Feature Selection using Particle Swarm Optimization for Thermal Face Recognition

Ayan Seal, Suranjan Ganguly, Debotosh Bhattacharjee, Mita Nasipuri and Consuelo Gonzalo-Martin

Abstract This paper presents an algorithm for feature selection based on particle swarm optimization (PSO) for thermal face recognition. The total algorithm goes through many steps. In the very first step, thermal human face image is preprocessed and cropping of the facial region from the entire image is done. In the next step, scale invariant feature transform (SIFT) is used to extract the features from the cropped face region. The features obtained by SIFT are invariant to object rotation and scale. But some irrelevant and noisy features could be produced with the actual features. Unwanted features have to be removed. In other words, optimum features have to be selected for better recognition accuracy. Since PSO is an optimization method, which works with the principle of local as well as global searches for finding optimum set of features. Here, this process has been implemented to select a subset of features that effectively represents original feature extracted for better classification convergence. Finally, minimum distance classifier is used to find the class label of each testing images. Minimum distance classifier acts as an objective function for PSO. In this work, all the experiments

A. Seal (✉) · S. Ganguly · D. Bhattacharjee · M. Nasipuri
Department of Computer Science and Engineering, Jadavpur University, Kolkata, India
e-mail: ayan.seal@gmail.com

S. Ganguly
e-mail: suranjanganguly@gmail.com

D. Bhattacharjee
e-mail: debotoshb@hotmail.com

M. Nasipuri
e-mail: mitanasipuri@yahoo.com

A. Seal · C. Gonzalo-Martin
Center for Biomedical Technology, Universidad Politecnica de Madrid, Madrid, Spain
e-mail: consuelo.gonzalo@upm.es

R. Chaki et al. (eds.), *Applied Computation and Security Systems*, Advances in Intelligent Systems and Computing 304, DOI 10.1007/978-81-322-1985-9_2

have been performed on UGC-JU thermal face database. The maximum success rate of 98.61 % recognition has been achieved using SIFT and PSO for frontal face images and 90.28 % for all images.

Keywords Face recognition · Infrared face images · Scale invariant feature transform · Particle swarm optimization

1 Introduction

Among all biometric identification systems, face recognition is one of the most suitable methods due to its non-intrusive nature. Face recognition has appeared as one of the most exciting research problem due to its numerous practical applications. It is extensively used in the authentication, surveillance, security, and human computer interaction purpose. But the existing face recognition systems have several limitations. The performance of face recognition is outstanding in case of controlled conditions. The performances degrade significantly in an uncontrolled environment. The main reason is that the most of the existing face recognition methods are based on visual images. The quality of the visual images changes with lighting condition, as a result, the performances degrade. That means the performance of the face recognition system, based on visual images, captured in daylight situation is not same with the performance based on nightlight vision. This problem is known as illumination problem. This problem can be solved by the uses of thermal face images instated of visual images. A thermal face image is captured by the thermal infrared camera, and a thermal infrared camera is illumination independent. It can capture nearly same images in all the environments even in a dark situation because a thermal infrared camera concentrates on emitted energy from the object surface. It does not consider the environment temperature. Another significant advantage of face recognition based on thermal face images is that the tasks of face detection, location, and segmentation are relatively easier and more reliable than their visual counterpart [1].

Some general face recognition methods include eigenfaces [2, 3] and Fisher's discriminant analysis [4], which are sensitive to illumination and different facial expressions. These methods are also used for dimensionality reduction, when the dimensionality is curse for some applications. Sometimes, LDA gives a better result than eigenfaces, but it does not give the robust solution as their separable criterion is not relevant to classification precision [12]. Wavelet transform [5] is very good tool to analyze texture pattern in time and frequency domain, and these techniques work well for frontal faces only. But they are not robust against rotation variations because the whole-face-based process is highly sensitive toward translations and rotations. Another limitation of these approaches is that the size of the feature vector is too large to recognize. So these methods are computationally expensive.

Generally, a face image is represented by large number of features using many feature extraction methods. Some of the features have more discriminating power than the others. In other words, all features do not contribute equally to the face recognition process. It is not always true that the higher number of features lead to higher recognition rate. So optimum feature selection is a big issue in pattern recognition domain. Optimum feature selection has many advantages. It reduces the feature size and increases the recognition rate. Feature selection process first identifies the irrelevant features, discards them, and takes others which are treated as optimum features. Thus, feature selection is basically a search process. There are different search algorithms such as greedy [13], branch and bound [14], sequential search algorithms [15], mutual information [16], and tabu search [17] which have been used successfully. But these algorithms are quite computationally expensive. There are other kinds of population-based search algorithms in the literature, which is less expensive in times. Such algorithms are Genetic Algorithm (GA)-based method [18, 19, 20], and ant colony optimization (ACO)-based techniques have attracted a lot of attention [21] to the researchers.

In this paper, we propose an efficient scheme for scale and rotation invariant thermal face recognition using scale invariant feature transform (SIFT). The main contributions of this work are as follows:

- A complete scale and rotation invariant thermal face recognition system based on SIFT features are implemented.
- PSO-based feature selection algorithm is developed to search for the optimal features and to increase the recognition rate, as well.
- Evaluation of the proposed system using the UGC-JU thermal face database and comparing its performance with other FR techniques.

The outline of this paper is organized as follows: Sect. 2 describes the different steps of proposed approach including image preprocessing, features extraction, selection of features, and classification. The experiment and results are presented in Sect. 3. Finally, Sect. 4 concludes and remarks about some of the aspects analyzed in this paper.

2 Proposed Method

In this section, we have introduced a robust method for face recognition using thermal face images. The overall system includes image preprocessing, extraction of features, feature selection, and classification. All of these elements are explained in the following sections.

2.1 Preprocessing

Preprocessing is the first step of the proposed system. In this step, an intermediate image is produced from the 'just captured raw' image. A typical 24-bit color thermal face image is shown in Fig. 1. Here, each 24-bit color images have been converted into its corresponding grayscale images. Then, those converted grayscale images are again converted into binary images counterparts. After conversion of the binary image, it has been found that some white segments are present with a larger one in the binary image and this biggest part is the face area. So the largest part has been extracted from the binary image using connected component labeling algorithm [6], and other small components, which are other parts of the binary image, have been excluded. It has also been seen that some holes are created in the face area due to uneven distribution of thermal information which is nothing but temperature statistics. These temperature statistics has been excluded in the binarization process, which will make the face recognition process tougher.

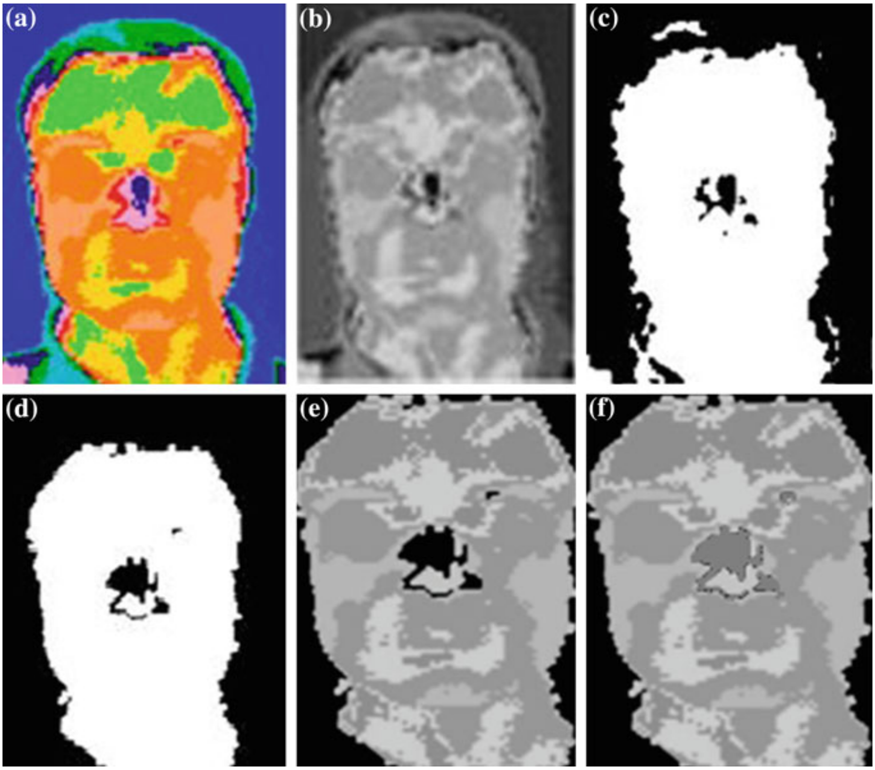


Fig. 1 The various outcomes of the preprocessing step. **a** A thermal face image. **b** Grayscale image. **c** Binary image. **d** Largest component as a face skin area. **e** Extracted face skin area in *gray* level. **f** Restored image

Principal component analysis (PCA) is reported to be robust for the problem of contaminated pixels [7]. But it cannot restore the lost information. So GPCA [8, 9] method is used to store the missing temperature statistic information. The outcomes of the preprocessing step are shown in Fig. 1.

2.2 Extraction of Features

The performance of any face recognition system highly depends on the selected features. So a good feature extraction algorithm is very necessary for face recognition which helps to identify probably the best discriminating power and which are less sensitive to variations in pose, scale, and illumination, and facial expressions etc. SIFT [10] is one such algorithm. SIFT is used for extraction of distinctive invariant features from the objects, which can be used to carry out the matching process. The features obtained by SIFT are invariant to object rotation and scale. It reduces the probability of reduced extraction due to occlusion and noise. First Gaussian function, $G(x, y, \sigma)$, is used to convolve with an input image $I(x, y)$ in order to get a scale-space image, $L(x, y, \sigma)$ image. The Gaussian function and scale-space image are found by the Eq. (1) and (2), respectively.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)} \tag{1}$$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2}$$

where ‘*’ is the convolution operation in the x and y directions. So the initial image is incrementally convolved with Gaussians to produce images separated by a constant factor k in scale-space. Here, each scale-space is divided into ‘s’ equal intervals, where ‘s’ is a natural number and hence $k = 2^{1/s}$. Then, difference-of-Gaussian function convolved with the image, $D(x, y, \sigma)$, that can be computed from the difference of two nearby scales separated by a constant multiplicative factor k is calculated for the detection of efficient, stable landmarks. The difference-of-Gaussian function is shown in Eq. (3).

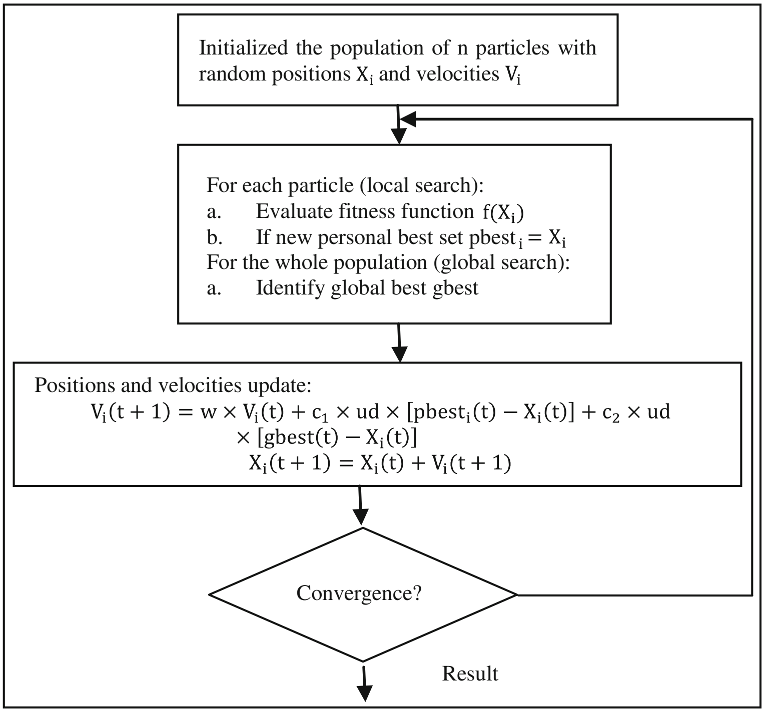
$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \tag{3}$$

Now, local maxima and local minima have been found from the image, $D(x, y, \sigma)$. First, a sample point from the current image has been chosen and compared to its eight neighbors in the current image and nine neighbors in the scale image above and below. It is chosen as a landmark only if it is larger (local maxima) than all of these neighbors or smaller (local minima) than all of them.

2.3 Feature Selection

After extraction of features, it has found that, within the extracted features, there are some features, which are irrelevant and noisy. These irrelevant and noisy features lead the misclassification rate. So the objective of feature selection step is to reduce the noisy data and exclude the irrelevant features as much as possible. In other word, find the optimal features from the original features including noisy and irrelevant features, which have higher discriminating power, to improve the recognition rate. Particle swarm optimization (PSO) is one such well-known tool to find the optimum characteristics with the help of local as well as global search in the feature search space in an iterative way. PSO proposed by Dr. Eberhart and Dr. Kennedy in 1995 [11]. In PSO, swarm consists of a group of random particles, which move around the solution space of the problem by updating through iterations for an optimum solution and go until convergence is achieved. A flowchart of the PSO-based system is given below:

Flowchart 1. PSO based feature selection



In this work, '*n*' number of random particles is chosen initially from the features space. Each particle having *c* parameters that are obtained, after feature extraction using SIFT operator, and their corresponding random velocities form a position matrix $X[n, c]$. Now, the threshold should be selected for the first round of selection of these random velocities and its corresponding positions by the following functions $V[i, j] = \mathrm{e}(X[i, j])$ where $1 \leq i \leq n$ and $1 \leq j \leq c$ and it is assumed to be 0.5 for this work. The velocity of the ith particle is described by the $V_i = (v_{i1}, v_{i2}, \ldots, v_{ic})$, and its corresponding state is represented by $X_i = (x_{i1}, x_{i2}, \ldots, x_{ic})$. If the newly computed velocity is greater than the threshold value (0.5), then this velocity and its location is selected for the next iteration. It is expected that, after each iteration, the recognition rate of the face recognition system increases with the newly selected features from the features space. So the success rate is calculated by an objective function known as the fitness function in PSO. The minimum distance function [5] is used here as a fitness function for this work. Here, minimum distance classifier concentrates both local and as well as global information of the features obtained from SIFT operator. The fitness function is evaluated for each particle in the swarm and is compared to the fitness of the best previous (pbest) result for that particle and to the fitness of the best particle (gbest) among all particles in the swarm. After finding the two best values (pbest and gbest), the particles start updating their velocities and positions according to the Eqs. (4) and (5), respectively.

$$V_i(t+1) = w \times V_i(t) + c_1 \times ud \times [\mathrm{pbest}_i(t) - X_i(t)] + c_2 \times ud \times [\mathrm{gbest}(t) - X_i(t)] \tag{4}$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{5}$$

where 'i' = 1, …, n and 'n' is the population size, 'ud' is another random number between 0 and 1, 'c_1' and 'c_2' are cognitive and social parameters, respectively, bounded between 0 and 1. In the velocity update equation, the + sign divides the whole equation into three components named as inertial component, a cognitive component, and social component, respectively. The inertia weight w is a factor used to control the balance of the search algorithm between exploration (=0.15) and exploitation (=1); the second element is the 'cognitive' section representing the local knowledge of the particle itself; the third component is the 'social' part, representing the cooperation among the particles. The iterative steps will go on until the process reaches the termination condition. It is experimentally found that thirty iterations are well enough to identify the optimum features from the features space, which leads the success rate to a great extent.

3 Experiment and Results

The performance of the proposed system is evaluated on newly created UGC-JU thermal face database [9]. UGC-JU thermal face database consists of 84 subjects each is having 39 different face images with Exp1 (happy), Exp2 (angry), Exp3 (sad), Exp4 (disgusted), Exp5 (neutral), Exp6 (fearful), and Exp7 (surprised) are considered. Different pose changes about x-axis, y-axis, and z-axis are also

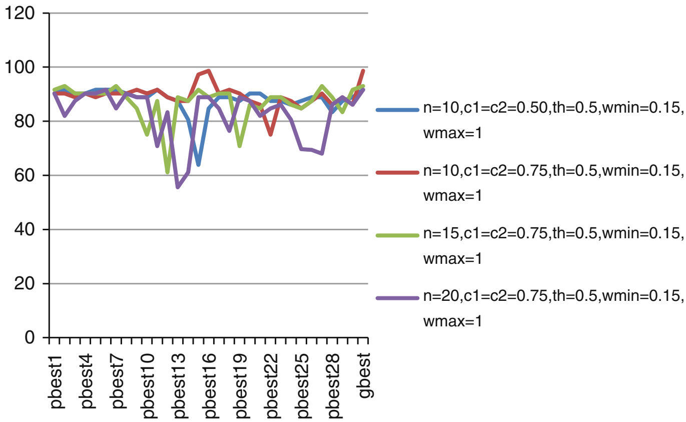


Fig. 2 The different values of ‘pbest’ and ‘gbest’ with different parameters of PSO for frontal faces

considered. Resolution of each image is 320×240, and the images are saved in JPEG format. First, all are processed and restored missing information using GPCA scale to 256×256 resolution for further processing. Then, SIFT is used to extract some points. After extraction of some points, total image is divided into several blocks of size 16×16 and find the number of points in each block and store them in a row vector. Hence, one row corresponds to one image and column corresponds to feature. So total number of features of a particular face or a row is $(256 \times 256)/(16 \times 16) = 256$. But the 256 extracted features contain some noisy and irrelevant features. As a result, the performance of the system degrades. So selection of features which are relevant, free from noise, and redundancy is essential to improve the performance of the thermal face recognition system. PSO is used to find the optimal features. Ten optimum features are taken from 256 features of a face image using PSO so that they have enough discriminating power. Within this feature selection step, minimum distance classifier is used as a fitness function which identifies the class label of the testing images in successive iterations. Total four sets of experiments have been performed. First set of experiments has been performed on frontal thermal face images which includes various facial expressions. In this experiment, features coming from SIFT operator are considered and PSO is not used. In the second set of experiments, all thermal face images including pose changes about x-axis, y-axis, and z-axis, and occlusion with the frontal thermal face images are considered, and the features selection tool is not used. So, in the above two cases, the size of the features of each face image is 256. In the third set of experiment, SIFT and PSO are used for feature extraction and feature selection, respectively, on the frontal thermal face images only. Different parameters for PSO like number of population ‘n,’ c_1, c_2 are varied, and the obtained results are shown in Fig. 2. The x-axis of this figure is represented by ‘pbest’ in each iteration, ‘gbest’ and the y-axis demonstrated the percentage of recognition rate in each iteration. Finally, SIFT and PSO are used on all thermal face images. In all these cases, total

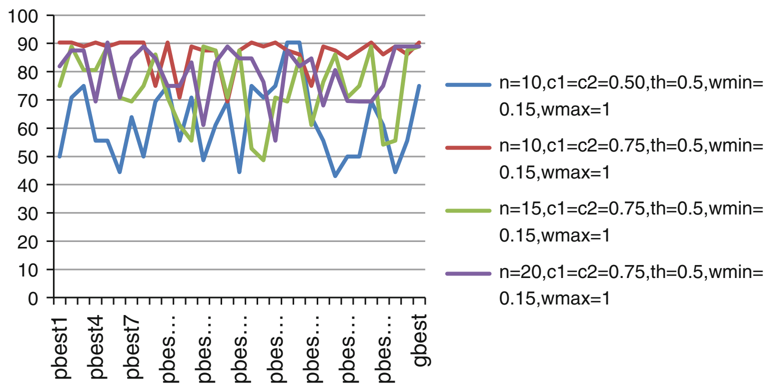


Fig. 3 The different values of 'pbest' and 'gbest' with different parameters of PSO for all faces

Table 1 The recognition rate in percentage

Method	Frontal thermal face images	All thermal face images
SIFT only	94.44	86.84
SIFT + PSO	98.61 (gbest)	90.28 (gbest)
Culter [2]	72.22	70.83

images are divided into two sets. First set is used for training, and other is used for testing purpose. Thus, from the Figs. 2 and 3, it is clear that the 'gbest' means global best recognition is 98.61 and 90.28 % when the value of 'n' is 10, c_1 and c_2 are 0.75 for frontal faces and all face images, respectively.

3.1 Comparison with Other Methods

The obtained recognition rate by the present method has been compared with one pioneer work on recognition of infrared face images by Culter [2]. The work has been implemented on UGC-JU thermal face images. The obtained results are shown in Table 1. All the results support the conclusion that the face recognition performances by the present method give better results in frontal images as well as all the images.

4 Conclusions

This paper presents a scale, translation, and rotation invariant thermal face recognition system using SIFT for face recognition systems. In this system, a PSO-based feature selection algorithm is efficiently utilized to search the optimum features which are not noisy and also not irrelevant. Experimental results based on the UGC-JU thermal face database which consists different facial expressions pose changes about 3 axes, occlusion. Four sets of experiments have been performed. Total number of images is divided into two parts of equal size. First part is used for training, and the other is used for testing. The maximum success of 98.61 % recognition has been achieved using SIFT and PSO for frontal face images and 90.28 % for all images, when the value of 'n' is 10, c_1 and c_2 are 0.75.

Acknowledgments Authors are thankful to a project supported by DeitY (Letter No.: 12(12)/2012-ESD), MCIT, Government of India and DST-PURSE Programme at Department of Computer Science and Engineering, Jadavpur University, India for providing the necessary infrastructure to conduct experiments relating to this work. Ayan Seal is grateful to Department of Science and Technology (DST), Government of India for providing him Junior Research Fellowship-Professional (JRF-Professional) under DST-INSPIRE Fellowship programme [No: IF110591]. Ayan Seal is also thankful to Universidad Politecnica de Madrid, Spain, for providing him scholarship under Erasmus Mundus Action 2 India4EU II.

References

1. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognition—a review. Comput. Vis. Image Underst. **97**, 103–135 (2005)
2. Cutler, R.: Face recognition using infrared images and eigenfaces. University of Maryland, Technology Report, College Park (1996)
3. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
4. Fisher, R.A.: The statistical utilization of multiple measurements. Ann Eugenics **8**, 376–386 (1938)
5. Bhattacharjee, D., Seal, A., Ganguly, S., Nasipuri, M., Basu, D.K.: A comparative study of human thermal face recognition based on haar wavelet transform and local binary pattern. Comput. Intell. Neurosci. doi:10.1155/2012/2610892012
6. Morse, B.S.: Lecture 2: Image Processing Review, Neighbors, Connected Components, and Distance (1998–2004)
7. Leonardis, A., Bischof, H.: Robust recognition using eigenimages. Comput. Vis. Image Underst. **78**(1), 99–118 (2000)
8. Everson, R., Sirovich, L.: Karhunen: Loeve procedure for gappy data. J. Opt. Soc. Am. A **12**(8), 1657–1664 (1995)
9. Seal, A., Bhattacharjee, D., Nasipuri, M., Basu, D.K.: UGC-JU face database and its benchmarking using linear regression classifier. Multimedia Tools Appl. (2013). doi:10.1007/s11042-013-1754-8
10. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)
11. Eberhart, R.C., Kennedy, J.: A new optimizer using particle swarm theory. In: Proceedings of the 6th International Symposium on Micromachine and Human Science, Nagoya, Japan, pp. 39–43 (1995)
12. Ramadan, R.M., Abdel-Kader, R.F.: Face recognition using particle swarm optimization-based selected features. Int. J. Signal Process. Image Process. Pattern Recogn. **2**(2), 51–65 (2009)
13. Kokiopoulou, E., Frossard, P.: Classification-specific feature sampling for face recognition. In: Proceedings of the IEEE 8th Workshop on Multimedia Signal Processing, pp. 20–23 (2006)

14. Narendra, P.M., Fukunage, K.: A branch and bound algorithm for feature subset selection. IEEE Trans. Comput. **6**(9), 917–922 (1977)
15. Pudil, P., Novovicova, J., Kittler, J.: Floating search methods in feature selection. Pattern Recogn. Lett. **15**, 1119–1125 (1994)
16. Roberto, B.: Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Networks **5**(4), 537–550 (1994)
17. Zhang, H., Sun, G.: Feature selection using Tabu search method. Pattern Recogn. Lett. **35**, 701–711 (2002)
18. Kim, D.-S., Jeon, I.-J., Lee, S.-Y., Rhee, P.-K., Chung, D.-J.: Embedded face recognition based on fast Genetic Algorithm for intelligent digital photography. IEEE Trans. Consum. Electron. **52**(3), 726–734 (2006)
19. Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A., Jain, A.K.: Dimensionality reduction using Genetic Algorithms. IEEE Trans. Evol. Comput. **4**(2), 164–171 (2000)
20. Fan, X., Verma, B.: Face recognition: a new feature selection and classification technique. In: Proceedings of the 7th Asia-Pacific Conference on Complex Systems, (2004)
21. Kanan, H.R., Faez, K., Hosseinzadeh, M.: Face recognition system using ant colony optimization-based selected features. In: Proceedings of the IEEE Symposium Computational Intelligence in Security and Defense Applications (CISDA 2007), pp. 57–62 (2007)
22. Seal, A., Bhattacharjee, D., Nasipuri, M., Basu, D.K.: UGC-JU Face Database and its Benchmarking using Linear Regression Classifier, Multimedia Tools and Applications, Springer, US, 10.1007/s11042-013-1754-8, (2013)

http://www.springer.com/978-81-322-1984-2

Applied Computation and Security Systems
Volume One
Chaki, R.; Saeed, K.; Choudhury, S.; Chaki, N. (Eds.)
2015, X, 213 p. 155 illus., Softcover
ISBN: 978-81-322-1984-2