A Novel Face Recognition Approach Based on Genetic Algorithm Optimization

Mourad MOUSSA[1]*, Maha HMILA[2], Ali DOUIK[2]

[1] Department of Informatics, Sciences Faculty of Gafsa,
University of Gafsa, Tunisia
Mourad.enim@yahoo.fr (**Corresponding author*)

[2] Department of Industrial Informatics, National engineering school of Sousse
University of Sousse, Tunisia
ali.douik@enim.rnu.tn

Abstract: In the field of image processing and recognition, discrete cosine transform (DCT) and principal component analysis (PCA) are two widely used techniques. In this paper we present a face recognition approach based on them. Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. It is the most important step that affects the performance of a face recognition system. Genetic Algorithms (GA), one of the most recent techniques in the field of feature selection, are a type of evolutionary algorithms that can be used also to solve this issue. The application of a GA in the resolution of a problem requires the coding of the potential solutions to this problem in finite bit chains in order to constitute the chromosomes coming from a population formed by candidate points. The aim is to find a selective function allowing good discrimination between chromosomes and to define the genetic operators that will be used. In this sense, this approach seeks to develop a system of face recognition using Genetic Algorithm and a DCT-PCA combination for feature selection and dimensionality reduction, to be applied to an archive of images of human faces. The proposed approach is applied on various Face Databases. Experimental results demonstrate the effectiveness of this approach compared to state of the art in face recognition.

Keywords: Face Recognition, Discrete Cosine Transform (DCT), Principal Component Analysis (PCA), Genetic Algorithm (GA).

1. Introduction

The field of facial recognition is meant to propose methods to identify people on the basis of visual information. For this purpose, facial recognition is one of the biometric methods used to identify individuals on the basis of biological information, many studies being done, such as [3], [25], [37], [2] and [28]. Other methods belonging to this category are, for example, capture of the iris image [8] or fingerprints [13]. In recent years, the problem of facial recognition has generated a great deal of research with abundant publications because of its many applications. These include systems for monitoring [22], man-machine interaction [7] and many more.

Over the last few decades, automatic face recognition and detection [36] witnessed considerable progress. However, it is a very challenging problem when the number of variables is large, the training examples are few, or the conditions of capture are unconstrained, resulting in face images varying widely in orientation, expression, and illumination. In all facial recognition methods, the most delicate point concerns the extraction and selection of the most relevant facial features, namely the features that best represent the information carried by a face. In fact, the facial features play a very important role in the task of classification and face recognition. Therefore, the selection of the appropriate features is necessary because some raw data may be redundant or irrelevant to this task. In this regard, in some cases, the performance of the recognition system is degraded due to the presence of redundant features. The problem of feature selection in classification usually arises when the number of variables that can be used to explain the class of an individual is very high. The needs have evolved a lot in recent years with the manipulation of a large number of variables in many fields such as genetic data or image processing. However, if data described by a large number of variables is to be processed, conventional methods of analysis, learning or data mining may prove ineffective or may lead to inaccurate results. Genetic Algorithms (GAs) are original methods that can be used in many fields, i.e. like [1] as well as for feature selection [5]. Often, selection methods that use GA are based on some wrapper methods for the evaluation of individuals. In this paper, we propose an approach based on reducing the initial size of

https://doi.org/10.24846/v27i1y201813

the data and selecting sets of relevant variables, using (PCA) combined with (DCT) and then GA. At first, the DCT transform is applied to convert the image into a frequency domain, and then a first reduction of the dimensionality is performed by the rejection of the high-frequency components by PCA. The genetic algorithm is then used to select the most significant and useful coefficients. The rest of the paper has been arranged as follows. In Section 2 we describe some crucial previous works in pattern recognition using genetic algorithm. Further on Section 3 presents the feature extraction and selection used techniques. Then Section 4 details the proposed approach, the experimental results and discussion. Finally, the conclusions inferred from the results are presented in Section 5.

2. Related works

Several face recognition methods have been proposed in recent years, based on two main axes: recognition from fixed images [38] and recognition from image sequences [29]. Studies in this field, under different lighting conditions, facial expressions and orientations, can be classified into two distinct categories depending on whether they relate to a geometric approach or to a global approach. We particularly emphasize the dimensional reduction methods which are part of the global approaches. Into a space with reduced dimensionality methods fall within the broader framework of data processing in general and object recognition. Global approaches take the image of the face as a whole and use well-known statistical analysis techniques. The idea is generally to project the input image of the face, previously vectorized, into a space of smaller dimensions, where recognition is supposed to be easier. The projection is often designed to select only the important features and sufficiently discriminating to differentiate the people between them. One of the advantages of global methods is that they are quick to implement, and their calculations are based on relatively simple matrix operations. However, since they consider the face as a whole, they are sensitive to the conditions of luminosity, pose or facial expression. The global methods can be broken down into two types of techniques: linear techniques and nonlinear techniques. The linear techniques realize a linear projection of the faces on a space of smaller dimension. The best known among these approaches is the technique known as Eigenfaces presented by Turk and Pentland in [34]. A PCA is performed on a training set of images of faces. The main eigenvectors resulting from the PCA define the new space. The images of faces are then projected onto this space, and the vectors obtained are used for classification. Many works have been carried out on the choice of the vectors to retain to define the new space. Thus, Kirby et al. [16] propose a criterion based on the energy of the eigenvalues associated to the eigenvectors. The eigenvectors corresponding to the greatest eigenvalues are retained until the sum of the eigenvalues exceeds a certain threshold of the total energy. Martinez et al. show in [23] that the recognition rates can be improved by ignoring the first eigenvectors (those whose associated eigenvalues are the largest), which often encodes the variations of illumination. Another well-known approach presented by Ming et al. [20] performs a Linear Discriminant Analysis (LDA), often called Fisher faces. Indeed, this technique consists in maximizing on a learning set the Fisher criterion, namely the quotient of the inter-class variance by the intra-class variance. Thus, contrary to the technique of Eigenfaces where the best representation (that maximizes the variance) is sought, the goal here is a better separation of the classes. However, since the number of images is often smaller than their dimension, the intra-class variance matrix may be singular, and its inversion is therefore problematic. Other linear techniques have also been used for the computation of feature vectors such as Independent Component Analysis in [11], the factorization of non-negative matrices in [9], the bilinear discriminant analysis in [35]. Other global non-linear techniques have been developed, often using linear techniques. Thus, the Kernel-PCA analysis and the Kernel-LDA [15], [10] use the mathematical notion of kernels to extend the linear PCA and LDA techniques. Genetic algorithms have also successfully demonstrated their great ability to solve optimization problems. They have also been used in the field of feature selection [18]. Many of the studies reported in the literature have shown that methods that use GAs as a research technique have yielded better results compared to other methods of selection [12], [32]. The application of GAs on selection problems of subsets of feature was put in place by Ferri et al. [6]. They showed that the use of GAs is well suited for selection on sets of medium-sized features. Kudo et al. [17] demonstrated the possibility of using GAs for selection on large-scale sets by adjusting the number of generations, the size of the population and the probabilities of genetic operations on the one hand and the evaluation function on the other hand.

3. Feature extraction and selection

The feature selection methods are techniques for choosing the most interesting and relevant features of a given system to enhance its performance. This phase is usually an important module of a complex system. The fields of application of feature selection techniques are varied, for example, modeling [27], classification [4], automatic learning (Machine Learning) [14] and data mining. In the specialized literature, the authors set out a list of three objectives to achieve a selection of features for classification: reducing the task of extracting features, improving the accuracy of the classification module and improving the reliability of the performance estimation.

3.1 Feature extraction using DCT

In recent years, some researchers have explored the possibility of extracting features in the frequency domain using Discrete Cosine Transform (DCT) like [26], [21]. The results showed that this technique is promising and allows feature discriminating in the frequency domain. Also, it has been concluded that even the most dominant features can degrade the performance of the recognition system due to variations in laying, lighting and expression. As a result, the extraction of features is an important step before classification. The DCT is a mathematical function that allows changing the representation domain of a signal. Thus, a temporal or spatial signal can be transformed into an identical representation in the frequency domain, making certain of these properties exploitable. DCT is widely used in signal and image processing, especially in compression. The DCT has indeed an excellent property of "regrouping" the energy: the information is mainly carried by the low-frequency coefficients. High frequencies are reserved for rapid changes in the pixel intensity; therefore, they are generally minimal in an image. Thus, it is possible to represent the entire information of the image on very few coefficients. The DCT is applied on a square matrix and the result is shown in a matrix of the same size. The low frequencies are located at the top left of the matrix, and the high frequencies at the bottom right. The matrix transformation DCT being orthogonal, it is accompanied by a method of inversion in order to be able to return in the spatial domain. Thus, after making modifications in the frequency domain and eliminating variations of the image that are almost invisible to the human eye, we return to a representation in the form of pixels. The following equation shows DCT Matrix:

$$DCT(i,j) = \frac{1}{\sqrt{2}} C(i)C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} pixel(x,y) \cos\left(\frac{(2x+1)i\pi}{2N}\right) \cos\left(\frac{(2y+1)j\pi}{2N}\right)$$

3.2 Feature selection using PCA

Principal Component Analysis (PCA) is an old and widely used method of data analysis, and it is widely known in statistics. We suppose to have a set X of N data, each described by P attributes. If we consider the data as points in a P-dimensional Euclidean space, the objective of the PCA is to construct the most characteristic and economic Euclidean space to represent these points. Thus, the objective of PCA is to move from the data space to a feature space.

3.3 Genetic algorithm

GAs, initiated in the 1970s by John Holland, are optimization algorithms based on derivated techniques of the genetics and nature evolution mechanisms. Recent works in the field of face recognition use evolutionary algorithms for feature selection [33]. GAs are a type of evolutionary algorithms that can be used to solve this kind of problem [19]. To use the GAs, we start with an arbitrarily selected initial population of chromosomes and evaluate the relative fitness of each chromosome. A GA is an iterative optimum search algorithm that manipulates a population of constant size. The constant size of the population causes a phenomenon of competition between the chromosomes. Each chromosome represents the coding of a potential solution to the problem to be solved, and consists of a set of elements called genes, which can take several values belonging to an alphabet that is not necessarily digital. At each iteration, called generation, a new population with the same number of chromosomes is created. This generation consists of chromosomes better "adapted" to their environment as represented by the selective function. As the generations progress, the chromosomes will tend towards the optimum of the selective function. The creation of a new population from the previous one is done by applying the genetic operators that are: selection, crossing and mutation.

3.3.1 Selection

The selection is a process by which a chromosome is copied into the new population according to the values of the function to be optimized for this chromosome.

3.3.2 Crossing

The simple crossing or a point crossing consists firstly in choosing a pair of chromosomes with a probability p, then as a second step, the representative chains are cut at an identical random position in both parents. This produces two "head segments" and two "tail segments". Finally, we switch the two tail segments of the parents to obtain two children who inherit some characteristics of their parents.

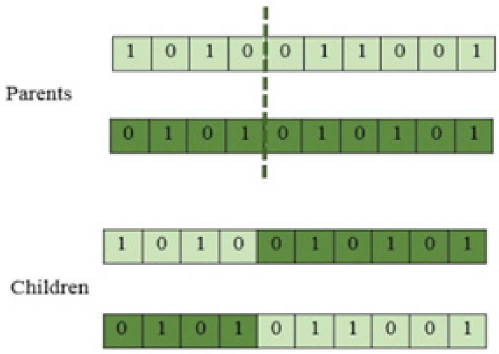

Figure 1. Crossing

3.3.3 Mutation

A mutation is defined as the inversion of a bit in a chromosome. This is equivalent to changing the value of a parameter randomly. Mutations play the role of noise and prevent evolution from freezing. They allow a global as well as a local search, according to the weight and the number of mutated bits. Moreover, they guarantee mathematically that the global optimum can be achieved.

Algorithm1: genetic algorithm
Initialization
Creation of the initial population
Evaluation of the initial population
repeat
Selection of individuals
Application of genetic operators on these individuals
Evaluation of the new individuals
until
obtaining a satisfactory solution or reaching the predefined number of generations.

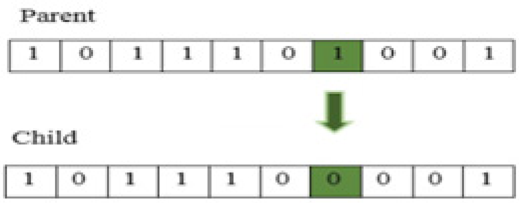

Figure 2. Mutation

4. Experiments and discussion

4.1 Experiments

In order to evaluate our procedure presented by Figure 3, three benchmark databases (Yale, ORL and UMIST), whose description is given in Table 1 and whose samples appear in Figure 4, have been used. Each database was randomly divided into training and, respectively test sets. All possible configurations of training and test sets have been experimented. The results presented are the average ones after all these experiments. While experimenting, we have used the minimum Euclidian distance (ED) as a measurement, for decision making on classification.

Table 1. Databases details

Database	Number of classes	Images per class	Size of image
Yale	15	11	243*320
ORL	40	10	112*92
UMIST	20	24	variable

The experiments were realized using MATLAB R2015a. As part of our study, several tests were carried out to adopt the best parameterization. The number of genes to be selected was set to 30. The other parameters we have set for the genetic exploration are shown in Table 2.

Table 2. Genetic Algorithm parameters

Parameters	Value
Population size	50
Number of generations	100
Crossing probability	0.5
Mutation probability	0.1

4.2 Results and discussion

In the learning phase, the genetic algorithm was used with a population of 50 chromosomes that reproduce for 100 generations, according to the results. The coefficients have been selected by the

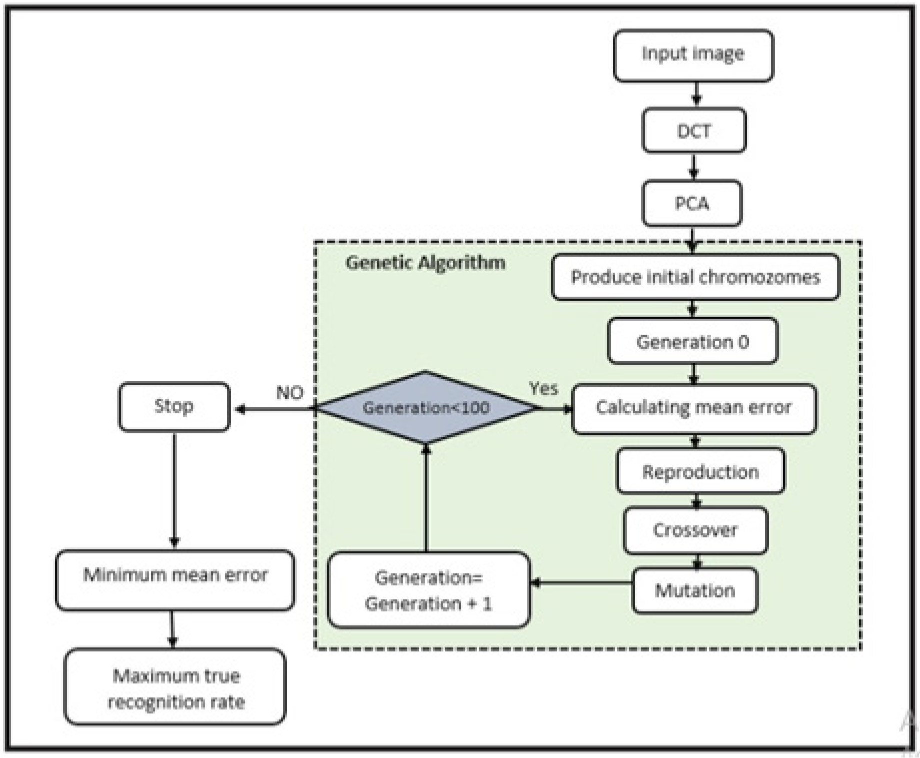

Figure 3. Whole procedure of our face recognition approach

Figure 4. Samples of Yale, ORL and UMIST Databases

Table 3. Comparison of our approach with some other works

Database	Number of classes	Number of train cases	Number of test cases	Recognition rate of previous works	Recognition rate of our works
ORL	40	3	7	92.5%	92.62%
ORL	20	6	4	97.5%	98.45%
UMIST	20	24	variable	91.66%	99.4%
YALE	15	5	6	93.33%	96.5%
YALE	15	4	7	95.23%	95.5%

genetic algorithm. In our case, DCT transforms discrete, integer image data from the spatial domain to the frequency domain. After this transformation, the distribution of the coefficients becomes more concentrated, such that the main data of the framework focused on the weak amplitude frequencies.

Using the combination of DCT and PCA, made it possible to obtain a final identification rate of over than 99% in some experiments. We can see that our method of selection shows better performances than many other previous works like [31] and [24]. That is, a gain of more than 8% compared to the best result obtained by the method [30]. This allows us to emphasize especially the contribution of our method for relevant coefficient selection.

The results obtained by our procedure are compared with other related works in Table 3.

5. Conclusion

Features extraction and optimization is an important task in computer vision and pattern recognition, in fact it contributes to improving results obtained by many popular applications like classification, tracking and segmentation.

The quality of the processing system depends directly on the correct choice of the relevant extracted data, in the form of feature vectors. However, in many cases the resolution becomes difficult enough due to bad choice of the feature extraction and selection methods, resulting in the loss of some useful information and then affecting the true recognition rate. In this paper, we propose a new approach of face recognition using the genetic algorithm and a DCT-PCA combination to select the features and reduce dimensionality. Experiments have shown that our method is fast, and has the ability to select a small number of features while maintaining very satisfactory classification rates. The performance of the proposed method is demonstrated through a comparison with other methods from the specialized literature.

REFERENCES

1. Ben Hassen, N., Saadaoui, K. & Benrejeb, M. (2017). Lead-lag Controller Design for Time Delay Systems Using Genetic Algorithms, *Studies in Informatics and Control*, *26*(1), 87-94.

2. Benzaoui, A. & Boukrouche, A. (2014). Face Analysis, Description and Recognition using Improved Local Binary Patterns in One Dimensional Space, *Journal of Control Engineering and Applied Informatics*, *16*(2014), 52-60.

3. Chen, Z., Liu, C., Chang, F., Han, X. & Wang, K. (2014). Illumination Processing In Face Recognition, *International Journal of Pattern Recognition and Artificial Intelligence*, *28*(2014), 42-58.

4. Dash, M. & Liu, H. (1997). Feature selection for classification, *Intelligent Data Analysis*, *1*(1997), 131-156.

5. El Ferchichi, S., Laabidi, K. & Zidi, S. (2009). Genetic Algorithm and Tabu Search for Feature Selection, *Studies in Informatics and Control*, *18*(2), 181-187.

6. Ferri, F. J., Kadirkamanathan, V. & Kittler, J. (1993). Feature Subset Search using Genetic Algorithms. In *IEE/IEEE Workshop on Natural Algorithms in Signal Processing*.

7. Gautam, R. & Singh, P. (2015). Human machine interaction, *International journal of science, technology & management*, *4*(1), 188-193.

8. Gupta, S. & Gagneja, A. (2014). Proposed Iris Recognition Algorithm through Image Acquisition technique, *International Journal of Advanced Research computer Science and Software Engineering*, *4*(2), 268-271.

9. Guillamet, D., Vitri, J. & Schiele, B. (2003). Introducing a weighted non-negative matrix factorization for image classification, *Pattern Recognition Letters*, *24*(3), 2447-2454.

10. Huang, J., Yuen, P. C., Chen, W.-S. & Lai, J. H. (2004). Kernel subspace LDA with optimized kernel parameters on face recognition. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*.

11. Hyvrinen, A. & Oja, E. *One-unit Learning Rules for Independent Component Analysis*, *Technical Report.* Helsinki University of Technology, Laboratory of Computer and Information Science.

12. Ishibuchi, H. & Nakashima, T. (2000). Multi-objective pattern and feature selection by a genetic algorithm. In *Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation* (pp. 1069-1076).

13. Jain, A. K., Arora, S. S. & Cao, K. (2016). Fingerprint Recognition of Young Children, *IEEE Transactions on Information Forensics and Security*, *12*(7), 1501-1514.

14. Joachims, T. (2005). Text categorization with Support Vector Machines: Learning with many relevant features, *Machine Learning, ECML-98*, 137-142.

15. Kim, K. I., Jung, K. & Kim, H. J. (2002). Face recognition using kernel principal component analysis, *IEEE Signal Processing Letters*, *9*(2), 40-42.

16. Kirby, M. & Sirovich, L. (1990). Application of the KL Procedure for the Characterization of Human Faces, *IEEE Trans. Pattern Analysis and Machine Intelligence*, *12*(3), 103-108.

17. Kudo, M. & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers, *Pattern Recognition*, *33*(2000), 25-41.

18. Kuncheva, L. I. & Jain, L. C. (1999). Nearest neighbor classifier: Simultaneous editing and feature selection, *Pattern Recognition Letters*, *20*(3), 1149-1156.

19. Leardi, R., Boggia, R. & Terrile, M. (1992). Genetic algorithms as a strategy for feature selection, *Journal of Chemometrics*, *6* (1992), 267-281.

20. Li, M. & Yuan, B. (2005). 2D-LDA: A statistical linear discriminant analysis for image matrix, *Pattern Recognition Letters*, *26*(5), 527-532.

21. Manikantan, K., Govindarajan, V., Sasi Kiran, V. V. S. & Ramachandran, S. (2012). Face Recognition using Block-Based DCT Feature Extraction, *Journal of Advanced Computer Science and Technology*, *1*(2012), 266-283.

22. M'halla, A., Dutilleul, S. C., Craye, E. & Benrejeb, M. (2010). Monitoring of a Milk Manufacturing Workshop Using Chronicle and Fault Tree Approaches, *Studies in Informatics and Control*, *19*(4), 379-390.

23. Martinez, A. M. & Kak, A. C. (2001). PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(2), 228-233.

24. Pandaa, R. & Naik, M. K. (2015). A novel adaptive crossover bacterial foraging optimization algorithm for linear discriminant analysis based face recognition, *Applied Soft Computing*, *30*(2015), 722-736.

25. Raghavendra, R., Raja, K. B. & Busch, C. (2015). Presentation Attack Detection for Face Recognition Using Light Field Camera, *IEEE Transactions on Image Processing*, *24*(2015), 1060-1075.

26. Rashidi, S., Fallah, A. & Towhidkhah, F. (2012). Feature extraction based DCT on dynamic signature verification, *Scientia Iranica*, *19*(2012), 1810-1819.

27. Ribeiro, P. C. & Victor, J. S. (2005). Human Activity Recognition from Video: modeling, feature selection and classification architecture. In *International workshop on human activity recognition and modelling* (*2005*).

28. Sagar, G. V., Barker, S. Y., Raja, K. B., Babu, K. S. & Venugopal, K. R. (2015). Convolution based Face Recognition using DWT and feature vector compression, *Journal of Engineering and Technology*, *2*(2015), 320-324.

29. Sharma, P., Yadav, R. N. & Arya, K. V. (2016). Face recognition from video using generalized mean deep learning neural network. In *4th International Symposium on Computational and Business Intelligence*.

30. Subban, R., Mankame, D., Nayeem, S., Pasupathi, P. & Muthukumar, S. (2014). Genetic Algorithm based Human Face Recognition. In *Int. Conf. on Advances in Communication, Network, and Computing, CNC* (*2014*).

31. Sukhija, P., Behal, S. & Singh, P. (2016). Face Recognition System Using Genetic Algorithm, *Procedia Computer Science, 85*(2016), 410-417.

32. Sun, Z., Bebis, G. & Miller, R. (2004). Object detection using feature subset selection, *Pattern Recognition*, *37*(4), 2165-2176.

33. Tangour, F. & Borne, P. (2008). Presentation of Some Metaheuristics for the Optimization of Complex Systems, *Studies in Informatics and Control*, *17*(2), 169-180.

34. Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, *3*(1991), 71-86.

35. Visani, M., Garcia, C. & Jolion, J.-M. (2005). Normalized radial basis function networks and bilinear discriminant analysis for face recognition. In *IEEE International Conference on Advanced Video and Signal Based Surveillance* (*AVSS*) (pp. 342-347).

36. Vrejoiu, M. H. (2017). A Pattern Matching Method and Algorithm for Face Detection, *Studies in Informatics and Control*, *26*(1), 75-86.

37. Wang, W. & Chang, F. (2016). A Facial Expression Recognition Method Based on Feature Blocks, *Journal of Control Engineering and Applied Informatics*, *18*(2016), 77-85.

38. Xu, Y., Zhang, Z., Lu, G. & Yang, J. (2016). Approximately symmetrical face images for image pre-processing in face recognition and sparse representation based classification, *Pattern Recognition*, *5*(4), 68-82.