

# Face Recognition Using Fuzzy Clustering and Kernel Least Square

**Essam Al Daoud**

Computer Science Department, Zarqa Private University, Zarqa, Jordan  
Email: [essamdz@zpu.edu.jo](mailto:essamdz@zpu.edu.jo)

Received October 2014

Copyright © 2015 by author and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Over the last fifteen years, face recognition has become a popular area of research in image analysis and one of the most successful applications of machine learning and understanding. To enhance the classification rate of the image recognition, several techniques are introduced, modified and combined. The suggested model extracts the features using Fourier-Gabor filter, selects the best features using signal to noise ratio, deletes or modifies anomalous images using fuzzy c-mean clustering, uses kernel least square and optimizes it by using wild dog pack optimization. To compare the suggested method with the previous methods, four datasets are used. The results indicate that the suggested methods without fuzzy clustering and with fuzzy clustering outperform state-of-art methods for all datasets.

## Keywords

Face Recognition, Fuzzy Clustering, Kernel, Least Square, Gabor Filters

---

## 1. Introduction

Facial recognition has been an active research topic since the early nineties. There have been several advances in the past few years in terms of face detection and tracking, feature extraction mechanisms and the related machine learning techniques. Face recognition has drawn the attention of researchers in fields from image analysis and processing, computer vision, to psychology and security [1]. In spite of more than fifteen years of extensive re-search, large number of papers published in journals and conferences dedicated to this area, we still cannot claim that automatic face recognition is comparable to human performance [2]. The difficulty of automatic face recognition is due to various effects like aging, facial expressions, occlusions, lighting and viewpoint changes induced by body movement [3]. A successful face recognition mythology depends heavily on the particular choice of the features, observation reduction and classification method. In this paper four algorithms have in-

roduced, modified and combined to enhance the face recognition and detection [4]. The features are extracted using Fourier-Gabor filter and selected by generalizing signal to noise equation. Modified fuzzy clustering algorithm is used to reduce the misclassified faces by excluding or modifying the low membership faces. Furthermore a new non-linear kernel is used to kernel the least square method and the wild dog pack optimization is implemented to find the optimal parameters. The suggested method is implemented by using four datasets: AT & T database of faces (ATT), Indian Face Database (IFD), Faces95 from Essex university database and Yale face database (Yale).

## 2. Previous Studies

Principal component analysis (PCA) converts a set of faces images of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. PCA can be used to capture as much as possible of the variability of the face image where the eigenfaces matrix can be used to recognize the new faces by minimizing [5]:

$$\|U^T(x - \mu) - U^T\delta_b\|^2 \quad (1)$$

where  $U$  is the eigenfaces matrix or features matrix,  $\mu$  is the mean image and  $\delta_b$  is the difference between the input image  $x_b$  and the mean. Linear discriminant analysis (LDA) selects eigenvectors  $U$  in such a way that the ratio of the between-class scatter and the within class scatter is maximized which characterizes or separates two or more classes of faces images

$$U_{opt} = \arg \max_U |U^T S_B U| / |U^T S_W U| \quad (2)$$

where  $S_B$  and  $S_W$  are the between class scatter matrix and the within class scatter matrix respectively and  $U_{opt}$  can be found by solving the generalized eigenvalue problem [6]. Independent component analysis (ICA) separates a multivariate signal into additive subcomponents. ICA uses centering, whitening, and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem for the actual iterative algorithm [7]. The whitening matrix is twice the inverse square root of the covariance matrix, this removes the first and the second-order statistics of the data [8]. Fourier transform is another important method, which plays a critical role in a broad range of image processing applications. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. For a square image of size  $n \times n$ , the two-dimensional DFT for the point  $a$  and  $b$  in spatial domain is equivalent to the point  $m$  and  $n$  in frequency domain, this transform can be calculated as following:

$$\text{Fourier}(m, n, \text{img}A) = \frac{1}{n^2} \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} \text{img}A(a, b) e^{-i2\pi(ma/n + nb/n)} \quad (3)$$

Gabor filters have been used to solve the configuration, orientations and emotions problems. A filter bank consisting of Gabor filters with various scales and rotations can be created and combined with different methods to enhance the recognition rate. The most commonly used filter in face recognition have the form [9]:

$$\begin{aligned} \text{Gabor}(x, y, \mu, \nu) &= \rho^2 / \sigma^2 \exp\left(-\rho^2(x^2 + y^2) / 2\sigma^2\right) \\ &\quad * \left(\exp(i\rho * (x, y) - \exp(-\sigma^2/2))\right) \end{aligned} \quad (4)$$

where

$$\rho = \pi / 2 (\sqrt{2})^\nu e^{i\pi\mu/8} \quad (5)$$

The output locations for each Gabor sub-matrix are specified by  $x$  and  $y$ .  $\mu$  and  $\nu$  define the orientation and scale of each sub-matrix.

## 3. The Suggested Method

The suggested method modifies and combines several techniques to enhance the face recognition. The main

steps can be summarized as following:

- 1) Extract features from each image using Fourier-Gabor filter
- 2) Select the best features using signal to noise ratio (SNR)
- 3) Delete or modify anomalous images using fuzzy c-mean cluster
- 4) Model the training images using the least square method
- 5) Use a new non linear kernel
- 6) Optimize the weight vector using wild dog pack optimization (WDPO)

### 3.1. Feature Extraction and Selection

There are several methods have been used to extract the features from the face images, one among the best methods is Fourier-Gabor filter which is introduced in [8]. Therefore the same method will be used in this study to extract the features from the face images. However different size, orientation and scale values are adopted in this study. The main steps are as following:

- 1) Resize the images to  $40 \times 40$
- 2) Transform the images to frequency domain by applying Fourier transform in Equation (3)
- 3) Use Equation (4) to prepare 8 Gabor filters for orientation and 10 filters for scaling. Thus 80 different filters will be constructed
- 4) Multiply the result matrix in step 2 by each matrix in step 3.
- 5) Resize each matrix in step 4 as one row
- 6) Construct the feature vector for each face by concatenation all the rows in step 5.

The number of the extracted features is  $40 \times 40 \times 80 = 104,000$ . To delete insignificant features, Signal-to-Noise Ratio (SNR) method is applied [10]:

$$SNR = \frac{|\mu_- - \mu_+|}{|\sigma_- + \sigma_+|} \quad (6)$$

where  $\mu$  is the mean,  $\sigma$  is the standard deviation and the signs  $\{+, -\}$  are indicate to the class. But for multiclass, as in our case, one against all procedure is repeated for all classes and the total is calculated. The features that have lowest total are deleted.

### 3.2. Face Reduction and Modification

If face  $x$  in a class is closer to the faces in another class, this will lead to misclassification of the new faces that are close to the face  $x$ . Therefore, fuzzy clustering is used to modify the faces that have low membership degree in the correct class and delete the faces that have high membership degree in the wrong classes. Thus the fuzzy c-mean clustering will be modified as following:

- 1) Initialize the membership randomly  $u_{ij}$
- 2) Calculate the centers

$$c_j = \frac{\sum_{i=1}^n u_{ij}^2 x_i}{\sum_{i=1}^n u_{ij}^2} \quad (7)$$

- 3) Update the memberships

$$v_{ij} = \sum_{k=1}^m \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^2 \quad (8)$$

$$u_{ij} = 1/v_{ij} \quad (9)$$

- 4) Go to step 2 until no significant improvement.
- 5) For all  $i = \{1 \dots n\}$  and for all  $j = \{1 \dots m\}$ 
  - Consider  $x_i \in \text{cluster}_k$  and  $x_i \in \text{class}_t$
  - If  $u_{ij} > s$  and  $k \neq t$
  - Delete  $x_i$  and update the parameter
  - If  $u_{ij} < \rho$  and  $k = t$
  - Modify  $x_i = (x_i + c_k)/2$
- 6) Go to step 2 until the above conditions become invalid.

$u_{ij}$  is the degree of membership of  $x_i$  in the cluster  $j$ ,  $x_i$  is the  $i^{\text{th}}$  face in the dataset,  $c_j$  is the center of the cluster,  $\sigma$  and  $\rho$  are parameters to be tuned in the experimental section,  $n$  is number of images and  $m$  is number of classes or clusters (the fuzzy c-mean clustering before modification can be found in [11]).

### 3.3. Kernel Least Square Method

Once the features and the observations have prepared, several classifiers can be used such as neural networks, support vector machine (SVM) and regularized least squares classification (RLSC) [10]. In this section RLSC will be used with a new non linear kernel and will be minimized by wild dog packs optimizer (WDPO). The general formulization of non-linear regularized least squares classification as following:

$$\min_{w \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^n \left( y_i - \sum_{j=1}^n w_j K(x_i, x_j) \right)^2 + \lambda / 2 w^T K w \quad (10)$$

$w$  is the weight vector,  $y$  is the target class,  $\lambda$  is the regularization parameter and  $K$  is a kernel. In this study the following kernel is suggested:

$$k(u, v) = 1 - \sin(\mu \|u - v\|) \quad (11)$$

To train RLSC the parameters  $\lambda$ ,  $\mu$  and the vector  $w$  must be optimized. Therefore wild dog pack optimizer is used to find the best values. **Algorithm 1** summarizes the main steps of WDPO [12]:

**Algorithm 1:** Wild dog pack optimization (WDPO) Generate  $n$  dogs randomly  
**while** ( $t < iter$ )

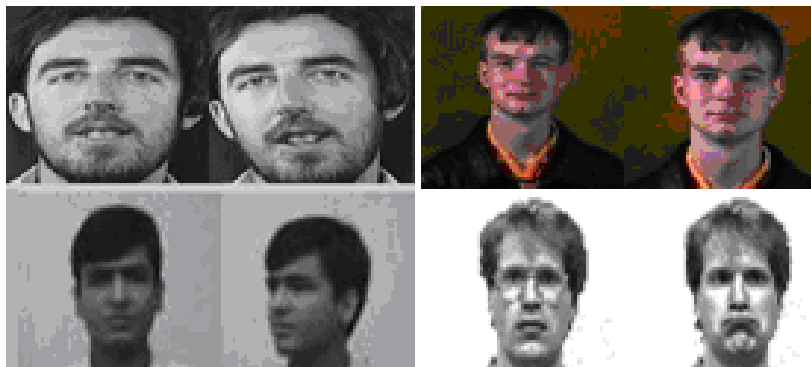
- Evaluate Locations using the updated parameters
- Choose the best location as alpha  $\alpha(t)$
- Every few iterations update the parameters using self competition
- Update the dogs locations with regard to alpha using Equation (12):

$$dog_i(t+1) = dog_i(t) + c \text{ rand}(\alpha(t) - dog_i(t)) + c \text{ rand}(n) (\alpha(t) - dog_i(t)) \quad (12)$$

- $c$ : is used to control the diversity from alpha
- Evaluate dogs
- Choose the best dog as alpha
- If no improvement for  $v$  iterations
- Use Hoo procedure to escape from the local minimum

## 4. Experimental Results

In this study four datasets are used: AT & T database of faces (ATT), Indian Face Database (IFD), Faces95 from Essex university database and Yale face database (Yale) [13]-[16]. **Figure 1** show samples faces from each dataset. All the datasets have many subjects and several images per individual, different facial expressions, configuration, orientations and emotions are used. **Table 1** summarizes the content of each dataset.



**Figure 1.** Faces from ATT, Yale, IFD and Faces95.

All the results in this section are found by using 10-fold cross validations. The code is written by using Matlab 10. The experiments are divided into three stages. The first stage is used to find the best number of features using signal to noise method. The implementation in this stage is repeated several times, each time the number of the features is reduced by 10,000. **Table 2** shows sample results of different number of features. The best number of the features according to this stage is 80,000. The second stage is to find the best values of the clustering parameters  $\sigma$  and  $\rho$ . **Table 3** indicates that the best values are  $\sigma = 50$  and  $\rho = 30$ . The third stage is to compare the proposed method with three famous methods: Linear discriminant analysis (LDA), Principal component analysis (PCA), and Gabor-SVM. Moreover, to the proposed method is implemented with and without fuzzy clustering. **Table 4**, **Table 5** show the classification rate using 30% and 50% for training respectively. The results indicate to that the suggested methods without fuzzy clustering and with fuzzy clustering outperform state-of-art methods for all datasets.

**Table 1.** Face dataset description.

<i>Datasets</i>	<i>Images</i>	<i>Subjects</i>	<i>Total</i>	<i>Variations</i>
<b>ATT</b>	10	40	400	lighting and expressions
<b>IFD</b>	11	40	440	Face direction
<b>Faces95</b>	20	72	1440	Distance and scale
<b>Yale</b>	165	15	2475	Expressions

**Table 2.** Sample results by using different number of features.

#Features	Datasets			
	<i>Yale</i>	<i>Face95</i>	<i>IFD</i>	<i>ATT</i>
<b>104,000</b>	96.3	92.2	94.0	94.7
<b>80,000</b>	97.2	<b><u>94.8</u></b>	<b><u>95.8</u></b>	<b><u>96.9</u></b>
<b>50,000</b>	<b><u>97.3</u></b>	93.0	92.8	95.4
<b>10,000</b>	93.0	91.6	89.7	92.1
<b>5000</b>	88.9	84.2	78.1	83.0

**Table 3.** Sample results by using different values for the clustering parameters  $\sigma$  and  $\rho$ .

Clustering parameters	Datasets			
	<i>Yale</i>	<i>Face95</i>	<i>IFD</i>	<i>ATT</i>
$\sigma = 20$ and $\rho = 5$	94.7	91.5	90.3	91.6
$\sigma = 30$ and $\rho = 10$	96.1	93.0	91.0	93.1
$\sigma = 40$ and $\rho = 20$	96.3	<b><u>95.9</u></b>	94.5	92.8
$\sigma = 50$ and $\rho = 30$	<b><u>97.4</u></b>	95.7	<b><u>96.7</u></b>	<b><u>98.0</u></b>
$\sigma = 60$ and $\rho = 40$	93.1	92.1	93.1	95.2

**Table 4.** The classification rate using 30% for training.

Method	Datasets			
	<i>Yale</i>	<i>Face95</i>	<i>IFD</i>	<i>ATT</i>
<b>PCA</b>	91.2	81.1	80.2	94.2
<b>LDA</b>	95.1	89.2	85.6	95.7
<b>Gabor SVM</b>	95.6	86.2	86.4	95.6
<b>Proposed without clustering</b>	95.7	90.1	88.6	95.1
<b>Proposed with clustering</b>	<b><u>97.1</u></b>	<b><u>92.5</u></b>	<b><u>91.8</u></b>	<b><u>96.0</u></b>

**Table 5.** The classification rate using 50% for training.

Method	Datasets			
	Yale	Face95	IFD	ATT
PCA	92.4	82.7	89.9	95.7
LDA	96.2	92.6	93.4	96.2
Gabor SVM	96.2	90.2	92.3	96.6
Proposed without clustering	96.5	92.1	93.9	95.9
Proposed with clustering	<b>98.2</b>	<b>94.5</b>	<b>96.1</b>	<b>97.6</b>

## 5. Conclusion

Face recognition presents a challenging problem in the field of image analysis and computer vision, and as such has received a great deal of attention over the last few years because of its many applications in various domains. The performance of the proposed method was demonstrated on various datasets that contain several images per individual, different facial expressions, configuration, orientations and emotions. The results are quite promising; the suggested method outperforms the previous methods for all datasets. The future work will be dedicated to apply and combine the other features that are extracted by various techniques.

## References

- [1] Zeng, Z., Pantic, M., Roisman, G.I. and Huang, T.S. (2009) A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **31**, 39-58. <http://dx.doi.org/10.1109/TPAMI.2008.52>
- [2] Aleksic, P.S. and Katsaggelos, A.K. (2006) Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs. *IEEE Trans. Information Forensics and Security*, **1**, 3-11. <http://dx.doi.org/10.1109/TIFS.2005.863510>
- [3] Kotsia, I. and Pitas, I. (2007) Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. *IEEE Trans. Image Processing*, **16**, 172-187. <http://dx.doi.org/10.1109/TIP.2006.884954>
- [4] Kotsia, I., Buciu, I. and Pitas, I. (2008) An Analysis of Facial Expression Recognition under Partial Facial Image Occlusion. *Image and Vision Computing*, **26**, 1052-1067. <http://dx.doi.org/10.1016/j.imavis.2007.11.004>
- [5] Sahoozadeh, A.H., Heidari, B.Z. and Dehghani, C. (2008) A New Face Recognition Method Using PCA, LDA and Neural Network. *Proceedings of World Academy of Science Engineering and Technology*, **31**, 7-12.
- [6] Pang, S., Ozawa, S. and Kasabov, N. (2005) Incremental Linear Discriminant Analysis for Classification of Data Streams. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, **35**, 905-914. <http://dx.doi.org/10.1109/TSMCB.2005.847744>
- [7] Bartlett, M.S., Movellan, J.R. and Sejnowski, T.J. (2002) Face Recognition by Independent Component Analysis. *IEEE Transaction on Neural Networks*, **13**, 1450-1462. <http://dx.doi.org/10.1109/TNN.2002.804287>
- [8] Al Daoud, E. (2009) Enhancement of the Face Recognition Using a Modified Fourier-Gabor Filter. *Int. J. Advance. Soft Comput. Appl.*, **1**, 119-131.
- [9] Vukadinovic, D. and Pantic, M. (2008) Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers. *IEEE International Conference on Systems, Man and Cybernetics Waikoloa, Hawaii*, 10-12 October 2005, 1692-1698.
- [10] Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Edition, Springer, New York. <http://dx.doi.org/10.1007/978-0-387-84858-7>
- [11] Wang, X.Y. and Garibaldi, J.M. (2005) Simulated Annealing Fuzzy Clustering in Cancer Diagnosis. *Informatica*, **29**, 61-70.
- [12] Al daoud, E., Alshorman, R. and Hanandeh, F. (2014) A New Efficient Meta-Heuristic Optimization Algorithm Inspired by Wild Dog Packs. *International Journal of Hybrid Information Technology*, **7**, 1-17.
- [13] AT & T. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [14] Saha, K., Saha, P., Bhowmik, M.K., Bhattacharjee, D. and Nasipuri, M. (2013) Creation of North-East Indian Face

Database for Human Face Identification. *Proc. SPIE 8659, Sensors, Cameras, and Systems for Industrial and Scientific Applications XIV*, Vol. 8659, 19 February 2013, 86590Q. <http://dx.doi.org/10.1117/12.2002028>

[15] Faces95 from Essex University. <http://dces.essex.ac.uk/mv/allfaces/>

[16] Yale Face Database. <http://vision.ucsd.edu/content/yale-face-database>