with the University of Electro-Communications, Tokyo, Japan, on a Japan Society for the Promotion of Science (JSPS) Post-Doctoral Fellowship for Foreign Researchers. In March to May, 2004, and in May to June, 2009, he visited University of Paris XII, Val de Marne, France, as an Enseignant-Invité (invited teacher). He has authored/coauthored more than 130 technical articles, including 80 international journal papers. He has coauthored a book and coedited two books, all published by Springer-Verlag, Germany. He presently serves as an editor/associate editor of three IEEE Transactions and two Elsevier journals. He is a member of the Technical Committee on “Imaging Measurements and Systems” of *IEEE Instrumentation & Measurement Society*, USA. He is a fellow of IETE (India), a fellow of Institution of Engineers (India), and a senior member of the IEEE (USA). His research interests include fuzzy-based nonlinear control, intelligent instrumentation, computer vision, robotics, image processing and pattern recognition, signal processing, and stochastic optimization techniques.

Anisha Halder

is currently serving as an associate professor in the Department of Electronics and Communication Engineering, Heritage Institute of Technology (HIT), Kolkata. Before joining HIT, Dr Halder received her PhD degree on Emotion Recognition Using Type-2 Fuzzy Sets from Jadavpur University, Kolkata. She is an author of over 30 publications, covering invited book chapters, journal papers, and conference proceedings. Her current research interest includes cognitive and brain sciences and human-computer interface design.

Amit Konar

is currently a professor in the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India. He is an author of eleven books and over 350 publications, including invited book chapters, journal papers, and conference proceedings, all in the broad area of machine intelligence. He supervised 16 PhD and over 140 ME theses, and has served as an associate editor/editor of many top international journals, published by the IEEE, Elsevier, and IOS press. His current research interest includes cognitive psychology, neuro-prosthetics, and computational biology with a focus to offer rehabilitative support to people suffering from neurological and psychological disabilities.