Knowledge-Based Systems 111 (2016) 248–267

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Intelligent facial emotion recognition using moth-firefly optimization

Li Zhang[a,*], Kamlesh Mistry[a], Siew Chin Neoh[b], Chee Peng Lim[c]

[a] Computational Intelligence Research Group, Department of Computing Science and Digital Technologies, Faculty of Engineering and Environment, University of Northumbria, Newcastle, NE1 8ST, UK
[b] Faculty of Engineering, Technology and Built Environment, UCSI University, Malaysia
[c] Institute for Intelligent Systems Research and Innovation, Deakin University, Waurn Ponds, VIC 3216, Australia

ARTICLE INFO

Article history:
Received 18 March 2016
Revised 18 August 2016
Accepted 19 August 2016
Available online 22 August 2016

Keywords:
Facial expression recognition
Feature selection
Evolutionary algorithm
Ensemble classifier

ABSTRACT

In this research, we propose a facial expression recognition system with a variant of evolutionary firefly algorithm for feature optimization. First of all, a modified Local Binary Pattern descriptor is proposed to produce an initial discriminative face representation. A variant of the firefly algorithm is proposed to perform feature optimization. The proposed evolutionary firefly algorithm exploits the spiral search behaviour of moths and attractiveness search actions of fireflies to mitigate premature convergence of the Levy-flight firefly algorithm (LFA) and the moth-flame optimization (MFO) algorithm. Specifically, it employs the logarithmic spiral search capability of the moths to increase local exploitation of the fireflies, whereas in comparison with the flames in MFO, the fireflies not only represent the best solutions identified by the moths but also act as the search agents guided by the attractiveness function to increase global exploration. Simulated Annealing embedded with Levy flights is also used to increase exploitation of the most promising solution. Diverse single and ensemble classifiers are implemented for the recognition of seven expressions. Evaluated with frontal-view images extracted from CK+, JAFFE, and MMI, and 45-degree multi-view and 90-degree side-view images from BU-3DFE and MMI, respectively, our system achieves a superior performance, and outperforms other state-of-the-art feature optimization methods and related facial expression recognition models by a significant margin.

1. Introduction

Facial expression recognition, which plays an important role in pattern recognition, computer vision, and human computer interaction, is widely used in personalised healthcare, video games, surveillance systems, humanoid service robots, and multimedia. In recent studies, many algorithms focusing on face recognition, gender and age estimation, and facial emotion classification have been developed. However, high dimensionality is still a challenging issue for such applications. Although many feature dimensionality reduction techniques have been proposed, it is still difficult to identify the most significant discriminating features that best represent within and between class variances for emotional facial expression.

In this research, we propose a facial emotion recognition system with a variant of evolutionary firefly algorithm for feature optimization. The main aim of the proposed system is to identify the most significant discriminative characteristics for each emotion category, in an attempt to address the above challenges. It integrates the spiral search behaviour of the moths, attractiveness search actions of the fireflies, and Simulated Annealing (SA) embedded with the Levy flights to increase local exploitation and global exploration and, at the same time, mitigate the premature convergence problem of the Levy-flight firefly algorithm (LFA) [1] and the moth-flame optimization (MFO) algorithm [2].

The proposed system is composed of three key steps, as illustrated in Fig. 1. Firstly, a novel texture descriptor is proposed, which incorporates the use of Local Binary Patterns (LBP), Local Gabor Binary Patterns (LGBP), and LBP variance (LBPV) [3] to capture local spatial patterns and contrast measures of local texture to retrieve an initial discriminative facial representation. Secondly, the proposed variant of the firefly optimization algorithm is used to identify the most significant and discriminative features of each emotion category. Thirdly, single and ensemble classifiers are used for recognizing seven expressions (happiness, fear, disgust, surprise, sadness, anger, and neutral) based on the derived optimal feature subsets. Evaluated with frontal-view images from CK+ [4], MMI [5] and JAFFE [6], and 45-degree multi-view and 90-degree

* Corresponding author.
E-mail addresses: li.zhang@northumbria.ac.uk (L. Zhang), kamlesh.mistry@northumbria.ac.uk (K. Mistry), u_jane80@yahoo.co.uk (S.C. Neoh), chee.lim@deakin.edu.au (C.P. Lim).

http://dx.doi.org/10.1016/j.knosys.2016.08.018

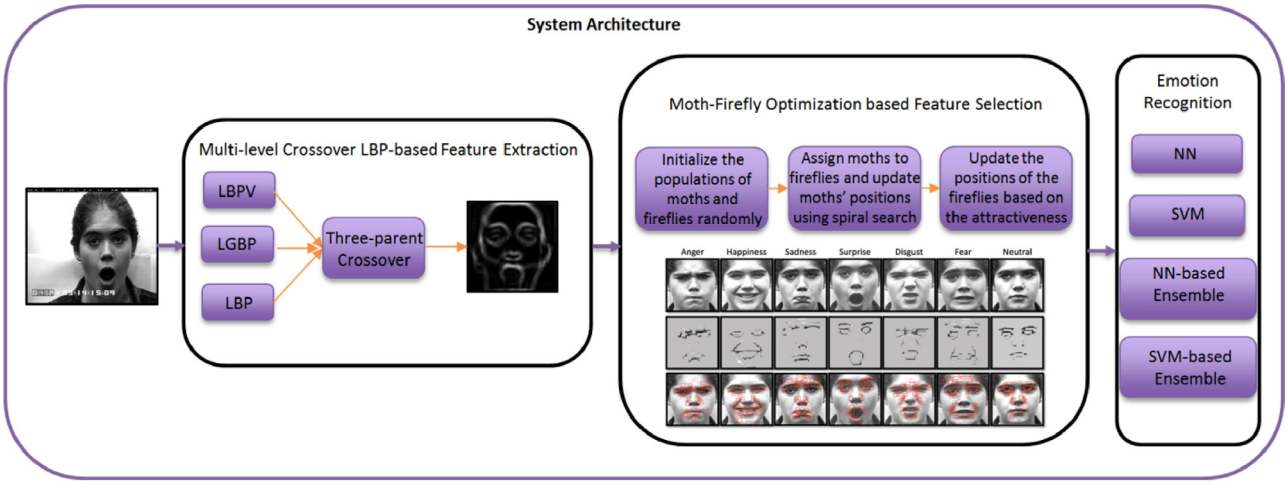

Fig. 1. The system architecture.

side-view images from BU-3DFE [7] and MMI, respectively, the empirical results indicate that our system shows a superior performance, and outperforms other state-of-the-art optimization methods and related facial expression recognition models by a significant margin.

The contributions of this research are summarized as follows.

1. A novel texture descriptor incorporating LBP, LGBP, and LBPV is proposed to derive an initial discriminative facial representation. The proposed descriptor is able to extract spatial patterns and contrast measures of each image region for facial analysis, in order to better deal with illumination changes, rotations, and scaling differences. The empirical findings indicate that it outperforms conventional texture descriptors for facial expression recognition.
2. A LFA variant, known as M-LFA, is proposed for feature optimization. It explores the spiral search behaviour of the moths and attractiveness search actions of the fireflies to mitigate the premature convergence problem of the conventional LFA and MFO models. Specifically, it employs the logarithmic spiral search of the moths to increase local exploitation of the fireflies. In comparison with the flames in MFO, the fireflies not only represent the best solutions identified by the moths, but also act as the search agents guided by the attractiveness function to cause sudden movements of the fireflies and associated moths to increase global exploration. Therefore, it increases local exploitation of LFA and global exploration of MFO to guide the search process towards global optima. SA-embedded Levy flights diversification is also used to further improve local exploitation of the identified current global best solution. Overall, the proposed strategies work cooperatively to avoid premature stagnation while guiding the search process towards global optima.
3. Evaluated with frontal-view images extracted from CK+, MMI, and JAFFE databases and multi-view and side-view images from BU-3DFE and MMI, respectively, the empirical results indicate that the proposed system shows superior capabilities of finding local and global optima simultaneously, and outperforms a number of conventional and state-of-the-art metaheuristic search methods such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), MFO, LFA, and other PSO and firefly algorithm (FA) variants, non-evolutionary feature selection algorithms, as well as other related facial expression recognition models reported in the literature by a significant margin.

The paper is organised in the following way. The related work on facial expression recognition and state-of-the-art feature optimization techniques are discussed in Section 2. Section 3 describes the key stages of the proposed system, which include facial feature extraction using the proposed LBP descriptor in Section 3.1, and the proposed M-LFA algorithm for feature optimization in Section 3.2. Section 4 presents the evaluation of the proposed system using frontal-view images extracted from CK+, MMI, and JAFFE and multi-view and side-view images from BU-3DFE and MMI, respectively. Section 5 presents some concluding remarks of this research, and identifies a number of directions for further work.

2. Related work

In this section, we review state-of-the-art research on evolutionary feature optimization and facial expression recognition.

2.1. Facial expression recognition

Many facial expression recognition applications have been proposed recently. Zhang et al. [8] proposed a multimodal learning method to learn the joint representation from texture and landmark modalities of facial images. A structured regularization (SR) in combination with an auto-encoder was proposed in their work to learn sparsity and density from each modality to generate the joint representation. Feature extraction and classification were also combined together in their proposed model. The proposed method was also capable of dealing with expression recognition tasks with head pose variations. Evaluated with CK+ and NVIE databases, the work showed superiority over other methods. Ali et al. [9] proposed a facial expression recognition system with empirical mode decomposition (EMD) based feature extraction. The EDM method decomposed 1-D facial signal into a set of intrinsic mode functions (IMFs), in which the first IMF was considered as facial features for expression classification. Their work employed three feature dimensionality reduction techniques, i.e., Principal Component Analysis (PCA) with Linear Discriminant Analysis (LDA), PCA with Local Fisher Discriminant Analysis (LFDA), and Kernel LFDA (KLFDA), to perform optimization of the extracted EMD-based features. Several classifiers including the Extreme Learning Machine with Radial Basis Function (ELM-RBF) were used to classify seven facial expressions. JAFFE and CK+ databases were employed for system evaluation. Shojaeilangari et al. [10] proposed a spatio-temporal descriptor named Histogram of Dominant Phase Congruency (HDPC) for facial expression recognition from video sequences. This proposed descriptor extended the phase congruency concept to 3D, and incorporated histogram binning to describe both motion and appear-

ance based features. The experimental results indicated that the proposed system was capable of dealing with expression recognition tasks with scaling variations and illumination changes. Evaluated with CK+ and AVEC 2011 video sub-challenge, the work showed impressive performance in terms of robustness and accuracy.

Siddiqi et al. [11] proposed a facial expression recognition system that applied a stepwise linear discriminant analysis (SWLDA) for feature extraction and the hidden conditional random fields (HCRFs) model for expression recognition. In their work, facial expressions were divided into three categories, i.e. lips-based, lips-eyes-based, and lips-eyes-forehead-based. The first step was to use SWLDA and HCRF to classify an input image into one of the three categories. Secondly, SWLDA and HCRF trained only for a particular category were used to determine an emotion label. Evaluated with CK+, JAFFE, MMI and Extended Yale B Face datasets, the work achieved significant improvement on recognition accuracy, but with expensive computational costs. Neoh et al. [12] proposed a facial expression system with direct similarity and Pareto-based feature optimization. The former was integrated with the concept of micro GA to identify feature subsets that could represent the common features of each expression. The latter took both within and between class variations into account for multi-objective feature optimization. Integrated with diverse ensemble classifiers, the work achieved impressive performances when tested with CK+ and MMI. The Pareto-based approach was also proven to be more efficient in dealing with challenging feature optimization for frontal and side-view images. Zhang et al. [13] conducted real-time 3D facial Action Unit (AU) intensity estimation and expression recognition. Their work employed the minimal-redundancy-maximal-relevance (mRMR) criterion to identify a set of 16 feature subsets among the initially extracted raw facial features, which were then used to estimate the intensities of 16 diagnostic AUs. A novel ensemble classifier was integrated with a clustering algorithm for classification of six universal facial expressions and detection of newly arrived, unseen novel emotion classes (those not included in the training dataset). In their work, a distance-based clustering method and the uncertainty measures of the base classifiers within each ensemble were used for novel class detection. Evaluated using the Bosphorus 3D database and real human subjects, the system achieved impressive performances for identification of six emotions and novel unseen expressions.

2.2. Feature selection and optimization algorithms

Evolutionary algorithms have been widely employed for feature optimization because of their impressive search capabilities [14,15]. In this section, we discuss state-of-the-art evolutionary feature optimization algorithms, which include the most popular conventional search algorithms such as PSO, GA, and SA, and other hybrid models.

2.2.1. Genetic algorithm

Motivated by the Darwinian principle of 'survival of the fittest', the GA was developed by Holland [16]. It employs three evolutionary operators, i.e. crossover, mutation, and selection. The crossover operator generates two offspring by exchanging part of a chromosome with the corresponding part of another. We employ the single-point crossover operator in this research. The mutation operator randomly changes one or more bits of an offspring chromosome in order to produce new genetic characteristics. Selection ensures the highest quality chromosomes will be selected and propagated to the next generation to enhance the convergence property of the algorithm. Crossover helps local exploitation to enhance convergence, while mutation brings search diversity and increases global exploration. According to theoretical studies, a higher crossover probability in the range of [0.6, 0.95] and a lower mutation probability in the range of [0.001, 0.05] are usually recommended. These settings enable a higher level of local exploitation and a lower degree of global exploration to reach global optimality [14].

2.2.2. Particle swarm optimization

Motivated by swarm behaviours such as bird flocking and fish schooling, PSO was proposed by Kennedy and Eberhart [17]. In PSO, a number of particles move in the search space by following the swarm leader, in order to find the optimal solutions. It records the best position ever achieved by a particle as the personal best, *p_best*, and the best position of the overall swarm as the global best, *g_best*. PSO employs the following strategies for updating the position and velocity of each particle.

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{1}$$

$$v_{id}^{t+1} = w \times v_{id}^{t} + c_1 \times r_1 \times \left(p_{id} - x_{id}^{t}\right) + c_2 \times r_2 \times \left(p_{gd} - x_{id}^{t}\right) \tag{2}$$

where x_{id}^{t+1} and v_{id}^{t+1} represent the position and velocity of each particle in the $(t+1)^{\text{th}}$ iteration in the d^{th} dimension, respectively. An inertia weight, w, is also introduced to adjust the effects of the previous velocity to the current one. Moreover, p_{id} and p_{gd} represent the personal best (*p_best)* and global best (*g_best*) in the d^{th} dimension, respectively. c_1 and c_2 denote the learning parameters or acceleration constants, whereas r_1 and r_2 indicate two random parameters between [0, 1].

Overall, PSO is a widely used swarm-based algorithm owing to its simplicity and flexibility. However it has limited exploration capability, and tends to be trapped in local optima [14].

2.2.3. Simulated annealing

Proposed by Kirkpatrick et al. [18], SA simulates the annealing process in material processing that usually requires a careful control of temperature and its cooling rate. SA has been widely used in diverse optimization problems. SA employs a random search technique for global exploration. It accepts not only better solutions but also those less ideal solutions with a probability,p,as defined in Eq. (3).

$$p = exp\left(-\frac{\Delta f}{T}\right) > r \tag{3}$$

where Δf denotes the change of the fitness function between the new and previous solutions. T represents the temperature for controlling the annealing process with r as a random value uniformly distributed in [0, 1]. As an example, in a minimization problem, a new solution with a higher fitness value than that of the current solution will be accepted with probability p

In SA, the annealing schedule (i.e. the cooling schedule), which controls the decreasing rate of the temperature, plays an important role in influencing local exploitation and global exploration. In this research, we employ a geometric cooling schedule, i.e. $T = \alpha T$, where the temperature is decreased by a cooling factor $\alpha \in [0, 1]$ [14]. In practice, SA is able to attain global optimality, but at the expense of a high computational cost [14].

2.2.4. Variants or hybrid optimization methods

Xue et al. [19] proposed two PSO-based multi-objective feature selection algorithms. The first algorithm incorporated non-dominated sorting into PSO (NSPSO) while the second algorithm integrated the concepts of crowding, mutation, and dominance into PSO (CMDPSO) to address feature optimization problems. They compared the performances of both algorithms with those of non-

dominated sorting genetic algorithm II (NSGAII), Strength Pareto Evolutionary Algorithm2 (SPEA2), Pareto Archived Evolutionary Algorithms (PAES) on twelve benchmark datasets. The experimental results indicated that CMDPSO selected the smallest number of features, and outperformed NSPSO, NSGAII, SPEA2, and PAES in terms of classification performance and computational efficiency. Zhang et al. [15] proposed a PSO variant called GM-PSO to identify the most discriminative bodily characteristics from static and dynamic motion features for the regression of arousal and valence dimensions for bodily expression perception. GM-PSO integrated PSO with GA and mutation techniques of Gaussian, Cauchy, and Levy distributions to overcome the premature convergence problem of conventional PSO. It outperformed GA, PSO, and other PSO variants in selecting the discriminative features and finding the global optimum. Goodarzi and Coelho [20] proposed a FA-based variable selection method for application to spectroscopy. They employed FA, PSO, and GA integrated with partial least squares (PLS) for spectroscopic data selection respectively. Evaluated with three spectroscopic datasets, FA identified the smallest number of wavelengths while maintaining a similar prediction performance. Alweshah and Abdullah [21] proposed two hybrid FA methods to optimize the weights of a probabilistic neural network (NN) to improve its classification performance. The first method integrated FA with SA (SFA), where SA was used to improve the final solution of FA. The second method combined SFA with Levy flights (LSFA) to further improve the global best solution. Tested with eleven benchmark datasets, LSFA outperformed SFA and LFA, and achieved impressive classification accuracy. Verma et al. [22] proposed a modified FA incorporating opposition-based and dimensional-based methodologies, known as ODFA, to deal with high dimensional optimization problems. The proposed ODFA model used opposition-based learning to perform initialization of the candidate solutions by including initialization of the opposite position of each firefly. It also employed the dimensional-based method to update the position of the global best firefly along each dimension. Evaluated with multidimensional standard functions, OFDA outperformed FA, PSO, and Differential Evolution (DE) significantly.

There are also other hybrid or modified FA algorithms that deal with diverse engineering optimization problems. As an example, Coelho et al. [23] proposed a modified FA model combined with chaotic maps to improve the convergence rate of the original FA model for solving reliability-redundancy optimization problems. Yang [1] proposed a Levy-flight Firefly Algorithm (i.e. LFA) to increase global exploration, which outperformed classical search algorithms such as PSO and GA. Abdullah et al. [24] proposed a hybrid FA model by combining FA with DE for high dimensional and nonlinear biological parameter optimization. A multi-objective FA model was also proposed by Arsuaga-Ríos and Vega-Rodríguez [25] by adding multi-objective properties into classical FA to deal with workload scheduling problems for minimizing energy consumption in grid computing. Kazem et al. [26] proposed a chaotic FA model that combined chaos theory with FA to identify optimal hyper-parameter settings of Support Vector Regression (SVR) for stock market price forecasting. Instead of randomly generating the initial population, the chaotic mapping operator (CMO) was used to produce the initial swarm to increase population diversity. Fireflies with a lower light intensity employed a chaotic movement to move towards those with a higher light intensity. Especially, the firefly with the highest light intensity purely used the chaotic movement, rather than a random walk behaviour, for exploring the search space. In comparison with GA-based SVR, chaotic GA-based SVR, FA-based SVR, NN, and the adaptive neuro-fuzzy inference system, chaotic FA outperformed these related methods in the evaluation of several most challenging stock market datasets from NASDAQ.

3. The proposed facial expression recognition system

In this section, we introduce the proposed facial expression recognition system, which includes a new LBP descriptor for feature extraction, a new M-LFA algorithm for feature optimization, as well as single and ensemble classifiers for facial emotion recognition.

3.1. Feature extraction using the proposed LBP descriptor

We propose a new texture descriptor, which combines LBP, LGBP, and LBPV, for feature extraction, in order to better deal with illumination changes, rotations, and scaling differences. The proposed LBP variant descriptor combines the discriminative capabilities of LBP, LGBP, and LBPV, and depicts great efficiency in extracting discriminative spatial patterns and contrast information for texture classification.

Proposed by Ojala et al. [27], LBP is a well-known texture descriptor. The basic idea of LBP is to threshold each group of 3×3 neighbouring pixels against the centre pixel to generate a sequence of binary outputs. It has been further extended to use various numbers of circular neighbouring pixels. The LBP descriptor can be denoted as LBPs,r, where s is the number of sampling points in the neighbourhood and r is the radius. The advantage of LBP is its invariance to monotonic gray-scale changes. LBP is efficient in extracting rotation invariant texture features from a local region. However, its drawback is loss of neighbourhood contrast and global information for texture description. LGBP [28] combines Gabor filters with LBP to improve the discriminative capability of LBP. It has excellent representation and discriminating power of spatial information of the face, and shows great robustness in dealing with illumination changes, misalignment, and scaling differences.

Introduced by Guo et al. [3], LBPV is a rotation invariant descriptor that focuses on exploiting local contrast information to further improve the discriminative capability of LBP. It combines two aspects of complementary texture information, i.e. local spatial structure extracted by LBP and the contrast, and generates a simplified joint representation. Specifically, the rotation invariant contrast measure (i.e. the variance of local image texture (VAR)) is calculated from a local region, and used as an adaptive weight to fine tune the impact of the LBP code to generate histograms. Therefore, enriched with contrast measures, LBPV possesses more robustness and discriminative capability than that of LBP for texture classification. It also has efficient computational complexity, and possesses the same feature dimensions as those of LBP. When pairing with global matching mechanisms, LBPV is able to outperform more complex, state-of-the-art joint LBP and contrast distribution descriptors such as LBP/VAR.

In this research, we combine the above three well-known texture descriptors to gain additional discriminating power to improve texture classification and better deal with rotations, illumination, and scaling differences. Moreover, the proposed LBP variant employs a three-parent crossover scheme for histogram generation. First of all, each of the three descriptors is applied to a gray-scale input image with a size of 75×75. The three texture descriptors generate a binary pattern for each sub-region of the test image, respectively. To combine the output patterns, a three-parent crossover scheme is applied, i.e., an offspring is derived from three parents. Because of the impressive discriminating capabilities of LBPV and LGBP in comparison with that of LBP, LBPV and LGBP are selected as the dominating parents, whereas LBP is used a reference parent to supply basic reference information when LBPV and LGBP disagrees. The proposed descriptor compares each bit of the first parent pattern generated by LBPV with the corresponding bit of the second parent generated by LGBP. If they are the same, this bit is inherited by the offspring. Other-

wise, the corresponding reference bit from the third parent generated by LBP is inherited by the offspring. As an example, suppose the 1st parent (generated by LBPV) = 11,010,001, the 2nd parent (produced by LGBP) = 0,110,0100, and the 3rd parent (obtained from LBP) = 11,011,010. After applying the three-parent crossover, the newly generated offspring is 11,010,000. Moreover, it is worth to point out that although the differences between the results obtained by the proposed LBP descriptor and the three baseline descriptors (e.g. LBPV) in the above example are minor, they are obtained from a small local region of 3×3. These differences could increase drastically to 625 (25×25) histograms when an image with a size of 75×75 is used. In comparison with the parent patterns, the experimental results indicate that the newly generated offspring possesses more discriminative capabilities. Evaluated using images with illumination and scaling differences derived from the CK+ database and multi-view and side-view images from BU-3DFE and MMI respectively, the proposed LBP operator shows superior performance, and outperforms the above three baseline descriptors significantly in a number of diverse test cases. Detailed evaluation results are provided in Section 4.1. Overall, the proposed descriptor is more capable of retrieving discriminative facial features such as edges and corners, and shows great robustness and efficiency in dealing with images with low contrast ratios.

3.2. Background and the proposed feature optimization algorithm

In this section, we introduce the proposed M-LFA algorithm for feature optimization. To provide the necessary background information, the original FA, LFA and MFO models are first introduced, as follows.

3.2.1. Firefly algorithm

Proposed by Yang in 2008 [14], FA is a nature inspired population-based metaheuristic search algorithm. FA applies the following three principles in its search process. Firstly, all fireflies are unisex, with one attracted to all others. Secondly, attractiveness is proportional to the brightness of a firefly. Therefore, the firefly with less brightness moves towards the one with stronger illuminations. If no brighter fireflies exist, a random walk behaviour is conducted. Thirdly, the light intensity of each firefly denotes the solution quality. Studies indicate that FA demonstrates promising superiority over other algorithms, such as PSO and GA [14].

In FA, the light intensity variation and attractiveness are two important aspects. They both decrease as either the distance to the light source or the distance between two fireflies increases. They also vary with the degree of light absorption. The light intensity variation with respect to the distance, r, and media absorption is defined in Eq. (4) [14].

$$I = I_0 e^{-\gamma r} \tag{4}$$

where I_0 is the original light intensity when $r = 0$, while γ denotes a fixed light absorption coefficient.

The attractiveness factor, $\beta(r)$, of a firefly is proportional to the light intensity, as defined in Eq. (5) [14].

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{5}$$

where β_0 is the initial attractiveness at $r = 0$.

Moreover, the distance between two fireflies i and j is computed in accordance with the Cartesian distance, as shown in Eq. (6).

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d} \left(x_{i,k} - x_{j,k}\right)^2} \tag{6}$$

where x_i and x_j represent the positions of fireflies i and j, respectively. $x_{i,k}$ indicates the kth dimension of position x_i for the ith firefly while d denotes the dimensions of a given problem.

In conventional FA, randomization is conducted using a Gaussian or uniform distribution. A Levy-flight Firefly Algorithm (i.e. LFA) was also proposed by Yang [1]. Levy flights are used to implement randomization to further enhance performance. The movement of firefly i towards a brighter firefly j is defined in Eq. (7).

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j - x_i\right) + \alpha \ \text{sign}\left[\text{rand} - \frac{1}{2}\right] \oplus \text{Levy} \tag{7}$$

where the second term indicates the movement due to attraction, and the third term denotes randomization using Levy flights. Note that α is the randomization parameter, while sign[rand $- \frac{1}{2}$] represents a random direction with the random step length following a Levy distribution, where rand generates a random number in the range of [0, 1].

In FA and LFA, the light absorption coefficient, γ, plays a very important role in characterising attractiveness and determining the convergence speed of the algorithms. When $\gamma \to 0$, attractiveness remains constant and the light intensity does not decrease. In this case, the behaviours of FA and LFA are very similar to that of PSO, where the whole population of individuals is visible in the search space, and the global optimum can be easily identified. When $\gamma \to \infty$, attractiveness becomes nearly non-existence, where each firefly performs the random walk operation (e.g. Levy flights for LFA and Gaussian mutation for FA) without interacting with other individuals in the search space. In this way, FA and LFA are equivalent to a random search algorithm such as SA. FA and LFA usually work between these two extreme cases, where different fireflies work independently to search for the optimal solutions. The experimental results indicate that FA and LFA are capable of finding global and local optima simultaneously. Studies also indicate that they automatically divide the population into subgroups, show impressive capability of dealing with multimodal optimization problems, and outperform PSO and GA in terms of accuracy and computational efficiency [1].

3.2.2. Moth-flame optimization

Introduced by Mirjalili [2], MFO is inspired by the transverse orientation navigation behaviours of moths. It possesses local and global search capabilities, and is efficient in solving optimization problems with constrained and unknown search spaces. In MFO, both moths and flames are employed to represent solutions and their positions denote the problem variables in the search space. Specifically, moths are designated as the search agents to identify optimal solutions in the search space, whereas flames are employed to indicate the best positions of the moths obtained so far.

MFO employs a logarithmic spiral function defined in Eq. (8) to update the position of a moth [2].

$$S\left(M_i, F_j\right) = D_i . e^{bt} . \cos\left(2\pi t\right) + F_j \tag{8}$$

where M_i represents the ith moth while F_j denotes the jth flame. $D_i = |F_j - M_i|$ is the distance between the ith moth and the jth flame, while b indicates a constant that defines the shape of the logarithmic spiral. t indicates how close the next position of the moth is to the flame, which is a value in the range of [-1, 1] where -1 and 1 indicate the closest and farthest to the flame, respectively. The spiral equation is equipped with efficient local and global search capabilities. It enables a moth to explore the search space, but not necessarily in the space between a moth and a flame, and to balance well between local exploration and global exploration. In other words, exploration is achieved when the optimal solution is found outside the space between a moth and a flame whereas exploitation occurs when the fitter solutions are found between them.

In addition, flames are ranked based on their fitness values in each iteration. In order to avoid local optimum and increase global

exploration, each moth updates its position with respect to a specific flame. The best flame is used for updating the position of the first moth, whereas the worst flame is employed for updating the position of the last moth. Overall, the updated flames are used for updating the positions of the moths in each generation, in order to explore the search space more effectively.

To increase local exploitation of the optimal solution vectors, the number of flames, *f_no*, in MFO is decreased adaptively. Eq. (9) defines the operation.

$$f\_no. = round\left(N - m \times \frac{N-1}{T}\right) \tag{9}$$

where m represents the current number of iterations with N and T indicating the maximum number of iterations and flames, respectively. Overall, MFO employs the logarithmic spiral search defined in Eq. (8) and the flame decrement strategy defined in Eq. (9) to balance between global exploration and local exploitation to achieve global optimality. It shows superior capabilities of dealing with multimodal and unimodal optimization functions, as well as other challenging optimization problems with unknown search spaces [2].

3.2.3. The proposed variant of the LFA algorithm

We propose a variant of the LFA algorithm, which integrates LFA with the concept of MFO for discriminative feature optimization. The resulting algorithm is known as M-LFA. The proposed M-LFA algorithm benefits from both spiral search behaviour of the moths and attractiveness search actions of the fireflies to mitigate the premature convergence problem of the original LFA and MFO models. It employs the logarithmic spiral search process of the moths to increase local exploitation of the fireflies to avoid stagnation. In comparison with the flames in MFO, the fireflies not only represent the best solutions identified by the moths, but also act as the search agents based on the attractiveness function to increase search diversity. Therefore, it increases local exploitation of LFA and global exploration of MFO to guide the search process towards the global optimum. The identified best solution is also further mutated by the SA operation with Levy flights, in order to generate an offspring further away from its parent and to increase exploitation. Overall, the above mechanisms work in a cooperative manner to overcome premature convergence and guide the search process towards global optimality. Algorithm 1 illustrates the proposed M-LFA algorithm. Fig. 2 shows the flowchart of the proposed algorithm.

As illustrated in Algorithm 1, the proposed M-LFA algorithm firstly initializes a population of fireflies and a swarm of moths, respectively. Then, the fitness of each moth and each firefly is evaluated using a fitness or objective function, $f(x_j)$. Two separate arrays are created to store and rank the corresponding fitness values for the fireflies and moths, respectively. This enables preservation of the best solutions obtained by the moths and fireflies, respectively. The next step is to initialize the light intensity of each firefly, where $I_j = f(x_j)$, and the constant light absorption coefficient, γ.

As discussed earlier, the main mechanism of the proposed algorithm consists of two search strategies. The first strategy employs the moth spiral concept to improve the exploitation capability of the fireflies, while the second strategy employs the attraction and attractiveness behaviour of LFA to enable sudden and optimal movement of the fireflies in the search space to diversify the search process. The first search strategy is similar to that of MFO, where each moth is assigned to a specific firefly that is ranked based on the fitness value. Specifically, the first moth is assigned to the best firefly, while the last moth is assigned to the worst firefly. The sequence of the fireflies is also updated based on the best solution found by the moths and the attractiveness impact between

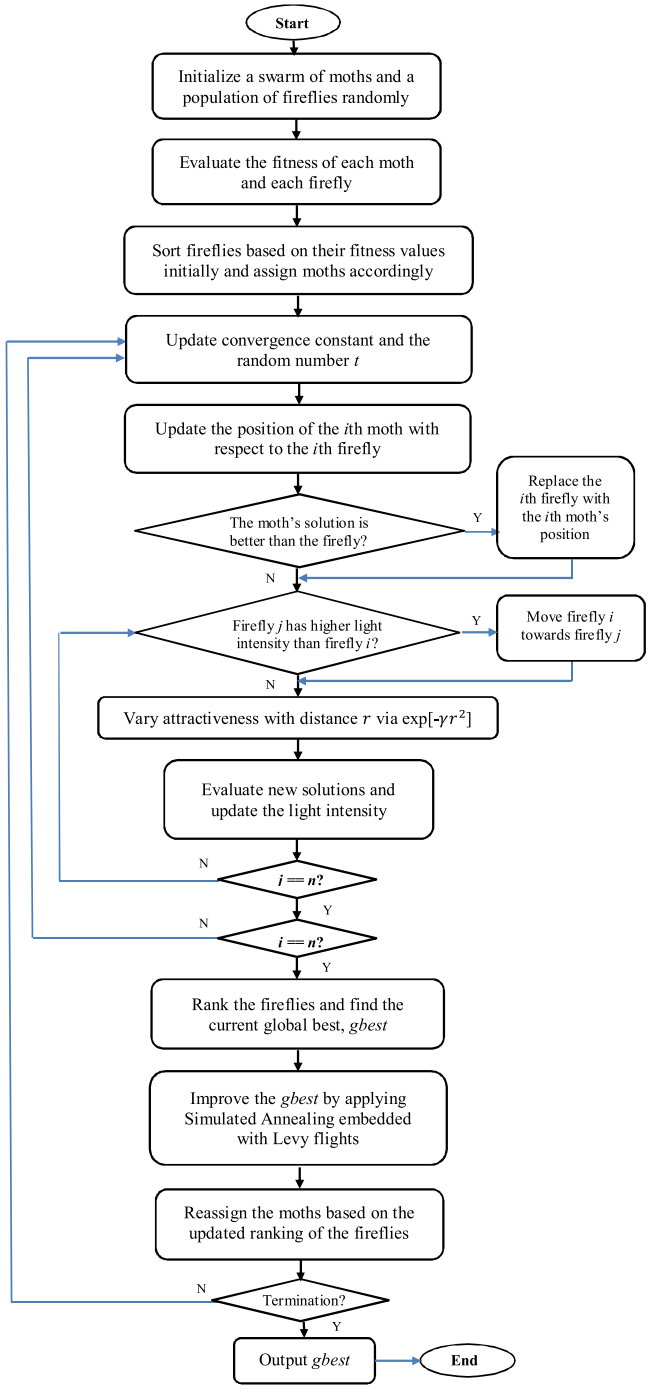

Fig. 2. Flowchart of the proposed M-LFA feature optimization algorithm.

the fireflies in each generation. Therefore, the strategy requires different moths to move around different fireflies, which enhances global exploration and reduces the probability of premature convergence. The updated logarithmic spiral movement defined in Eq. (10) is used to update a moth's position with respect to a firefly.

$$S\left(M_i, x_j\right) = D'_i.e^{bt}.\cos\left(2\pi t\right) + x_j \tag{10}$$

where M_i represents the ith moth while x_j is the position of the jth firefly, and $D'_i = |x_j - M_i|$ represents the distance between the ith moth and the jth firefly. In this way, each moth performs a spiral search around each firefly to exploit its neighbourhood. If the solution obtained by the moth has a better fitness, it is used to replace the position of the current firefly. Overall, the fireflies are

Algorithm 1: Pseudo-code of proposed moth-firefly algorithm.

```
1. Start
2. Initialize a population of moths and a population of fireflies randomly;
3. Generate two arrays to store the fitness values for moths and fireflies respectively;
4. Evaluate each moth and each firefly using the fitness/objective function f(x);
5. Set light intensity I = f(x) and light absorption coefficient γ;
6.
7. Sort fireflies based on their fitness values initially and assign moths accordingly;
8. While (Stopping criterion is not satisfied)// until it finds the optimal solution or the maximum number of iterations is met.
9. {
10.     For i = 1 to n do //for each moth attached with each firefly
11.         {
12.             //Using spiral search of moths to guide the search
13.             Update convergence constant and the random number t; // t parameter controls how close the next position of the moth is to the firefly, e.g.
                -1 is the closest and 1 is the farthest.
14.             Calculate the distance between the ith firefly (x_i) and the ith moth (M_i) using D_i = |x_i − M_i|;
15.             Update the position of the ith moth with respect to the ith firefly using Eq. (10);
16.             If (the moth's solution is better than the firefly)
17.             {
18.               Replace the ith firefly with the ith moth's solution;
19.             } End If
20.             //Using attractiveness function of fireflies to guide the search
21.             For j=1 to n do //for all fireflies
22.             {
23.               If (I_j > I_i)
24.             {
25.               Move firefly i towards firefly j using Eq. (7);
26.             } End IF
27.             Vary attractiveness with distance r via exp[-γr^2];
28.             Evaluate new solutions and update the light intensity;
29.             } End For
30.         } End For
31.     Rank the fireflies and find the current global best;
32.     Improve the global best by applying Simulated Annealing embedded with Levy flights;
33.     Reassign the moths based on the updated ranking of fireflies;
34. } End While
35. Output optimal solution(s);
36. End
```

updated with the most optimal solutions identified by the moths in each iteration.

In the proposed LFA variant, both moths and fireflies are regarded as the search agents. As such, they move around in the search space to search for the optimal solutions. As explained earlier, each firefly is updated not only by the fitter solutions found by the corresponding moths in the neighbourhood, but also moves towards other more attractive fireflies in the search space. Therefore, the second search strategy employs attractiveness and attraction action defined in Eq. (7) to move a firefly towards a brighter one, in order to increase search diversity. Since the moths perform spiral movement around the fireflies, and the fireflies move towards more attractive ones in each iteration, the second search strategy increases global exploration of the moths in MFO. The algorithm then updates the attractiveness and light intensity with respect to the distance.

Importantly, the above two search mechanisms work in a collaborative manner to enable the algorithm to escape from the local optimum. As an example, if the spiral search process of the moths stagnates and does not find fitter solutions for some of the fireflies, the attractiveness function still enables less fit fireflies to move towards better solutions, and reach more optimal search regions, therefore avoiding stagnation. In addition, if the fireflies fail to communicate or interact with each other because of a similar light intensity or foggy situations, the moths conduct spiral search around different fireflies to increase global exploration and fine tune the optimal solution vectors to overcome the local optimum. These strategies work cooperatively to mitigate premature convergence and guide the search process to attain global optimality.

After conducting the abovementioned two search processes, the set of fireflies is subsequently ranked based on their fitness values. The most promising solution (firefly) among the population is identified in each generation. To improve local exploitation of the current best solution, the SA operation defined in Eq. (11) is employed to perform mutation.

$$x_{t+1} = x_t + \varepsilon \tag{11}$$

where x_{t+1} and x_t represent the newly generated and the originally identified promising solutions, respectively, and ε indicates a standard random walk operation such as a Gaussian, Cauchy, or Levy distribution. In this research, Levy flights are used to mutate the original solution and generate an offspring further away from its parent. If the new offspring solution, *Nsol*, has a better fitness value, it is used to replace the current global best, *Csol*. However, if the new solution is worse than the current best solution, the SA accepts the new solution if it satisfies the following probability rule [14].

$$\exp\left(-\frac{f(Nsol) - f(Csol)}{T}\right) > random[0,\ 1] \tag{12}$$

where T represents the current temperature for controlling the annealing process (see Section 2.2.3). T is decreased in each iteration by a cooling factor $\alpha \in [0, 1]$ as defined in Eq. (13).

$$T = \alpha T \tag{13}$$

Subsequently, the new global best solution is used to guide the search process in the next generation. The algorithm iterates until the termination criterion is met, i.e. the maximum number of iteration is reached or the optimal solution is found. In this way, the algorithm is able to benefit from the optimal solutions obtained from both moths and fireflies simultaneously.

In the proposed algorithm, the following fitness function is applied to evaluate the fitness of each moth and firefly. It consists of two criteria, i.e. the number of selected features and classification accuracy. Note that classification accuracy refers to the accuracy rate obtained for each emotion category, rather than a combined accuracy score for all emotion classes, in order to avoid bias.

$$f(x) = w_a \times acc_x + w_f \times (number\_feature_x)^{-1} \quad (14)$$

where w_a and w_f represent the weights for classification accuracy and the number of selected features, respectively, with $w_f = 1 - w_a$. In this research, we consider classification accuracy is more important than the number of selected features, therefore w_a is set to a higher value (0.9) while w_f assumes a smaller value (0.1).

The setting of 0.9 and 0.1 as the weights for classification accuracy and the number of selected features, respectively, is obtained from the empirical studies of this research, although such a setting is also recommended in other studies [12,15,19,29]. In this research, we have also used an empirical demonstration by changing w_a:w_f from 0.9:0.1, 0.8:0.2, 0.7:0.3, 0.6:0.4 to 0.5:0.5. The setting of 0.9:0.1 has been selected owing to its performance in producing the best trade-off between classification accuracy and the number of selected features. This observation is also consistent with other findings, e.g. to further increase classification accuracy, more redundant features need to be removed. Indeed, the experimental results indicate that the selected feature subsets by M-LFA are more discriminative than those obtained by other state-of-the-art PSO and FA variants (e.g. GM-PSO [15], chaotic FA [26]), and non-evolutionary feature selection methods [9, 13] (see the evaluation details in Sections 4.2 and 4.3).

Overall, the proposed M-LFA algorithm employs spiral search, attractiveness action, and random walk operations to diversify the search process and increase local exploitation and global exploration. The empirical results indicate that it has superior capabilities in finding the global optimum, and outperforms metaheuristic search methods such as PSO, GA, MFO, LFA, GM-PSO [15], LSFA [21], ODFA [22] and chaotic FA [26], significantly.

For the experimental study, we employ NN and Support Vector Machine (SVM) for recognition of seven expressions. Optimal settings of NN and SVM are identified using the trial-and-error and grid search methods, respectively. The AdaBoost procedure is also used to construct two ensemble classifiers for expression classification, i.e. NN-based and SVM-based ensemble models, where the former employs three NNs as the base classifiers with the latter using three SVMs as the base classifiers. Each ensemble model employs the weighted majority voting method to combine the outputs from the three base classifiers to generate the final classification result [12, 29]. The empirical results indicate that the SVM-based ensemble model outperforms both the NN-based ensemble and single classification models in diverse experimental settings.

4. Evaluation

We employ the frontal-view images from CK+ and images with 45-degree and 90-degree rotations from BU-3DFE and MMI, respectively, for evaluating the proposed LBP descriptor for feature extraction. Moreover, the frontal-view images from CK+, JAFFE, and MMI, and multi-view and side-view images from BU-3DFE and MMI are also used to evaluate the proposed M-LFA algorithm for feature optimization.

4.1. Evaluation of the proposed LBP descriptor for feature extraction

To evaluate the proposed LBP descriptor, three baseline LBP descriptors, i.e. LBP, LGBP and LBPV, have been employed for comparison. Distinctive sets of 250 and 175 images representing the seven facial expressions from the CK+ database are used for training and test, respectively. In this experiment, we use each LBP descriptor integrated with single and ensemble classifiers for expression recognition without applying any feature selection methods. Table 1 shows the results from the entire sets of raw features extracted by the proposed LBP variant and the original LBPV, LGBP and LBP descriptors, respectively.

As indicated in Table 1, the best results are produced using the SVM-based ensemble model, and the proposed LBP descriptor outperforms LBP, LGBP, and LBPV by 14.80%, 8.33%, and 5.25%, respectively. The empirical results indicate efficiency and superiority of the proposed LBP descriptor over LBP, LGBP, and LBPV.

To further evaluate the efficiency of the proposed LBP descriptor in dealing with rotations, illumination changes and scaling differences, we have generated four sets of images with 45-degree and 90-degree rotations, illumination changes, and scaling differences, respectively, for evaluation purposes. Firstly, all 175 test images from CK+ have been converted to those with illumination changes by using the brightness and contrast adjustment function provided by OpenCV [30], as follows.

$$g(i, j) = \alpha f(i, j) + \beta \quad (15)$$

where $g(i, j)$ denotes the output image pixels and $f(i, j)$ denotes the source image pixels with i as the row of pixels and j as the column of the pixels. α represents the gain and β represents the bias, which are used to control the contrast and brightness, respectively. The original test images are set into high and low brightness alternatively, using the above equation to generate the new test images.

Similarly, the original 175 test images have also been transformed to a set of 175 images with scaling differences using the OpenCV resize() function [30]. The same training set of 250 images from CK+ used in the previous experiment has been employed for training before newly generated test images with illumination changes and scaling differences are used for evaluation.

To evaluate the LBP descriptors using rotated images, 45-degree multi-view images from the BU-3DFE database and 90-degree side-view images from the MMI database are also extracted. Specifically, a total of 140 side-view images with 90-degree rotations are extracted from the video sequences of MMI, with half of them (i.e. 70) employed for training and the other half (i.e. 70) for test. A set of multi-view images with 45-degree rotations from BU-3DFE is also employed for evaluation, with 500 images for training and another 250 images for test. In each experiment, the corresponding LBP-based feature extraction method is applied and integrated with diverse classifiers without any feature selection process. The detailed evaluation results for cases of illumination changes, scaling differences, and rotations are summarised in Tables 2–5.

As indicated in Tables 2–5, the proposed LBP descriptor shows great robustness and efficiency in dealing with illumination changes, scaling differences, and rotations. It outperforms the three baseline LBP descriptors in the above diverse test cases significantly. Integrated with the SVM-based ensemble model, all the LBP descriptors achieve the highest accuracy rate in each experiment. When the SVM-based ensemble model is applied, the proposed LBP descriptor outperforms the three baseline descriptors by 10.17–16.62%, 10.18–17.83%, 3.77–8% and 5.9–7.1%, for images with illumination changes and scaling differences, and multi-view images with 45-degree rotations (BU-3DFE), and side-view images with 90-degree rotations (MMI), respectively. Moreover, the 90-degree side-view images from the MMI database pose the most challenging problem because of the dramatic information loss in side-view expressions. Our LBP descriptor, however, still shows more discriminating capabilities as compared with those from other LBP descriptors in handling such images. Overall, the empirical results indicate that the three strategies incorporated in the proposed LBP descriptor are able to better preserve the distinctiveness and dif-

Table 1
Classification performance of all features extracted by the proposed LBP descriptor and the original LBPV, LGBP, and LBP descriptors for the CK+ database.

Feature Extraction Methods	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
LBP	63.20	64.83	70.00	71.20
LGBP	68.12	68.98	75.00	77.67
LBPV	70.54	71.00	77.32	80.75
The proposed LBP descriptor	77.70	79.10	83.50	86.00

Table 2
Performance comparison between the proposed LBP descriptor and other LBP descriptors without any feature selection using 175 images with illumination changes derived from CK+.

Feature extraction models	NN %	SVM %	NN-based Ensemble %	SVM-based Ensemble %
LBP	59.45	62.75	65.00	66.05
LGBP	65.95	67.23	70.35	72.25
LBPV	68.35	69.15	71.88	72.50
The proposed LBP descriptor	75.50	77.07	80.33	82.67

Table 3
Performance comparison between the proposed LBP descriptor and other LBP descriptors without any feature selection using 175 images with scaling differences derived from CK+.

Feature extraction models	NN %	SVM %	NN-based Ensemble %	SVM-based Ensemble %
LBP	60.10	61.88	64.35	65.50
LGBP	66.00	68.00	71.67	73.15
LBPV	67.25	68.45	70.85	71.55
The proposed LBP descriptor	76.66	78.81	81.24	83.33

Table 4
Performance comparison between the proposed LBP descriptor and other LBP descriptors without any feature selection using 70 side-view images with 90-degree rotations extracted from MMI.

Feature extraction models	NN %	SVM %	NN-based Ensemble %	SVM-based Ensemble %
LBP	50.70	51.25	53.44	53.90
LGBP	51.00	51.50	54.00	54.00
LBPV	51.67	52.05	54.60	55.10
The proposed LBP descriptor	55.35	57.15	60.50	61.00

Table 5
Performance comparison between the proposed LBP descriptor and other LBP descriptors without any feature selection using 250 multi-view images with 45-degree rotations extracted from BU-3DFE.

Feature extraction models	NN %	SVM %	NN-based Ensemble %	SVM-based Ensemble %
LBP	70.00	72.33	74.00	74.00
LGBP	73.45	74.25	76.50	77.33
LBPV	74.10	74.75	76.00	78.23
The proposed LBP descriptor	79.00	79.00	81.50	82.00

ferentiate different local structures in the neighbouring pixels of an input image. Fig. 3 shows the example outputs of all the LBP operators for images with illumination changes, scaling differences and rotations.

4.2. Comparison of the proposed M-LFA feature optimization with other metaheuristic search methods

To evaluate the proposed M-LFA algorithm for feature selection, a number of state-of-the-art and conventional search methods are used for comparison purposes, which include PSO, GA, LFA, MFO, GM-PSO [15], LSFA [21], ODFA [22], and chaotic FA [26]. We employ the frontal-view images from CK+, JAFFE and MMI, multi-view images with 45-degree rotations from BU-3DFE, and side-view images with 90-degree rotations from MMI in the experimental study. Specifically, we employ 250 frontal-view images from the CK+ database for training, and 175 images extracted from each of the CK+, MMI, and JAFFE databases for test. Moreover, 500 and 250 multi-view images with 45-degree rotations from BU-3DFE are also used for training and test respectively. Another set of 140 side-view images with 90-degree rotations from MMI is also employed, with 70 images for training and the remaining 70 images for test. In each experiment, the proposed LBP descriptor is used to extract the initial features. Then, each feature optimization algorithm is

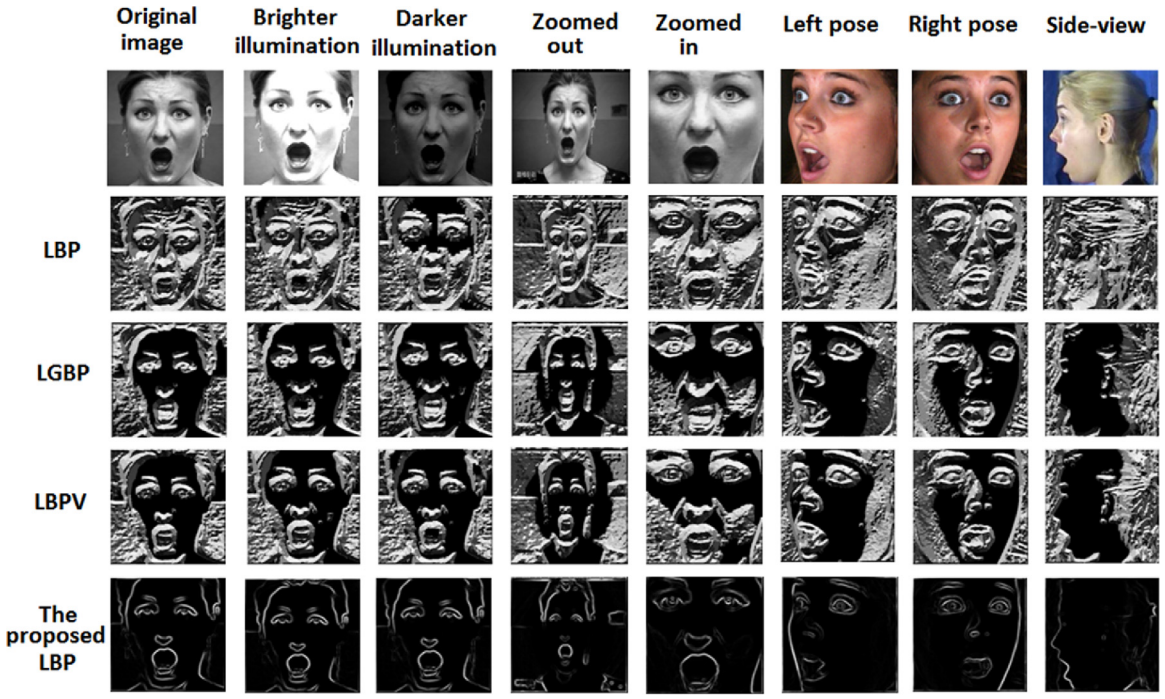

Fig. 3. Example outputs of all the LBP operators for images with illumination changes (the 2nd and 3rd columns derived from CK+), scaling differences (the 4th and 5th columns derived from CK+) and rotations (the 6th and 7th columns from BU-3DFE and the last column from MMI).

employed, before single and ensemble classifiers are used for expression recognition.

Firstly, the optimal settings of the proposed M-LFA algorithm and all other optimization methods are identified. As an example, to achieve the best trade-off between classification accuracy and the computational efficiency, LFA employs the following settings, i.e. population size=30, initial attractiveness=1.0, randomization parameter=0.2, absorption coefficient=1.0, Levy's index=1.5, and maximum iterations=500. These settings of LFA have also been applied to other FA variants (i.e. LSFA, ODFA and chaotic FA), and the proposed M-LFA algorithm, except that M-LFA has a population size of 30 moths plus 30 fireflies. Moreover, the following optimal settings have also been applied to PSO based on published studies and our empirical results, i.e. maximum velocity=0.6, inertia weight=0.78, population size=30, acceleration constants c_1=c_2=1.2, and maximum generations=500. The setting of classical GA is as follows: crossover probability=0.6, mutation probability=0.05, and maximum generations=500. The above PSO and GA parameters are also used as the optimal settings of GM-PSO.

4.2.1. Within database evaluation using Frontal-view images from CK+

Since the proposed and other feature optimization algorithms are stochastic in nature, we perform 30 trials to find the most discriminative feature subsets for each algorithm. In the first experiment, we employ 250 and 175 frontal-view images from the CK+ database for training and test, respectively. Empirically, the proposed M-LFA algorithm is able to converge within 200 to 300 iterations in most cases with a set of 30 to 50 features extracted. Moreover, we have compared the proposed M-LFA algorithm with other state-of-the-art and conventional metaheuristic search methods. Table 6 shows the average classification accuracy rates of each method integrated with diverse classifiers over 30 successive runs, respectively.

In comparison with all other methods, the proposed algorithm is able to extract the lowest number of features in the range of [30–50] and achieve the highest average accuracy rates when combined with all single and ensemble classifiers. When NN- and SVM-based ensemble models are applied, our algorithm achieves its best performance of 100% accuracy. Integrated with the NN-based ensemble model, our algorithm outperforms GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 21.12%, 18.67%, 14.27%, 8.54%, 10.7%, 16.11%, 9.1%, and 7.67%, respectively. In combination with the SVM-based ensemble model, the proposed algorithm outperforms GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 20%, 17.5%, 12.19%, 7%, 10%, 13.25%, 7.35%, and 5.5%, respectively. Moreover, the above accuracy rates obtained using the proposed M-LFA algorithm significantly outperform those using the entire set of raw features with the proposed LBP descriptor without any feature selection, as shown in Table 1.

Fig. 4 shows some examples of the generated optimized feature sub-regions for each expression of the images from CK+. Overall, significant discriminative characteristics are revealed for each expression, which correlate well with the emotional muscular activities defined in the Facial Action Coding System (FACS) [31]. As an example, the characteristics associated with the lip corner puller and cheek raiser are observed in the optimized sub-regions for "happiness", whereas mouth open, eye widened and the inner and outer brow raisers are explicitly illustrated in the facial sub-regions for "surprise" . The M-LFA algorithm also reveals features that are closely associated with the nose wrinkler, upper lip raiser, chin raiser and lips part for "disgust" . The significance of the lip stretcher, widened eyes, outer brow raiser and brow lowerer is observed in the selected facial sub-regions for "fear" . The significance of the brow lowerer, lid and lip tightener is explicitly demonstrated in the selected feature subset for "anger", while the inner brow raiser, brow lowerer and lip corner depressor are clearly indicated in the optimized facial regions for "sadness" . Overall, significant discriminative characteristics are revealed for each expression, which map closely to the AUs provided in FACS.

Table 6
Average classification results for different optimization models over 30 runs using the CK+ database.

Feature selection methods	Number of features	NN (30 runs) %	SVM (30 runs) %	NN-based Ensemble (30 runs) %	SVM-based Ensemble (30 runs) %
GA	100–200	74.60	76.90	78.88	80.00
PSO	110–200	76.33	78.70	81.33	82.50
LFA	55–80	80.00	80.00	85.73	87.81
MFO	40–90	87.80	89.10	91.46	93.00
GM-PSO [15]	45–80	83.76	86.45	89.30	90.00
LSFA [21]	60–85	79.00	79.50	83.89	86.75
ODFA [22]	50–80	87.22	88.55	90.90	92.65
Chaotic FA [26]	45–90	88.65	89.00	92.33	94.50
The proposed M-LFA algorithm	30–50	95.66	96.50	100	100

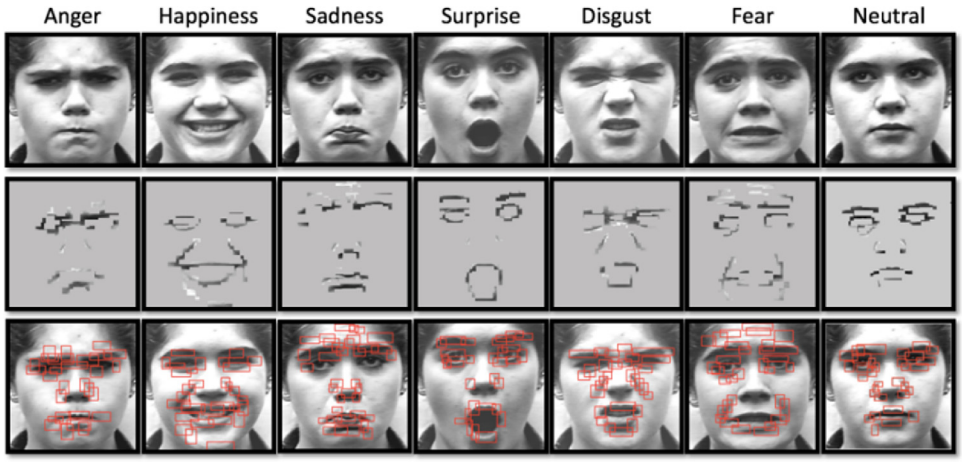

Fig. 4. The sub-regions and their distributions selected by M-LFA for each expression of the CK+ images.

Table 7
Average classification results over 30 runs for cross-database evaluation with JAFFE.

Feature selection methods	Number of features	NN (30 runs) %	SVM (30 runs) %	NN-based Ensemble (30 runs) %	SVM-based Ensemble (30 runs) %
GA	100–200	72.21	73.65	77.00	78.30
PSO	110–200	73.13	75.90	78.72	79.89
LFA	55–80	79.21	79.67	81.62	84.55
MFO	40–90	84.85	86.31	88.21	89.63
GM-PSO [15]	45–80	84.25	85.96	88.50	89.00
LSFA [21]	60–85	78.40	79.00	83.55	85.95
ODFA [22]	50–80	84.79	85.00	88.55	88.78
Chaotic FA [26]	45–90	88.00	88.00	90.75	91.45
The proposed M-LFA algorithm	30–50	94.21	95.30	100	100

4.2.2. Cross database evaluation using Frontal-view images from JAFFE and MMI

To further evaluate the efficiency of the proposed algorithm, a cross-database evaluation is conducted. A set of 175 frontal-view images is extracted from MMI and JAFFE, respectively, for test, while the above 250 frontal-view images from CK+ are employed as the training set. Table 7 shows the average performance of each algorithm in combination with single and ensemble classifiers for evaluation of 175 images extracted from the JAFFE database over 30 runs.

As illustrated in Table 7, in comparison with all other methods, the proposed M-LFA algorithm shows great robustness for cross-database evaluation with JAFFE. Integrated with single and ensemble classifiers, it achieves the best average accuracy over 30 runs. Integrated with the NN-based ensemble model, the average accuracy (i.e. 100%) of the proposed algorithm is 23%, 21.28%, 18.38%, 11.79%, 11.5%, 16.45%, 11.45% and 9.25% higher than those of GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA, respectively. Combined with the SVM-based ensemble model, the average accuracy (i.e. 100%) of the proposed algorithm outperforms those of GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 21.7%, 20.11%, 15.45%, 10.37%, 11%. 14.05%, 11.22%, and 8.55%, respectively.

Another cross-database evaluation is also conducted using 175 frontal-view images from the MMI database. The average classification results of each algorithm integrated with diverse classifiers over 30 trials are provided in Table 8. As shown in Table 8, trained with 250 images from CK+ and tested with 175 images from MMI, the proposed algorithm achieves the highest average accuracy rates in combination with all classifiers over 30 runs. When the SVM-based ensemble model is used, it achieves an average accuracy rate of 94.86%, and outperforms GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 17.35%, 16.8%, 8.46%, 5.07%, 6.88%, 7.81%, 5.98%, and 5.91%, respectively.

Table 8
Average classification results over 30 runs for cross-database evaluation with MMI.

Feature selection methods	Number of features	NN (30 runs) %	SVM (30 runs) %	NN-based Ensemble (30 runs) %	SVM-based Ensemble (30 runs) %
GA	100–200	69.60	69.79	75.34	77.51
PSO	110–200	71.03	73.11	77.51	78.06
LFA	55–80	79.90	81.76	83.33	86.40
MFO	40–90	84.83	85.30	87.16	89.79
GM-PSO [15]	45–80	83.24	83.77	86.45	87.98
LSFA [21]	60–85	79.33	81.45	85.75	87.05
ODFA [22]	50–80	82.13	83.06	86.95	88.88
Chaotic FA [26]	45–90	86.00	86.88	87.67	88.95
The proposed M-LFA algorithm	30–50	91.21	91.44	94.27	94.86

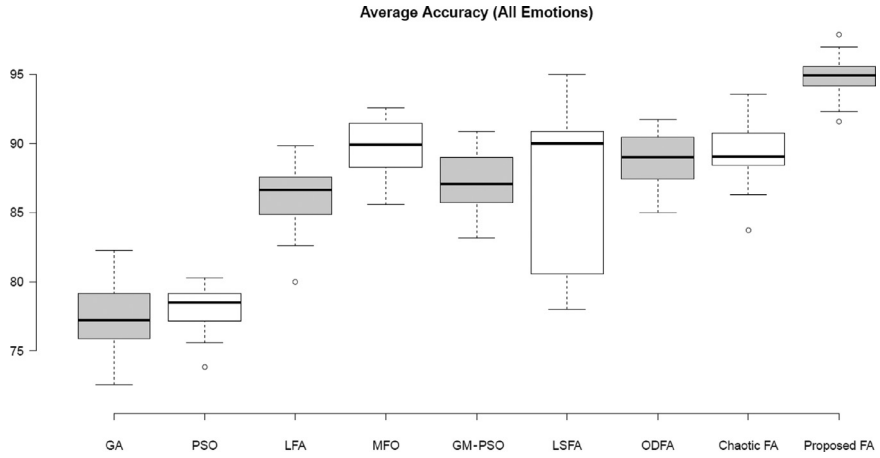

Fig. 5. The boxplot diagram for all the optimization methods integrated with SVM-based ensemble model over 30 runs using the MMI database.

Fig. 5 shows the boxplot diagram for the distribution of classification results over 30 runs of all optimization algorithms integrated with the SVM-based ensemble model using 175 images from MMI. As shown in Fig. 5, the proposed algorithm outperforms all other methods significantly. Nearly all the results of our algorithm are higher than the maximum accuracy rates of GA, PSO, LFA, MFO, GM-PSO, and ODFA with at least 75% of our results higher than the highest accuracy rate of chaotic FA, and 25% of our results higher than the maximum accuracy rate of LSFA. In comparison with other algorithms, the proposed algorithm also has a better accuracy distribution with comparatively smaller variations between the 25% and 75% percentiles. The median value of our algorithm (94.93%) is higher than those of GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 17.72%, 16.43%, 8.29%, 5%, 6.93%, 7.86%, 5.93%, and 5.86% respectively. Fig. 6 also illustrates the detailed performance variations of each emotion category over 30 runs for all algorithms using the MMI images.

As illustrated in Fig. 6, for each emotion category, the median value of our algorithm is higher than those of all other methods. For recognition of “sadness” and “fear” emotions, 25% of the results from our algorithm are higher than the maximum results of all other algorithms. For the “disgust” emotion, except for LSFA and ODFA, 25% of our results are also higher than the maximum results of the rest of the methods. For “happiness”, the minimum accuracy rate of our algorithm (with a lower whisker of 92%) is higher than 50% of the results of MFO, ODFA, and chaotic FA, 75% of the results of GA, PSO, GM-PSO, and LSFA, and all the results of LFA. For “anger”, the minimum accuracy rate of our algorithm (with a lower whisker of 89%) is higher than nearly 50% of the results of MFO, ODFA, and chaotic FA, 75% of the results of LFA, GM-PSO, and LSFA, and all the results of GA and PSO. For the “surprise” emotion, the median value of our algorithm (100%) is 5%, 5%, 5.5%, 6.5%, 7.5%, 8%, 15.5%, and 16% higher than those of chaotic FA, MFO, ODFA, LSFA, GM-PSO, LFA, PSO, and GA, respectively. For the “neutral” emotion, in comparison with all other methods, our algorithm has a better accuracy distribution with comparatively smaller variations between the 25% and 75% percentiles, and the minimum accuracy rate of our algorithm (with a lower whisker of 91%) is higher than at least 25% of the results of MFO and chaotic FA, 50% of the results of LSFA and ODFA, 75% of the results of LFA and GM-PSO, and all results of GA and PSO. Overall, the evaluation results indicate superiority of our algorithm. It outperforms all other methods by a significant margin.

Fig. 7 shows some examples of the generated optimized facial sub-regions for each expression pertaining to the images from MMI and JAFFE using the proposed M-LFA algorithm. Similar observations as those from CK+ can be explicitly observed in the example outputs. In general, significant discriminative characteristics associated with each expression are revealed, which indicate efficiency and superiority of the proposed M-LFA algorithm.

4.2.3. Evaluation using images with rotations from MMI and BU-3DFE

As indicated in our previous experiments, the 45-degree multi-view images from BU-3DFE and 90-degree side-view rotated facial images from MMI reveal significant information loss and pose great challenges to state-of-the-art facial expression recognition systems. Therefore, we employ such multi-view and side-view images from BU-3DFE and MMI, respectively, to further ascertain the robustness of our feature selection algorithm. In the experiment, we use the proposed LBP descriptor for feature extraction. Then, each feature selection algorithm is used for feature optimization before employing the single and ensemble classifiers.

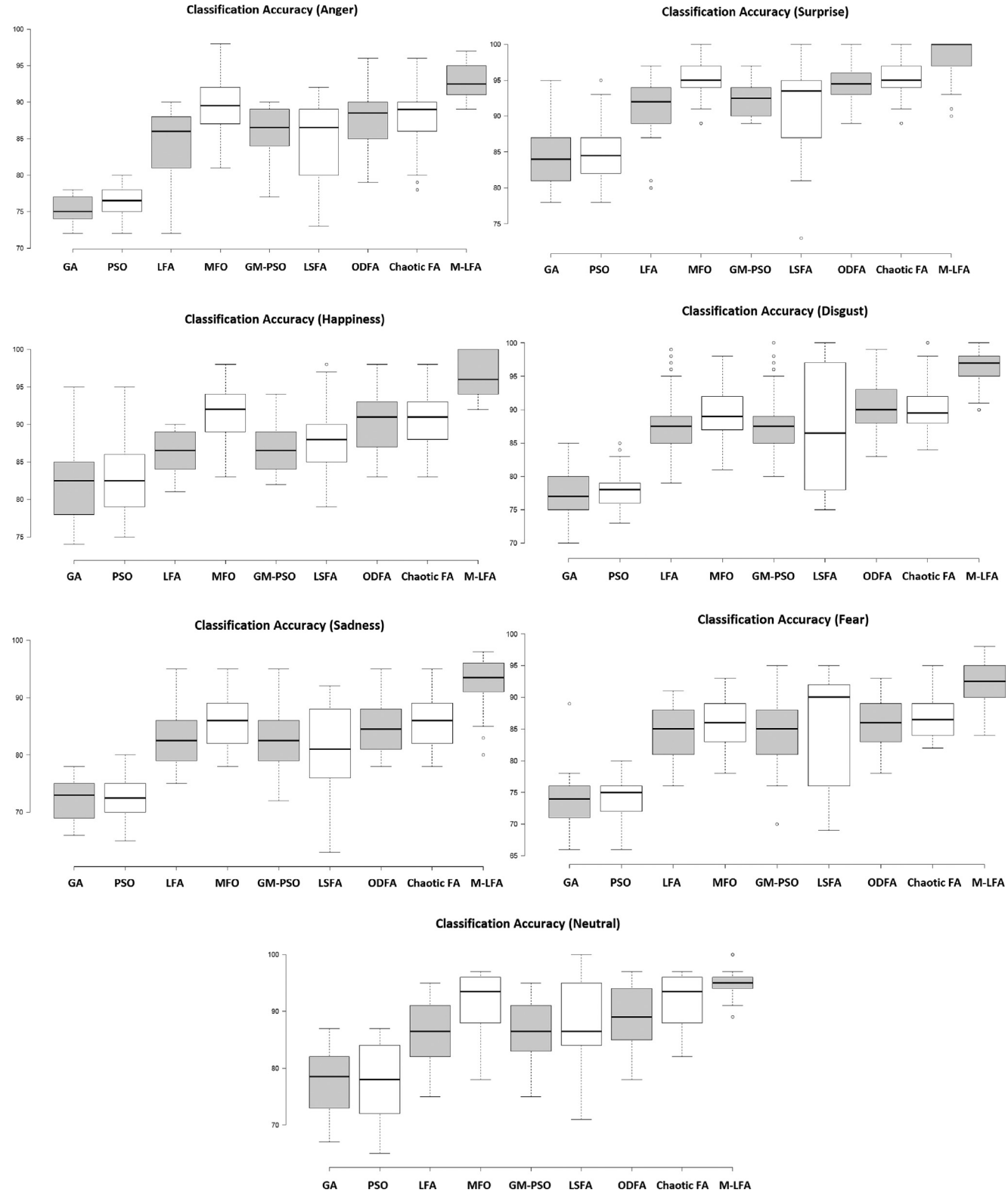

Fig. 6. The boxplot diagrams for each emotion category over 30 runs using the MMI database.

Firstly, we employ previous 140 side-view images from MMI for evaluation. A total of 70 images are used for training, with the remaining images for test. Table 9 shows the average classification results of each algorithm in combination with diverse classifiers over 30 runs. Some examples of the selected optimized facial subregions for each expression of the side-view images are illustrated in Fig. 8.

As illustrated in Table 9, the proposed algorithm achieves the best accuracy scores in combination with diverse classifiers. When the SVM-based ensemble model is applied, M-LFA achieves an average accuracy rate of 86.35%, and outperforms GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 21.23%, 20.45%, 15.6%, 11.35%, 10.25%, 12.68%, 11.35% and 9.9%, respectively. As indicated in Fig. 8, significant discriminative features associated with each

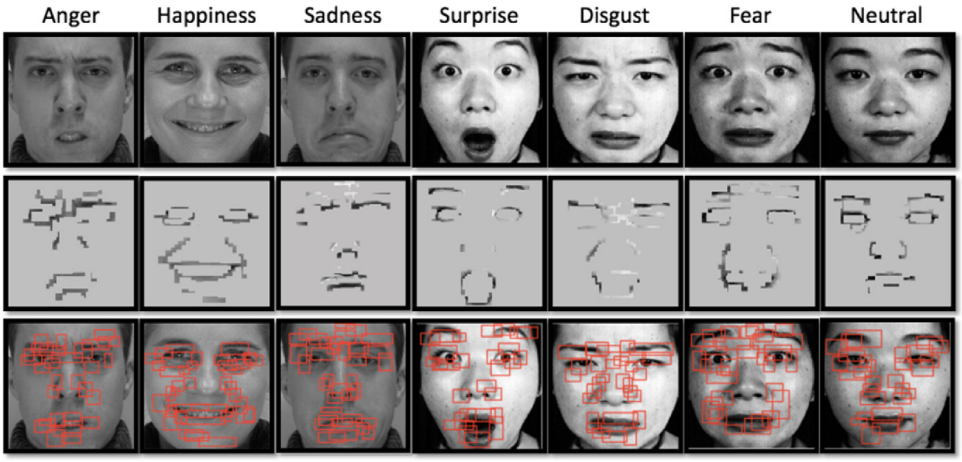

Fig. 7. The sub-regions and their distributions selected by the proposed M-LFA for each expression for images from MMI (the first three images) and JAFFE (the last four images).

Table 9
Average classification results over 30 runs for the 90-degree side-view images extracted from MMI.

Feature selection methods	Number of features	NN (30 runs) %	SVM (30 runs) %	NN-based Ensemble (30 runs) %	SVM-based Ensemble (30 runs) %
GA	120–250	60.44	60.77	64.43	65.12
PSO	100–195	61.85	61.95	65.00	65.90
LFA	45–90	66.99	68.19	70.09	70.75
MFO	50–85	71.20	71.88	73.50	75.00
GM-PSO [15]	45–80	72.33	72.50	74.75	76.10
LSFA [21]	65–85	70.25	71.00	72.34	73.67
ODFA [22]	50–70	70.05	70.78	72.45	75.00
Chaotic FA [26]	50–80	74.55	75.00	76.05	76.45
The proposed M-LFA algorithm	40–65	78.00	80.40	85.99	86.35

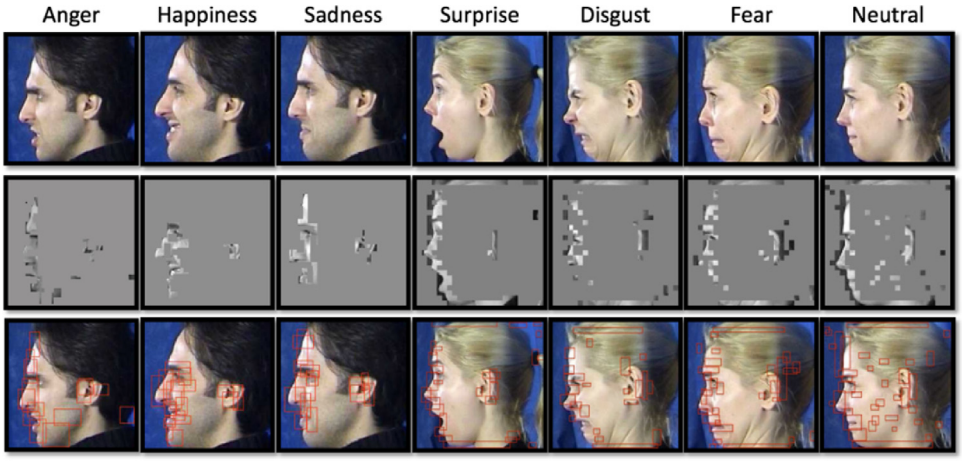

Fig. 8. The sub-regions and their distributions selected by the proposed M-LFA algorithm for each expression of the side-view images with 90-degree rotations extracted from MMI.

expression are also revealed for the 90-degree side-view images to indicate efficiency of the proposed M-LFA algorithm. The extracted features also map closely with the muscular activities recommended by FACS.

Moreover, we have also employed multi-view images from the BU-3DFE database to further assess our feature selection algorithm. The original 750 multi-view images from BU-3DFE are used for evaluation, with 500 images for training and 250 images for test. Table 10 shows the average classification performances of each algorithm in combination with diverse classifiers over 30 runs using multi-view images with 45-degree rotations. As shown in Table 10, the M-LFA algorithm achieves the highest accuracy rates in combination with diverse classifiers. Integrated with the SVM-based ensemble model, it obtains the best average accuracy rate of 100%, and outperforms GA, PSO, LFA, MFO, GM-PSO, LSFA, ODFA, and chaotic FA by 26.75%, 23.89%, 20%, 9.55%, 10.9%, 12.05%, 7.01%, and 3.3%, respectively.

The selected optimized facial sub-regions for each expression of the multi-view images are shown in Fig. 9. These selected optimal sub-regions around the mouth and eye areas associate strongly with the expression of seven emotions. They are highly correlated with the AUs provided in FACS too.

Table 10
Average classification results over 30 runs for multi-view images with 45-degree rotations extracted from BU-3DFE.

Feature selection methods	Number of features	NN (30 runs) %	SVM (30 runs) %	NN-based Ensemble (30 runs) %	SVM-based Ensemble (30 runs) %
GA	135–245	70.15	70.80	73.00	73.25
PSO	150–215	72.33	73.00	75.67	76.11
LFA	60–95	76.75	77.56	79.33	80.00
MFO	45–75	86.45	87.05	89.75	90.45
GM-PSO [15]	40–70	85.77	86.10	88.50	89.10
LSFA [21]	45–90	84.00	84.75	86.65	87.95
ODFA [22]	40–65	89.20	90.00	92.35	92.99
Chaotic FA [26]	40–70	92.50	92.89	94.30	96.70
The proposed M-LFA algorithm	25–55	95.87	96.25	100	100

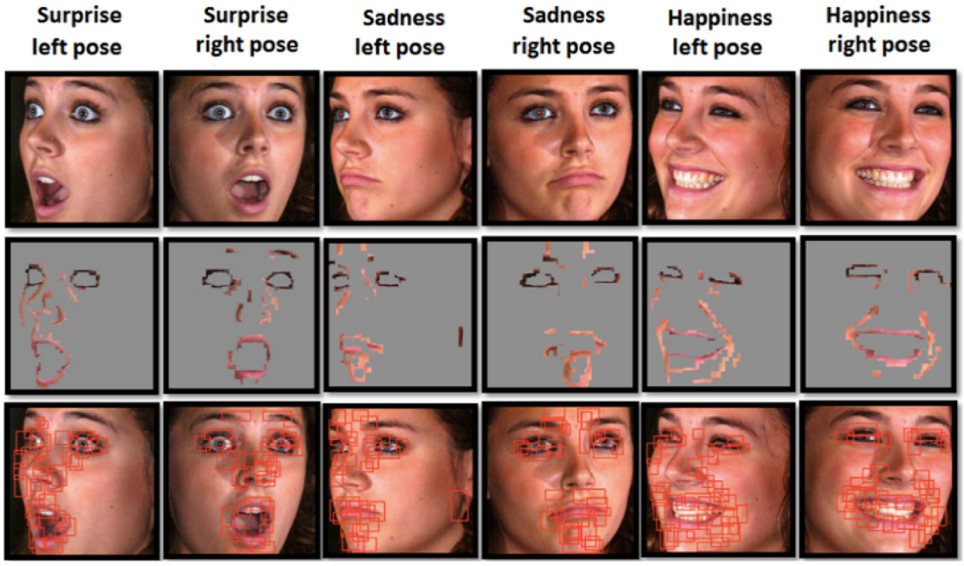

Fig. 9. The sub-regions and their distributions selected by the proposed M-LFA algorithm for each expression of multi-view images with 45-degree rotations extracted from BU-3DFE.

Some theoretical comparison between ODFA [22] and the proposed M-LFA algorithm is conducted, as follows. ODFA employs an opposition-based method for population initialization, which includes generating an opposite population of the original swarm. Then, a dimensional-based method is applied to generate the global best solution by identifying the most optimal value for each dimension individually. This global best solution is then used for updating the position of each firefly in each iteration. Although achieving an efficient computational cost, the search process of ODFA inherits the PSO concept, which is guided by the global best solution in each generation, rather than evolving through a set of optimal solutions based on the attractiveness and attraction behaviours of the original FA model. In addition, ODFA does not provide any mechanism to conduct long and short jumps of the identified global best solution to avoid local optimum. Therefore, it is more likely to lead to premature convergence. Another shortcoming of ODFA is its limitation in dealing with multimodal optimization problems that have multiple best solutions. GM-PSO [15] integrates PSO with the GA and three mutation techniques of Gaussian, Cauchy and Levy distributions to further enhance the swarm leader to identify the discriminative features for bodily expression regression. However, its search strategy relies heavily on the PSO mechanism where the search process is guided by the global best leader rather than multiple promising solutions in the search space, therefore more likely to be trapped in local optima. LSFA [21] incorporates LFA with SA, where the best solution identified by LFA is further enhanced by the SA. However, SA-based local exploitation is only applied to the global best solution. The algorithm does not include any other strategy to increase local exploitation of the overall population. Therefore, the search process lacks of diversity and shows limited capability in escaping from local optima. Moreover, chaotic FA [26] employs CMO for population initialization, in order to increase swarm diversity. A chaotic strategy is also employed by fireflies with a lower light intensity to move towards those with a higher light intensity in the neighbourhood. Especially, the firefly with the highest light intensity purely executes this chaotic movement, rather than a random behaviour, to exploit the search space. However, if this chaotic movement fails to generate a fitter offspring for the current global best leader, there is no search mechanism embedded in the algorithm to drive the search out of the local optimum and to overcome stagnation.

In comparison with the above methods, the proposed M-LFA algorithm employs fireflies and moths to follow multiple attractive solutions in the neighbourhood (rather than purely following the global best leader as in ODFA [22] and GM-PSO [15]) to mitigate premature convergence. It employs two search strategies, i.e. the spiral search of the moths to increase local exploitation of LFA and the attractiveness search actions of the fireflies to cause sudden movement of fireflies and their attached moths to increase global exploration of MFO. In each iteration, each firefly is guided by both search strategies simultaneously to find the optimal solution. These two strategies work cooperatively to overcome the local optimum. When there are no more attractive fireflies in the search space, the spiral search behaviour of the moths exploits the neighbourhood of

Table 11
Comparison between the proposed M-LFA algorithm and non-evolutionary feature selection methods using frontal-view images from CK+, JAFFE and MMI.

Feature selection methods	Average number of selected features	CK+ SVM-based Ensemble %	JAFFE SVM-based Ensemble %	MMI (Frontal-view) SVM-based Ensemble %
mRMR [13]	72	84.00	83.56	79.50
PCA+LDA [9]	61	97.88	97.05	92.33
PCA+LFDA [9]	55	98.75	98.25	92.50
KLFDA [9]	54	99.00	98.88	93.00
The proposed M-LFA algorithm	43	100	100	94.86

Table 12
Comparison between the proposed M-LFA algorithm and non-evolutionary feature selection methods using side-view images with 90-degree rotations from MMI.

MMI (90-degree side-view)		
Feature selection methods	Average number of selected features	SVM-based Ensemble %
mRMR [13]	164	70.85
PCA+LDA [9]	71	80.45
PCA+LFDA [9]	57	81.00
KLFDA [9]	57	83.20
The proposed M-LFA algorithm	50	86.35

Table 13
Comparison between the proposed M-LFA algorithm, and non-evolutionary feature selection methods using the multi-view images with 45-degree rotations from BU-3DFE.

BU-3DFE (45-degree multi-view)		
Feature selection methods	Average number of selected features	SVM-based Ensemble %
mRMR [13]	177	77.33
PCA+LDA [9]	90	96.00
PCA+LFDA [9]	84	96.35
KLFDA [9]	87	97.98
The proposed M-LFA algorithm	36	100

the fireflies, in order to avoid stagnation. On the other hand, when the spiral search process of the moths fails to identify a fitter solution, the attractiveness and attraction behaviours of the fireflies are able to guide the search process towards the optimal region. In addition, SA embedded with Levy flights is used to enable long-jump of the most optimal solution to avoid stagnation. The above search mechanism leads to superiority of the proposed algorithm over other state-of-the-art PSO and FA variants, i.e. GM-PSO [15], LSFA [21], ODFA [22], and chaotic FA [26]. Most importantly, our algorithm extends the natural multimodal optimization characteristics of the original LFA, and shows efficient abilities in dealing with multimodal optimization problems.

4.3. Comparison with Non-evolutionary feature selection methods

Besides the above theoretical evaluation and empirical comparison against evolutionary feature optimization methods, we further compare M-LFA with non-evolutionary feature selection techniques presented in [9] and [13]. As discussed earlier, Ali et al. [9] employed three feature dimensionality reduction techniques to identify the most discriminative features for each expression, i.e., PCA+LDA, PCA+LFDA, and KLFDA. Zhang et al. [13] developed a shape-based facial expression recognition system with mRMR-based feature selection. We have also implemented both non-evolutionary feature selection methods described in [9] and [13], i.e. mRMR, PCA+LDA, PCA+LFDA, and KLFDA, for comparison. In each experiment, we use the proposed LBP descriptor to extract initial facial features, and then apply each of the above feature selection methods for feature optimization. Table 11 shows the results for each set of 175 frontal-view images extracted from CK+, MMI, and JAFFE respectively. Note that 250 images from CK+ are used for training. We have also compared M-LFA with the above methods using the 90-degree side-view images from MMI, with 70 images for training and another 70 images for test. Multi-view images from BU-3DFE are also used for evaluation, with 500 and 250 images for training and test, respectively. The results from the side-view images from MMI and multi-view images from BU-3DFE are provided in Tables 12 and 13, respectively.

Theoretical comparison between M-LFA and the above non-evolutionary feature selection methods is conducted, as follows. Although mRMR is a popular feature optimization method, according to Zeng et al. [32], the incremental search scheme of mRMR only selects one feature at a time without considering the interaction between groups of features. Therefore, the experiments using the mRMR-based feature selection method shown in Tables 11–13 yield the least promising performance. As indicated in Tables 11–13, the three related methods, i.e. PCA+LDA, PCA+LFDA, and KLFDA in [9], show competitive performances to those achieved by the proposed M-LFA algorithm. However, a comparatively larger number of features are selected by the three methods than those selected by M-LFA for the above evaluation. Especially, for the evaluation of multi-view images from BU-3DFE, our algorithm selects a dramatically smaller number of 36 features in comparison with 90, 84, and 87 recommended by PCA+LDA, PCA+LFDA and KLFDA, respectively. The empirical results also indicate that the extracted smaller sets of features by our algorithm show more discriminative capabilities and reveal more relevant information pertaining to a specific emotion category owing to the spiral-based local exploitation and attraction-based global exploration. Therefore it outperforms PCA+LDA, PCA+LFDA and KLFDA consistently for the evaluation of frontal-view, multi-view and side-view images from diverse databases. Moreover, the performance of LFDA and KLFDA in [9] also relies heavily on the choice of the affinity matrix and the optimal selection, and parameter settings of the kernel functions respectively [33]. However, the proposed M-LFA algorithm does not require such complex kernel settings with smaller but more discriminating feature subsets identified. Therefore, M-LFA has better computational efficiency to fulfil real-time application requirements.

Furthermore, we compare the computation efficiency of our algorithm and all other metaheuristic and non-evolutionary feature selection algorithms in Table 14. The computational cost for each method includes the execution of the proposed LBP descriptor for feature extraction and the features selected by each corresponding method and the SVM-based ensemble classifier. This setting is selected because of its impressive performances for diverse test cases

Table 14
The computational costs of the proposed system and other related methods.

Feature selection methods applied	Average number of selected features	Computational cost (milliseconds)
GA	180	380
PSO	154	362
LFA	62	260
MFO	70	280
GM-PSO [15]	61	260
LSFA [21]	68	270
ODFA [22]	76	320
Chaotic FA [26]	57	250
mRMR [13]	72	285
PCA+LDA [9]	61	262
PCA+LFDA [9]	55	250
KLFDA [9]	54	250
The proposed M-LFA algorithm	43	235

for all methods. The results shown in Table 14 are obtained by averaging the computational efficiency and the number of selected features across all the testing conducted using frontal images from CK+, JAFFE and MMI.

As shown in Table 14, the computational costs are closely linked with the number of selected features for classification. As an example, we observe that GA and PSO have the longest processing time because larger sizes of features are selected, i.e. 180 for GA and 154 for PSO. The numbers of features extracted by ODFA (76), mRMR (72), MFO (70), LSFA (68), LFA (62), GM-PSO (61), PCA+LDA (61), PCA+LFDA (55) and KLFDA (54) are smaller, and they have lower computational costs. Comparatively, the proposed M-LFA algorithm has the smallest number of features (43) and the lowest computational cost.

4.4. Comparison with other facial expression recognition methods

We have compared our algorithm with other state-of-the-art facial expression recognition methods using CK+, MMI, and JAFFE. Table 15 shows the comparison results between our algorithm and other related methods with the CK+ database.

As shown in Table 15, among different related methods, Ali et al. [9], Zhang et al. [8], and Neoh et al. [12] achieved the highest accuracy rates with the CK+ database. Ali et al. [9] employed the non-linear technique of EMD to extract the initial features. Integrated with KLFDA and ELM-RBF, their proposed method achieved the highest recognition accuracy rate of 99.75%, when evaluated with the cross-validation scheme. Their results indicated that sometimes the “sad” expression could be misclassified as “surprise”. Zhang et al. [8] learnt a joint representation from the texture and landmark modalities to enhance expression recognition, and achieved an impressive average recognition accuracy rate of 99.3% over 5 runs using the hold-out validation method. However, their work required neutral images as references and also a sequence of six images as inputs for expression recognition. Neoh et al. [12] proposed a layered encoding cascade optimization model for facial expression recognition. Their feature optimization process revealed significant emotional facial texture information, and achieved an impressive accuracy rate of 97.4%. However, their multi-objective optimization strategy sometimes disregarded important mouth-related features associated with emotional expressions (e.g. lip tightener for anger). In comparison with the above-mentioned state-of-the-art methods, our proposed algorithm reveals significant discriminative facial features, correlated strongly with the facial AUs given in FACS, and achieves the highest average accuracy rate of 100% over 30 hold-out validations.

Table 16 illustrates the performance comparison between our algorithm and related methods using the JAFFE database. Ali et al. [9] employed three feature reduction techniques (PCA+LDA, PCA+LFDA, and KLFDA) in combination with EMD-based feature extraction, and achieved the highest accuracy rate of 100% using 10-fold cross validation. When trained with CK+ and tested with JAFFE, Shan et al. [34] employed boosted LBP and RBF-based SVM, and achieved an accuracy rate of 41.3%. In comparison with other methods, the proposed system achieves 100% accuracy over 30 runs when trained and tested using CK+ and JAFFE, respectively. Again, the result indicates superiority and robustness of our proposed algorithm.

We have also conducted performance comparison between our algorithm and related methods using the MMI database. As shown in Table 17, employing CK+ and MMI for training and test respectively, Liu et al. [41], Fan and Tjahjadi [36], and Shan et al. [34] obtained accuracy rates of 72.2%, 58.7%, and 51.1% using 10-fold cross validation, respectively. Zhong et al. [42] achieved F1-measure of 77.39% when trained and tested with MMI. In comparison with these methods, our algorithm achieves the highest average accuracy rate of 94.86% over 30 hold-out validations when trained with CK+ and tested with MMI. The proposed LBP descriptor for feature extraction and the M-LFA algorithm for feature optimization lead to superiority of our system.

4.5. Real-life deployment and evaluations

In this research, the proposed facial expression recognition system has been deployed in real-life settings to further ascertain its efficiency. We present the following case studies to address the practicality and robustness of the proposed system.

First of all, we integrate the proposed facial expression system with a vision-enriched intelligent virtual agent for health and emotion well-being monitoring for the elderly [43]. This intelligent agent has been developed to conduct object recognition, object/human attribute prediction (e.g. shape and colour for objects, and age and gender for human subjects), scene classification and facial expression recognition using live video stream inputs collected by the built-in camera of a personal computer or a tablet to perform health and emotion well-being monitoring [43]. The facial expression recognition function embedded in this intelligent agent has been performed by a basic version of our previous work [43]. The proposed facial expression recognition system has been used to replace our previous version and to work with the intelligent agent in facial expression perception in real-life settings. Speech recognition and synthesis functions are also integrated in the intelligent agent. The popular online encyclopaedia Wikipedia is also equipped to enable the conversation of the intelligent agent to literally cover any topics. User evaluation has been carried out with 10 subjects (4 British male, 2 British female, and 4 Indian male) aged between 20–30 to assess the newly integrated facial expression recognition function. The user evaluation session starts with greetings and an informal chat with the agent. Then, each user is asked to either pose the seven facial expressions, or show spontaneous expressions during their conversation with the facial expression classification results reported back to the user through a speech synthesis function. Overall, 70 (10 subjects * 7 emotions) facial expressions are captured and evaluated by the proposed system in real time. It achieves the following accuracy for each emotion category, i.e. 80% for “anger”, “fear”, and “sadness”, 100% for both “happiness” and “surprise”, and 90% for “disgust” and “neutral”, respectively. This intelligent health and emotion well-being monitoring system with posed and spontaneous facial expression recognition has been demonstrated in industrial showcases successfully. The above real-life deployment has also proved the superiority and efficiency of the proposed system in real-life settings. The system will also be further evaluated intensively by elderly

Table 15
Performance comparison using the CK+ database.

Studies	Methodology	Classes	Type of data	Evaluation Strategy	Recognition rate (%)
Zhang et al. [8]	Multimodal learning	6	Dynamic	5 hold-out validations with 33.3% for testing for each run	99.3
Ali et al. [9]	EMD+KLFDA+ELM-RBF	7	Static	10-fold cross validation	99.75
Neoh et al. [12]	Overlap LGBP+Pareto+Weighted majority vote	7	Static	42.8% for testing	97.4
Shan et al. [34]	Boosted LBP+SVM	7	Static	10-fold cross validation	91.40
Zhang et al. [35]	Facial landmark detection+neural networks	7	Static	Hold-out validation	75.83
Fan and Tjahjadi [36]	PHOG_TOP+Dense optical flow+SVM	7	Dynamic	Leave-one-out cross validation	83.7
Zhang et al. [37]	Unsupervised facial point detector+fuzzy c-means	7	Static	Hold-out validation	91.86
This research	The proposed LBP+M-LFA+ SVM-based ensemble	7	Static	30 hold-out validations with 46.6% for testing for each run	100

Table 16
Performance comparison using the JAFFE database.

Related Work	Methodology	Classes	Type of data	Evaluation Strategy	Recognition Rate (%)
Ali et al. [9]	EMD+KLFDA+ KNN/SVM/ELM-RBF	7	Static	10-fold cross validation	100
Shan et al. [34]	Boosted LBP+SVM	7	Static	10-fold cross validation	41.3 (trained with CK+ and tested with JAFFE)
Rahulamathavan et al. [38]	LFDA+kNN	7	Static	Leave-one-out cross validation	94.37
Zhang and Tjondronegoro [39]	Patch-based Gabor+SVM	6	Static	Leave-one-out cross validation	93.48
Zhao and Zhang [40]	LBP+ kernel discriminant isomap+ nearest neighbor	7	Static	Cross-validation	81.59
This research	The proposed LBP+M-LFA+ SVM-based ensemble	7	Static	30 hold-out validations	100 (trained with CK+ and tested with JAFFE)

Table 17
Performance comparison using the MMI database.

Related Work	Methodology	Classes	Type of data	Evaluation Strategy	Recognition Rate (%)
Shan et al. [34]	Boosted LBP+SVM	7	Static	10-fold cross validation	51.1(trained with CK+ and tested with MMI)
Fan and Tjahjadi [36]	PHOG_TOP+Dense optical flow+SVM	6	Dynamic	10-fold cross validation	58.70 (trained with CK+ and tested with MMI)
Liu et al. [41]	AUDN	6	Static	10-fold cross validation	72.2 (trained with CK+ and tested with MMI)
Zhong et al. [42]	Multitask sparse learning	6	Static	F1 MEASURE	77.39
This research	The proposed LBP+M-LFA+ SVM-based ensemble	7	Static	30 hold-out validations	94.86 (trained with CK+ and tested with MMI)

users in real-life settings with the support of UK industrial partners, NHS, and Age UK.

Another similar real-life deployment has also been conducted by integrating the proposed system with the C++ SDK of the latest humanoid NextGen H25 NAO robot, in order to bring benefits to real-life human robot interaction [29,35,37]. This humanoid robot has a powerful CPU processor and built-in camera sensors to enable real-time vision-based processing and facial expression perception. Related applications can also be found in our previous studies [29,35,37]. The proposed facial expression recognition system shows great potential to contribute to such intelligent service robot development for personalised healthcare and intelligent tutoring applications. Furthermore, the proposed LBP descriptor for feature extraction and M-LFA for feature optimization can also be easily applied to bioinformatics applications such as MRI brain tumour image classification and blood cancer detection from microscopic images [44–47]. As an instance, the proposed LBP descriptor is able to extract initial features from the MRI brain tumour or

microscopic lymphocytic images. Then, the most significant shape, colour, and texture features associated with the tumour region or the nucleus and cytoplasm of lymphocytes/lymphoblasts can be retrieved by the proposed M-LFA algorithm to improve classification accuracy.

5. Conclusions

We have proposed a new LBP descriptor for discriminative feature extraction and a novel LFA variant for feature optimization. Diverse classifiers have been employed for recognition of seven expressions. The proposed LBP descriptor integrates LBP, LGBP, and LBPV to extract local spatial patterns and contrast measures for texture description, in order to better deal with illumination changes, rotations, and scaling differences. The proposed M-LFA feature selection algorithm benefits from local exploitation of moths and attractiveness behaviours of fireflies simultaneously to identify local and global optimal solutions. It employs spiral search of the moths to increase local exploitation of LFA and the attractiveness search actions of the fireflies to cause sudden optimal movement of the fireflies and their attached moths to increase global exploration of MFO. SA-embedded Levy flights search diversification has also been used to increase exploitation of the current global best solution. Evaluated with the frontal-view images from CK+, MMI, and JAFFE and the multi-view and side-view images from BU-3DFE, and MMI respectively, the proposed system outperforms other state-of-the-art metaheuristic search and non-evolutionary feature selection methods by a significant margin. Moreover, the proposed system outperforms other state-of-the-art facial expression recognition methods reported in the literature significantly.

For future research, we will evaluate the proposed algorithm with diverse multimodal optimization problems to further ascertain its efficiency. To increase the convergence speed, adaptive parameter setting will also be explored to enable key parameters such as the randomization parameter in the attractiveness function to decrease gradually during the iterative process while approaching the global optima. Such a dynamic parameter setting enables the search process to have sufficient diversity in early generations as well as the capability of fine-tuning the solutions in final iterations. Moreover, we aim to evaluate the proposed algorithm in tackling multi-objective optimization problems. In terms of applications, the proposed algorithm will be also used for complex computer vision tasks, such as object tracking in video sequences and salient object detection, and bioinformatics applications such as retinal and skin disease detection.

References

[1] X.S. Yang, Firefly algorithm, levy Flights and global optimization, Res. Develop. Intell. Syst. 26 (2010) 209–218.
[2] S. Mirjalili, Moth-Flame optimization algorithm: A novel nature-inspired heuristic paradigm, Knowl. Based Syst. 89 (2015) 228–249.
[3] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recognit. 43 (3) (2010) 706–719.
[4] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended Cohn-Kanade dataset (CK+): a complete expression dataset for action unit and emotion-specified expression, in: Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis, USA, San Francisco, 2010, pp. 94–101.
[5] M. Pantic, M.F. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, in: Proceedings of IEEE Int'l Conf. Multimedia and Expo, The Netherlands, Amsterdam, 2005, pp. 317–321.
[6] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with gabor wavelets, in: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 200–205.
[7] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3D facial expression database for facial behavior research, in: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 211–216.
[8] W. Zhang, Y. Zhang, L. Ma, J. Guan, S. Gong, Multimodal learning for facial expression recognition, Pattern Recognit. 48 (10) (2015) 3191–3202.
[9] H. Ali, M. Hariharan, S. Yaacob, A.H. Adom, Facial emotion recognition using empirical mode decomposition, Expert Syst. Appl. 42 (3) (2015) 1261–1277.
[10] S. Shojaeilangari, W.Y. Yaub, E.K. Teoh, A novel phase congruency based descriptor for dynamic facial expression analysis, Pattern Recognit. Lett. 49 (2014) 55–61.
[11] M.H. Siddiqi, R. Ali, A.M. Khan, Y.T. Park, S. Lee, Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields, IEEE Trans. Image Process. 24 (4) (2015) 1386–1398.
[12] S.C. Neoh, L. Zhang, K. Mistry, M.A. Hossain, C.P. Lim, N. Aslam, P. Kinghorn, Intelligent facial emotion recognition using a layered encoding cascade optimization model, Appl. Soft Comput. 34 (2015) 72–93.
[13] Y. Zhang, L. Zhang, M.A. Hossain, Adaptive 3D facial action intensity estimation and emotion recognition, Expert Syst. Appl. 42 (3) (2015) 1446–1464.
[14] X.S. Yang, Nature Inspired Metaheuristic Algorithm (2008).
[15] Y. Zhang, L. Zhang, S.C. Neoh, K. Mistry, A. Hossain, Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles, Expert Syst. Appl. 42 (22) (2015) 8678–8697.
[16] J. Holland, Adaptation in Natural and Artificial Systems, MIT Press, Cambridge, MA, USA, 1975.
[17] J. Kennedy, R. Eberhart, Particle Swarm Optimization, in: Proceedings of IEEE Int. Conf. Neural Networks, 4, 1995, pp. 1942–1948.
[18] S. Kirkpatrick, C.D. Gelatt Jr, M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680.
[19] B. Xue, M. Zhang, W.N. Browne, Particle swarm optimization for feature selection in classification: a multi-objective approach, IEEE Trans. Cybern. 43 (6) (2013) 1656–1671.
[20] M. Goodarzi, L.S. Coelho, Firefly as a novel swarm intelligence variable selection method in spectroscopy, Anal. Chim. Acta. 852 (2014) 20–27.
[21] M. Alweshah, S. Abdullah, Hybridizing firefly algorithms with a probabilistic neural network for solving classification problems, Appl. Soft Comput. 35 (2015) 513–524.
[22] O.P. Verma, D. Aggarwal, T. Patodi, Opposition and Dimensional based modified firefly algorithm, Expert Syst. Appl. 44 (2016) 168–176.
[23] L. dos Santos Coelho, D.L. de Andrade Bernert, V.C. Mariani, A chaotic firefly algorithm applied to reliability-redundancy optimization, in: Proceedings of IEEE Congress of Evolutionary Computation (CEC), 2011, pp. 517–521.
[24] A. Abdullah, S. Deris, M. Mohamad, S. Hashim, A new hybrid firefly algorithm for complex and nonlinear problem, Distrib. Comput. Artif. Intell. 151 (2012) 673–680.
[25] M. Arsuaga-Ríos, M.A. Vega-Rodríguez, Multi-objective firefly algorithm for energy optimization in grid environments, *Swarm Intell.* LNCS 7461 (2012) 350–351.
[26] A. Kazem, E. Sharifi, F.K. Hussain, M. Saberic, O.K. Hussain, Support vector regression with chaos-based firefly algorithm for stock market price forecasting, Appl. Soft Comput. 13 (2) (2013) 947–958.
[27] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on featured distribution, Pattern Recognit. 29 (1) (1996) 51–59.
[28] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: Proceedings of 10th IEEE Int. Conf. Computer Vision, 2005, pp. 786–791.
[29] K. Mistry, L. Zhang, S.C. Neoh, C.P. Lim, B. Fielding, A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition, IEEE Trans. Cybern. 99 (2016) 1–14.
[30] OpenCV Tutorials. http://docs.opencv.org/2.4/doc/tutorials/tutorials.html, 2016 (assessed 03.07.16).
[31] P. Ekman, W.V. Friesen, J.C. Hager, Facial action coding system, the manual, Research Nexus division of Network Information Research Corporation, USA, 2002.
[32] Z. Zeng, H. Zhang, R. Zhang, Y. Zhang, A hybrid feature selection method based on rough conditional mutual information and Naive Bayesian classifier, ISRN Appl. Math. (2014) 1–11 Article ID 382738.
[33] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, J. Mach. Learn. Res. 8 (2007) 1027–1061.
[34] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vis. Comput. 27 (6) (2009) 803–816.
[35] L. Zhang, M. Jiang, D. Farid, A. Hossain, Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot, Expert Syst. Appl. 40 (13) (2013) 5160–5168.
[36] X. Fan, T. Tjahjadi, A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences, Pattern Recognit. 48 (11) (2015) 3407–3416.
[37] L. Zhang, K. Mistry, M. Jiang, S.C. Neoh, A. Hossain, Adaptive facial point detection and emotion recognition for a humanoid robot, Comput. Vis. Image Understand. 140 (2015) 93–114.
[38] Y. Rahulamathavan, R.C.-W. Phan, J.A. Chambers, D.J. Parish, Facial expression recognition in the encrypted domain based on local fisher discriminant analysis, IEEE Trans. Affective Comput. 4 (1) (2013) 83–92.
[39] L. Zhang, D.W. Tjondronegoro, Facial expression recognition using facial movement features, IEEE Trans. Affect. Comput. 2 (4) (2011) 219–229.
[40] X. Zhao, S. Zhang, Facial expression recognition based on local binary patterns and kernel discriminant isomap, Sensors (Basel) 11 (10) (2011) 9573–9588.

[41] M. Liu, S. Li, S. Shan, X. Chen, AU-inspired deep networks for facial expression feature learning, Neurocomputing 159 (2015) 126–136.
[42] L. Zhong, Q. Liu, P. Yang, J. Huang, D.N. Metaxas, Learning multiscale active facial patches for expression analysis, IEEE Trans. Cybern. 45 (8) (2015) 1499–1510.
[43] L. Zhang, B. Fielding, P. Kinghorn, K. Mistry, A vision enriched intelligent agent with image description generation, in: Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Singapore, 2016, pp. 1488–1490.
[44] G. Jothi, H.H. Inbarani, Hybrid tolerance rough set–firefly based supervised feature selection for mri brain tumor image classification, Appl. Soft Comput. 46 (2016) 639–651.
[45] S.C. Neoh, W. Srisukkham, L. Zhang, S. Todryk, B. Greystoke, C.P. Lim, A. Hossain, N. Aslam Nature Publishing Group., An intelligent decision support system for leukaemia diagnosis using microscopic blood images, Sci. Rep. 5 (14938) (2015) 1–14.
[46] A. Bourouis, M. Feham, A. Hossain, L. Zhang, An intelligent mobile based decision support system for retinal disease diagnosis, Decis. Support Syst. 59 (2014) 341–350.
[47] T.Y. Tan, L. Zhang, M. Jiang, An intelligent decision support system for skin cancer detection from dermoscopic images, In *Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2016)*, China, In Press.