Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 11 (7): 1213-1218, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.1213.1218

Statistical Model and Wavelet Function for Face Recognition

[1]Nadir Nourain and [2]Brahim Belhaouari Samir
[1]Department of Electrical and Electronic Engineering,
[2]Department of Fundamental and Applied Science, Universiti Teknologi Petronas,
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

Abstract: In this study, we proposed two statistical methods were proposed to reduce the number of features extracted by using multi-level decomposition of wavelet transform applied on facial image to extract the significant features and it make the classification process less sensitive to variation in pose, expressions and light. All coefficients of wavelet transform that do not contribute for face classification have been deleted but the most appropriate wavelet coefficients required for classification were considered. Also using statistical parameters in order to determine whether the coefficients have to be removed or kept derives the error probability. The coefficients were kept or removed based on the threshold limits of the statistical parameters. The simplest classifier, Euclidean Distance Method (EDM) was used in the classification process. The experiments have been performed on Olivetti Research Laboratory database (ORL) and Yale University database (YALE) with different resolutions; success rate of up to 99.33 and 88.48% have been achieved on ORL and YALE database, respectively. These methods brought about 40% improvements in comparison to the method that accounts the biggest coefficients from the four levels of decompositions.

Key words: Facial image compression, wavelet decomposition, extracting features, probability of error, Euclidean distance method

INTRODUCTION

The reduction of wavelet transform coefficients which represent the face in the image is the most significant problem due to the redundancy of features which are not required in Discrete Wavelet Transform (DWT) and these features in turn draws a negative effect on the classification process. This problem refers to the properties of wavelet functions (families) which are used in particular application to produce the coefficients to represent the features. In the case of face classification, some of these coefficients don't have face information that leads to increase the error rate of classification. For this problem, it's important to reduce the coefficients by choosing those coefficients that contain face information and ignoring the remaining.

Wavelet analysis, as opposed to Fourier analysis, offers extra freedom as the selection of atom of the transform inferred from the analysis of wavelet is left to the user. Moreover, according to the objectives of wavelet processing (Misiti *et al*., 2006), they prefer the continuous transform to the discrete transform, if the redundancy is useful to analyze the signal and make opposite choice, for signal compression. In the latter case, it is preferable to use the filters with wavelets whereas in the former case almost any zero integral function is appropriate. The Haar base (Karimi *et al*., 2006) which appeared at the beginning of the last century was replaced by its successors, Gaussian Morlet wavelets (Dan and Wentao, 2009), Meyer wavelets (Xia and Suter, 1996) and Daubechies wavelets (Vonesch *et al*., 2007). Among the new wavelets transforms, Daubechies wavelets are the most widely used. Besides, numerous wavelets regularly appear in the books and are made available in specialized software application. Construction of new wavelets was very intense in the first ten years after their introduction but recently, it has become less regular and bears on increasing specific goals, often associated with limited application contexts.

In facial images, the features extracted by wavelet function are used to classify the face. In the study of Ghosal *et al*. (2009), Gabor wavelets are used to extract the features for face recognition but dealing with all Gabor features is consider as a time consuming process. To resolve this problem, they proposed the use of Random Forest's variable importance computation feature. This computation feature is used to compute the most important Gabor features. This in turn, reduces the feature space by more than hundreds of features and speeds up the classification process. Eltoukhy *et al*. (2009)

Corresponding Author: Nadir Nourain, Department of Electrical and Electronic Engineering, Universiti Teknologi Petronas, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

proposed computer aided diagnosis system based on wavelet analysis multi-level decomposition using three types of wavelet function (db 8, sym 8 and bior 3.7) to extract the features and Euclidean Distance Method (EDM) as classifier to increase diagnostic accuracy. They also solve the problem of coefficients' redundancy by selecting the biggest hundred coefficients from each level of decomposition and then passing these coefficients to EDM classifier.

In this study, wavelet analysis multi-level of decomposition is used to transform the facial images by using Symlet wavelets in order to produce features coefficients and then the proposed statistical model is applied to these coefficients to reduce the redundancy and deleting all coefficients that don't have information. Finally, the last stage uses EDM as a simple classifier. Also the comparison between the method in (Eltoukhy *et al*., 2009) when it was used for face classification and the proposed method which uses two different databases and the result showed the increase in classification accuracy up to 40%.

SYMLET WAVELETS

Symlets (Guo-Sheng *et al*., 2006) form a family of almost similar wavelets proposed by Daubechies adjusting the construction of dbN. From the symmetry, the other properties of the two families are similar. An example of Symlets of order 2 to 8 (sym 1 is simply the Haar wavelet) is represented in Fig. 1.

From, Misiti *et al*. (2006), the idea of construction consists of re-using the m_0 function introduced for dbN, consider $| m_0 (\omega) |^2$ as function W of the variable $z = e^{i\omega}$. We can factorize W in various manners in the form of W (z) = U (z) U (z-1), since the roots of W with module

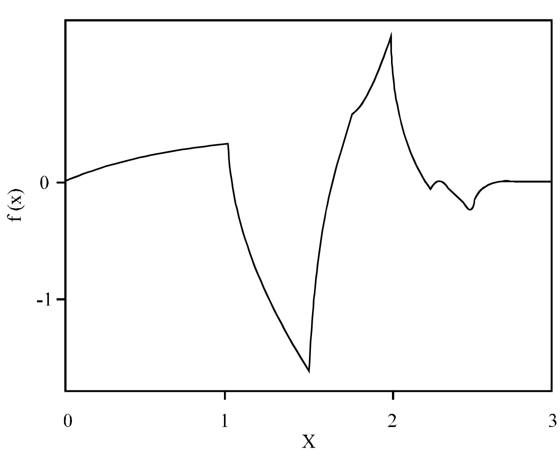

Fig. 1: Symlets: sym 2

different from 1 go in pair: if z_1 is a root then z_1-1 is also a root. By constructing U so that its roots are all of module<1, we construct the Daubechies wavelets dbN. The filter U has minimal phase. Another option, attained by optimizing factorization so that the filter U has almost linear phase and produces much more symmetric filter: the symlets.

The Euclidean Distance Method (EDM): Euclidean Distance Method (EDM) is the simplest classifier and easy in implementing as it was explained by Eltoukhy *et al.* (2009). For each class, the class core vector is the mean of 50% of the class vectors as in Eq. 1. For a new feature vector, the distances between feature vector and the class core vectors a re calculated using Eq. 2. The system automatically classifies the feature vector in the class for which the distance obtained is the smallest.

$$V_{core} = \frac{1}{N}\sum_{j=1}^{j} V_j \quad (1)$$

$$Dist = \sqrt{\sum_{i=1}^{k}\left(V^i_{core} - V^i_{test}\right)^2} \quad (2)$$

where, V_j is the coefficient vector for each training image, j is the index of vector N is the number of images used for training, Dist is the calculated distance between the tested image and every core vector, k is the length of vectors. V^i_{test} the feature vector of face to be classified and V^i_{core} is the vector core of each class. The facial images that are used in testing the effectiveness of the proposed system are decomposed into four levels of decomposition using sym 8. The resulting tables showed the distribution of used face images over different classes. For each experiment, a class core vector is calculated for each class using Eq. 1 and then all the features extracted from the ORL database and YALE database are used in the testing phase including those used to produce the class core vectors. Equation 2 is used to measure the distance between the coefficient vector and each available class core vectors.

THE PROPOSED STATISTICAL MODEL

Suppose, m_1, m_2 and m_3 are the mean of $class_1$, $class_2$ and $class_3$, respectively and m_T is mean total of all classes. Three matrices are as shown in Fig. 2.

From the Fig. 2 it can be seen that the coefficients extracted from the classes are good for classification process that leads to increase system accuracy.

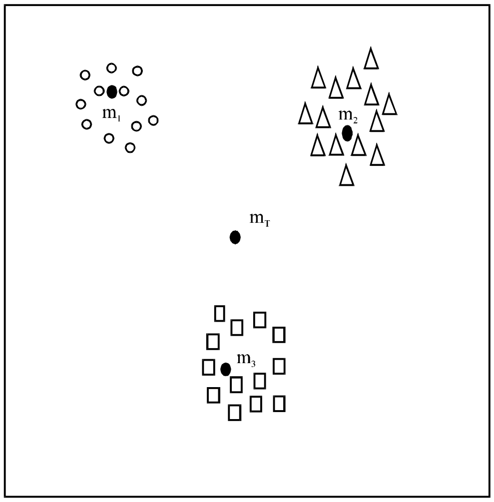

Fig. 2: For three classes without overlapping

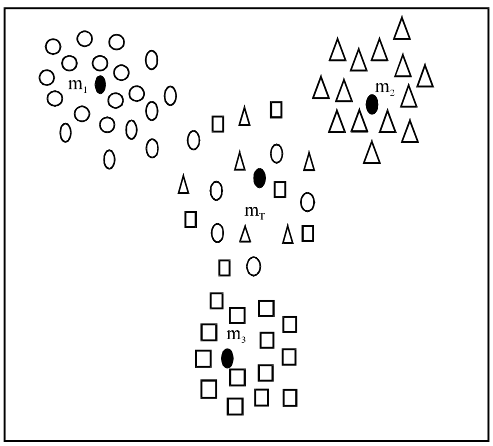

Fig. 3: For three classes with overlapping

From Fig. 3, it can be seen that the coefficients extracted from the classes are not good for classification process due to the overlapping between the classes that leads to increase the probability of error.

Now from Fig. 3, use the following formula:

$$m_T = \frac{\sum_{i=1}^{n} m_i}{n}$$

where, m_T is the total mean, m_i is class mean, i is index of the class and n the number of classes.

So the variance of m_i is:

$$Var_i = \frac{\sum_{j=1}^{n_i} \left(x_j^i - m_i\right)^2}{n_i} \qquad i = 1, 2, 3, \ldots, n_i \tag{4}$$

where, n_i is the number of the features in class i

The metric obtained from the Fig. 3 is not efficient for features extraction of wavelet transform due to overlapping between classes. So we will use another metric from Fig. 2 as follow:

$$Var\_mod = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{(m_i - m_T)^2}{Var_i}\right] \tag{5}$$

$$Var\_min = Min\left[\frac{(m_i - m_T)^2}{Var_i}\right] \tag{6}$$

where, Var_i variance of the class i

Now the way to select the desired features coefficients will be as follow:

If $Var\_mod \leq 1$ delete all coefficients belong to this column and they did not take in our consideration, otherwise we keep it.

To calculate the error probability we will take this example:

Suppose we have two classes as shown in Fig. 4:

From each class of any wavelet coefficients we can approximate it to follow Gaussian variables.

Chosen Feature Extraction of class1 ~ Gaussian (m_1, σ_1)

Chosen Feature Extraction of class2 ~ Gaussian (m_2, σ_2)

$$m_j = \frac{\sum_{i=1}^{i} x_i^j}{n} \qquad j = 1, 2 \tag{7}$$

where, x is the random variable, j^{th} is feature and i is class index.

Using our classifier, the probability of error to classify the variable X is:

$$P(error) = \sum_{i=1}^{2} P(x \in class_i).P(error / x \in class_i)$$

$$P(x \in classi) = ½ \text{ (by assumption)}$$

$$P(error / x \in class_1) = P(x < m_T / x \in class_1)$$

$$= P\left[\frac{x - m_1}{\sigma_1} < \frac{m_T - m_1}{\sigma_1}\right] = P\left[N(0,1) < \frac{m_T - m_1}{\sigma_1}\right] \tag{8}$$

$$= P\left[N(0,1) > \frac{m_1 - m_T}{\sigma_1}\right] = \varphi\left[\frac{m_1 - m_T}{\sigma_1}\right] \tag{9}$$

$$P(error / x \in class_2) = P(x < m_T / x \in class_2)$$

$$= P\left[N(0,1) > \frac{m_T - m_2}{\sigma_2}\right] \tag{10}$$

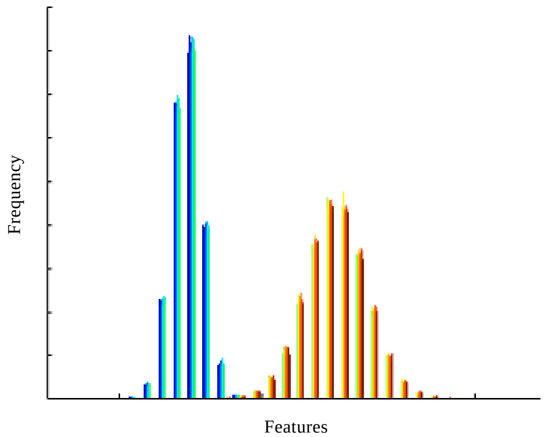

Fig. 4: Histogram of the features coefficients of two classes

$$= \varphi\left[\frac{m_T - m_2}{\sigma_2}\right] \tag{11}$$

$$P(\text{error} / x) = \frac{\varphi\left[\frac{m_1 - m_T}{\sigma_1}\right] + \varphi\left[\frac{m_T - m_2}{\sigma_2}\right]}{2} \tag{12}$$

$$\text{If we take } \left[\frac{m_T - m_2}{\sigma_2}\right] > 1 \text{ and } \left[\frac{m_1 - m_T}{\sigma_1}\right] > 1$$

$$P(\text{error}) < \phi(1) = 15\%$$

RESULTS AND DISCUSSION

The trend of low frequency coefficients of face images reduce the coefficients number and minimize the effect of varying facial expressions, shadow and variance in pose. These low frequency coefficients of ten images for randomly chosen 15 individuals out of a total of 40 individuals in ORL database of each class are retained as feature vector for classification while eleven images were chosen for each of the 15 individuals in YALE database. For matching and classification purpose in feature space, a natural choice is Euclidian distance metric:

$$D(V_{core}, V_{test}) = \sqrt{\sum_{i=1}^{k} (V_{core}^{i} - V_{test}^{i})^2} \tag{13}$$

The simulation result show that if we take the threshold to be equal to 1, the accuracy can be improved from 88% up to 99.33% on ORL database and from 63.03% up to 87.88% on YALE database.

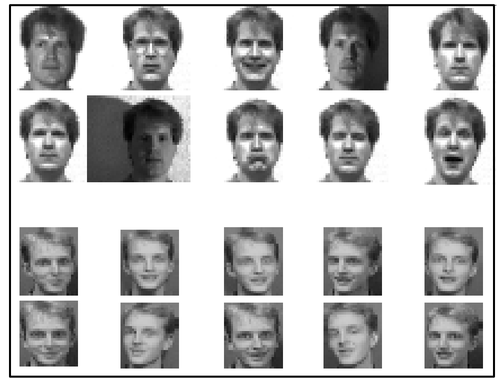

Fig. 5: Face images in the Yale face database (Top) and the ORL face database (Bottom)

Olivetti research laboratory face database: The presented statistical model is evaluated on the Olivetti Research Laboratory face database (ORL) (http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html). This database contains 10 different images for each of 40 people. The images of the same person are taken at different times, under slightly varying lighting conditions and different facial expressions. Some images have been captured with and without glasses. The head of the individuals in the images is slightly tilted or rotated.

Yale university database: The second database is YALE database which is established by Yale University (http://cvc.yale.edu/). Eleven images of 15 persons with total of 165 images for our experiment. The images of each person are with different facial expression or configuration: like center-light, with/without glass, happy, sad, left-light with/without glass, normal, right-light, sad and sleepy, surprised and wink. Few examples of these images are shown in Fig. 5.

In previous experiments the idea of Eltoukhy *et al*., (2009) for selecting the biggest coefficients from the each level of decomposition were tested. During the previous tests it was found that the Symelt wavelets were the best wavelets for face classification. The selection of Symlet wavelets is based on their properties (Guo-Sheng *et al*., 2006). The main problem in method (Eltoukhy *et al*., 2009) is that it is sensitive to variation in pose, illumination and expressions. The selection of the biggest coefficients doesn't represent the important face features and has an impact on the classification process in general. In the experiments, distribution of the features was given more attention and two statistical methods were used to select the significant features based on this distribution. Therefore, the effectiveness of the classification process is less sensitive to any type of variation.

Table 1: Comparison between the biggest coefficients method and the new two variance methods

Method	ORL database (%)	YALE database (%)
Chosen biggest 200 coefficients	88.00	63.03
Variance_mod	99.33	88.48
Variance_min	99.33	87.88

Table 2: The accuracy with different threshold values with YALE database

Threshold	No. of coefficients	Accuracy (%)
0.5	85757	86.67
0.7	78833	86.67
1	71644	87.88
1.2	68325	88.48
1.5	64607	87.88
1.7	62454	87.88
2	59595	87.88

Table 3: The accuracy with different threshold values with ORL database

Threshold	No. of coefficients	Accuracy (%)
0.5	8035	99.33
0.7	7344	99.33
1	6741	99.33
1.2	6455	99.33
1.5	6087	99.33
1.7	5872	99.33
2	5607	98.67

In Table 1 the comparison between the method by Eltoukhy *et al.* (2009) when it is used for face recognition (classification) and the proposed methods has been given. The simulation result shows that accuracy improvement has been achieved from 88 up to 99.33% on ORL database and from 63.03 up to 88.48% on YALE database.

The method in the first row takes the biggest 200 coefficients for classification process then the result shows that this method is not efficient for representing the face.

The two methods proposed in row 2 and 3 in Table 1 resulted in the increase of classification accuracy when they are applied on the same two databases and results were close to each other.

Table 2 and 3 show the optimum threshold to reduce features extracted by Symlet wavelets. In the simulation, the original features extracted by symlets wavelets were 125440 for YALE database and 22848 for ORL database. But this number of coefficients has been significantly reduced to the values indicated in Table 2 and 3 for different threshold limits by applying the statistical model. In Table 2, it can be seen that accuracy achieved in row 4 is the highest compared with the rest. In Table 3, the result shows that the accuracy values remained at 99.33% until a threshold value of 1.7 and reduced down to 98.67% for threshold value of 2.

In Table 4 and 5 other values of threshold have been tested to ensure that the optimum threshold values found in Table 2 and 3 is the optimum value. In addition, it can be seen that the highest threshold values were chosen for the test and the result indicates the decrease in the classification accuracy.

Table 4: The accuracy with other different threshold values with YALE database

Threshold	Coefficients	Accuracy (%)
0	125440	86.06
1	71644	87.88
10	37805	87.27
50	22398	79.39
100	17848	75.15
200	14552	72.735
500	11506	65.45
1000	9844	66.06

Table 5: The accuracy with other different threshold values with ORL database

Threshold	Coefficients	Accuracy (%)
0	22848	99.33
1	6741	99.33
10	3155	94.67
50	1237	92.67
100	818	88.67
200	585	86.00
500	73	78.00
1000	1	21.33

CONCLUSION

In this study wavelet analysis of multi-level of decomposition is used to transform the facial images by using Symlet8 wavelets in order to produce features' coefficients and two new statistical models were applied to these coefficients to reduce the redundancy ignoring all coefficients for which threshold value are lower. At last, EDM was used as a simple classifier. The results showed that, the two new variances increased the accuracy up to 40% which is more than the accuracy obtained when using the biggest wavelet coefficients. Besides, the simplest classifier affects the classification accuracy. From the final results, it was found that the optimum threshold values were (0.065) for the first method and (1.2) for the second method.

Moreover, works are underway to further increase the classification accuracy by using new statistical methods.

REFERENCES

Dan, Z. and S. Wentao, 2009. The optimal morlet wavelet and its application on mechanical fault detection. Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing, WiCom'09, Sept. 24-26, Beijing, pp: 1-4.

Eltoukhy, M.M., I. Faye and B.B. Samir, 2009. Breast cancer diagnosis in digital mammogram using multi-scale curvelet transform. Computerized Med. Imaging Graphics, 34: 269-276.

Ghosal, V., P. Tikmani and P. Gupta, 2009. Face classification using gabor wavelets and random forest. Proceedings of the 2009 Canadian Conference on Computer and Robot Vision, May 25-27, Kelowna, BC, pp: 68-73.

Guo-Sheng, H., R. Zhen and T. Yong-Jun, 2006. Symmetrical wavelet construction and its application for electrical machine fault signal reconstruction. Proceedings of the International Conference on Machine Learning and Cybernetics, Aug. 13-16, Dalian, China, pp: 3734-3737.

Karimi, R.H., J.P. Maralani, B. Lohmann and B. Moshiri, 2006. Haar wavelet-based optimal control of time-varying state-delayed systems: A computational method. Proceedings of the Computer Aided Control System Design, IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control, Oct. 4-6, Munich, pp: 1528-1533.

Misiti, M., Y. Misiti, Y.G. Oppenheim and J.M. Poggi, 2006. Wavelets and their applications, digital signal and image processing series. http://portal.acm.org/citation.cfm?id=1207627.

Vonesch, C., T. Blu and M. Unser, 2007. Generalized daubechies wavelet families. IEEE Trans., 55: 4415-4429.

Xia, X.G. and B.W. Suter, 1996. Vector-valued wavelets and vector filter banks. IEEE Trans. Signal Process., 44: 508-518.