Feature Extraction using Gabor and DWT for GMM based Face Verification Algorithms

Jesus Olivares-Mercado, Gabriel Sanchez-Perez, Mariko Nakano-Miyatake, Hector Perez-Meana

Postgraduate Section of Mechanical Electrical Engineering School, Instituto Politécnco Nacional, Av. Santa Ana no. 1000 Col. San Francisco Culhuacan, Mexico D.F. Mexico
hmpm@prodigy.net.mx

(Paper received on July 01, 2007, accepted on September 1, 2007)

Abstract. This paper presents two feature extraction methods for developing face verification algorithms based on the Gaussian Mixtures Model (GMM). The first one using the discrete wavelet transform, (DWT) while the second is based on the discrete Gabor transform (DGT). In both cases, firstly the feature extraction is carried out using either the DGF or the DWT. Next the Gaussian Mixture Model (GMM) is used to perform the face verification task. Evaluation results using the standard data bases with different parameters, such as the mixtures number, the number of faces used for training as well as the transform used for feature extraction show that proposed system provides better results that other previously proposed systems with a correctly detections larger than 95%, using any of these transforms,. Although, as happens in must face recognition systems, the verification rate decreases when the target faces present some rotation degrees.

1 Introduction

The development of security systems based on biometric features is currently a topic of active research, because it has a great importance in the development of the identity verification systems for access control, to enforce the security in restricted areas, and several other security applications. The terrorist attacks that have happened during the last decade have done evident the necessity of developing more reliable security systems, in offices, banks, companies, trades, etc. Among them the identity verification based on biometric methods appear to be a good alternative for the development of such security systems.

The biometrics systems consist of a group of automated methods for recognition or verification of people identity using physical characteristics or personal behavior of the person under analysis [1]. This technology is based on the fact that each person is unique and possesses distinctive features that can be used to identify her/him. Following these ideas several biometric based security systems have been developed using fingerprints, iris, voice hand and face features. Among them, the face verification systems appear to be a desirable alternative because in is non-invasive and its computational complexity is low, it is the biometric method easier of understanding

© H. Sossa, R. Barrón, E. Felipe (Eds.)

since for us the face is the most direct way to identify people; besides that the data acquisition of this method consists on taking a picture, doing it one of the biometric methods with larger acceptance among the users.

The recognition is a very complex activity of the human brain. For example, we can recognize hundred of faces learned throughout our life and to identify familiar faces at the first sight even after several years of separation with relative easy. However for a computer it is not a simple task. For instance, recently proposed face recognition systems, achieve a recognition rate of about 91% when the face image is not rotated or the rotation is relatively low. However although, this recognition rate is good enough for several practical applications, it may be so large for applications where the security should be extreme; such that we cannot tolerate a high erroneous recognition average. This paper proposes a face recognition algorithm that is able of achieving an erroneous verification rate below 9%. Several methods have been proposed for face recognition [2], [3] such as the methods based on statistical correlation of the geometry [4]; the face form which uses the distances among the position of the eyes, mouth, nose, etc. as well as those using the neuronal networks technology that trait to imitate the operation of the human brain [2]. Many of these systems can recognize a person even when they present some physical changes, such as the growth of the beard or mustache, changes in the color or the style of the hair, the use of glasses, etc. Although in general are sensitive to rotations of the face images.

Before starting the proposed methods analysis used for face recognition, it is necessary to point out the verification concept. In face verification, the person says to the system about his/her identity, presenting an identification card or writing a special password. The system holds the person's features (for example the persons face in this case), and then proceeds to solve if the person is who (his/her) claims to be.

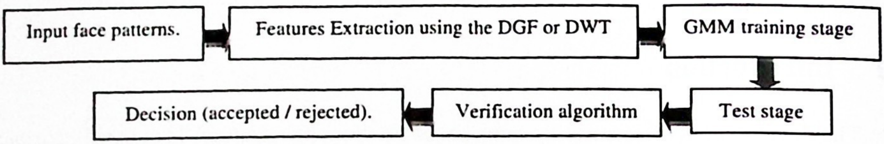


Fig. 1. Proposed face verification algorithm.

2 Proposed System

This section provides a detailed description of the proposed face verification algorithm which consists of three stages. Firstly a feature extraction of the face is carried out, using either the Gabor discrete transform (DGF) or discrete wavelet transform (DWT). Next using these features vectors, a model for each face is obtaining using a Gaussian Mixtures Model (GMM). Finally during the verification process, the GMM output is used to take the final decision. Figure 1 shows the block diagram of proposed algorithm.

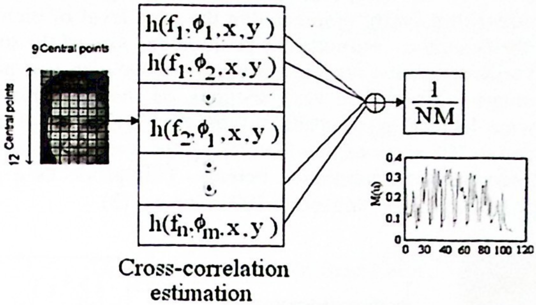


Fig. 2. DGT model.

2.1 Feature extraction with DGT

The feature extraction stage is one of the most important steps in any pattern recognition system. To this end, the proposed algorithm uses the DGT which has some relation with the human visual system (HVS). The two dimensional discrete Gabor functions (2D-DGF) depends on four parameters, two of them express their localization in the space (x, y) while the other two express the spatial frequency, f_m, and the orientation ϕ_n, where $m=1,2,..N_f$ and $n=1,2,..,N_\phi$ [5]. Thus to estimate the features vector, firstly the captured image (NxM) is divided in M_xM_y receptive fields each one of size $(2N_x+1)$x$(2N_y+1)$ (Fig. 2), where $N_x=(N-M_x)/2M_x$, $N_y=(M-M_y)/2M_y$. This fact allows that the number of elements in the features vector be independent of the captured image size. Next, the central point of each receptive field whose coordinates are given by (c_i,d_k), where $i=1,2,..,N_x$; $k=1,2,3,\ldots,N_y$ are estimated. Subsequently the first point of the cross-correlation $\psi(u,v)$ between each receptive field and the N_fN_ϕ Gabor functions $h_{m,\phi}(x,y)$ is estimated, where

$$h_{f,\phi}(x,y)=g(x',y')exp(2\pi jf_m(x'+y')) \tag{1}$$

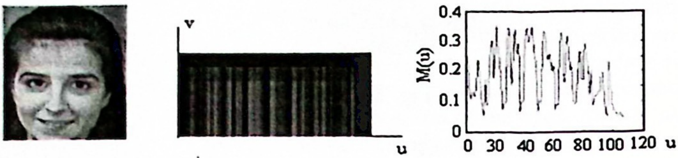
where denotes the Gabor function, and

$$(x',y')=((x\cos\phi_n+y\sin\phi_n),(-x\sin\phi_n+y\cos\phi_n)) \tag{2}$$

As shown in Fig. 1, N_fN_ϕ correlations are estimated for each receptive field, leading to an extremely large features vector. Thus to reduce the elements in the features vector, the first point of the total cross correlation between each receptive field and the set of DGF is used, which can be obtained taking the average of $\psi(u,v)$ with respect to v. Therefore the proposed algorithm features vector $M(u)$, is given by

$$M(u)=\frac{1}{N_v}\sum_{v=1}^{N_v}|\psi(u,v)| \tag{3}$$

where $N_v = N_f N_\phi$. To do the proposed method robust against changes of sizes and translation; the algorithm firstly assumes that the gray level of picture background is constant. Next the algorithm estimates the position and size of the image by analyzing the gray label variation on the image. Once the image size and position have been estimated, the image is divided in 12x9 sections, as shown in Fig. 2, whose central point will be always located in the space position (x, y), where x=0 and y=0. After the image was divided in 108 sections, the features vector was estimated with 9 phases and 6 normalized frequencies as mentioned before. This produces a matrix with 5832 elements that are subsequently reduced to 108 using eq. (3).



(a) (b) (c)

Fig. 3. a) Original image. b) Cross correlation matrix between DGF and receptive fields. c) Features vector obtained of proposed algorithm.

2.2 Feature extraction with DWT

The most commonly used set of discrete wavelet transforms was formulated by the Belgian mathematician Ingrid Daubechies in 1988. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is a half of that of the previous scale.

The DWT of a given signal *x* is estimated by passing it through a series of low pass and high pass filters (Fig. 4. First the samples are passed through a low pass filter with impulse response *g(n,m)* resulting in a convolution of the two. The signal is also decomposed simultaneously using a high-pass filter *h(n,m)*. The detail coefficients are the high-pass filter outputs and the approximation coefficients are the low-pass ones. It is important that the two filters, related to each other, are known as a quadrature mirror filter. However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are:

$$Y_{LOW}(n,m) = \sum_{j=-\infty} \sum_{k=-\infty} x(n,m) g(2n-k, 2m-j) \tag{4}$$

$$Y_{HIGH}(n,m) = \sum_{j=-\infty} \sum_{k=-\infty} x(n,m) g(2n-k, 2m-j) \tag{5}$$

This decomposition reduces the spatial resolution since only a quarter of each filter output allows characterizing the face image. However, because each output has band width equal to a quarter of the original one, the output image can be decimated to reduce the image size.

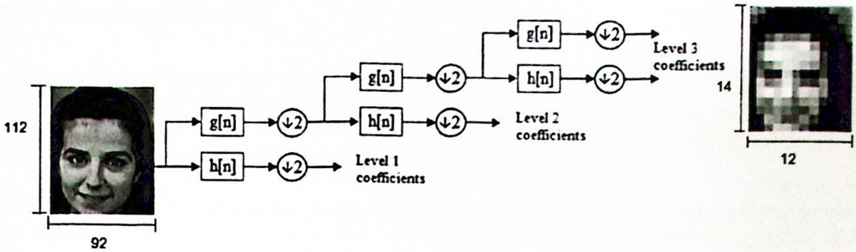


Fig 4. 3 level wavelet decomposition

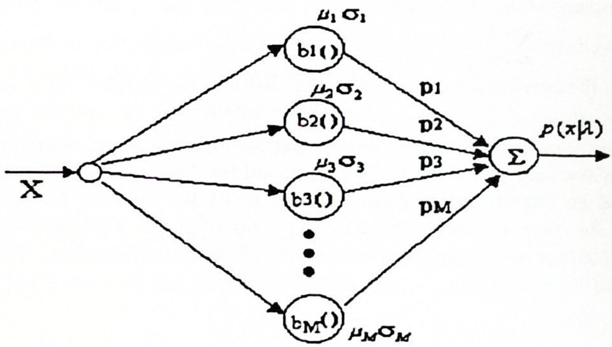


Fig. 5. Gaussian Mixture Model

Here only the approximation coefficients are used to characterize the face image. This decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled. This is represented as a binary tree with nodes representing a sub-space with different time-frequency localization. The tree is known as a filter bank.

At each level in the above diagram the signal is decomposed into low and high frequencies. Due to the decomposition process the input signal must be a multiple of 2^n where n is the number of levels. For example a signal with 32 samples, frequency range 0 to f_n and 3 levels of decomposition, 4 output scales are produced:

2.3 Face verification stage

To perform the face verification task a GMM will be used because, the GMM, which consists of a sum of M weighted Gaussian density functions, is able to approximate any probability distribution if the number of Gaussian components is large enough. Consider the GMM shown in Fig. 5 which is described by the following equation [6]:

$$p(\bar{x} / \lambda) = \sum_{i=1}^{M} p_i b_i(\bar{x}) \tag{6}$$

where $\bar{x}$ is a N-dimensional vector, $b_i(\bar{x}), i=1,2,...M$, are the density components and p_i, $i=1,2,...M$, are the mixture weights. Each density component is a D-dimensional Gaussian function given as:

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2}|\sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\bar{x}-\bar{\mu}_i)' \sigma_i^{-1}(\bar{x}-\bar{\mu}_i)\right\} \tag{7}$$

Where ( )' denotes the transpose vector, μ_i denotes the mean vector and σ_i the covariance matrix which is assumed to be diagonal; and p_i are the mixture weights which satisfies that $\sum_{i=1}^{M} p_i = 1$. The model of the distribution is determined by the mean vector, the covariance matrix and the distribution weights for a face model is given by $\lambda = \{p_i, \mu_i, \sigma_i\}$ $i = 1,2,..,M$. The estimation of the optimal parameter is a non-linear problem, such that we need an iterative algorithm to estimate the optimal parameters of the face verification algorithm, and the Maximum Likelihood algorithm (ML) is used to search the optimal parameters of the system, providing the best approach to the face model under analysis. So then the goal is to find the best parameters of λ that maximize the a posteriori probability distribution. For a sequence of vectors T of training X={x_1,...,x_T}, GMM likelihood can be written as [7]:

$$p(X / \lambda) = \prod_{t=1}^{T} p(X_t / \lambda) \tag{8}$$

Equation (8) is not lineal in relation of the parameters λ, then is necessary to carry out the estimation in an iterative way using the Expectation-Maximization algorithm (EM), in which starting from an initial set of parameters *λ(r-1)* and a new model is estimated *λ(r)*, where r denotes the r-th iteration, so that:

$$p(X / \lambda(r)) \geq p(X / \lambda(r-1)) \tag{9}$$

To achieve this goal, each T partial feature vectors $\mathbf{X}_t$, of the GMM parameters are updated [7]. During the testing phase we need to estimate the probability that one face under analysis corresponds to a given model, that is $P_r(\lambda/X)$. To achieve this, the Bayes theorem is used, then obtain:

$$\hat{R} = \sum_{t=1}^{T} \log_{10}(p(X_t / \lambda)) \tag{10}$$

where $p(X/\lambda)$ is the conditional probability of the face **X** given by face model λ, this is the GMM system response shown in the Fig. 6.

3 Evaluation Results

The evaluation of proposed system was carried out by computer simulations using the database created by Olivetti Research Laboratory in Cambridge, UK (ORL), which consists of images of 30 people with 10 images of each one which differs on, face rotation, different inclination, etc. The images have a size of 92 x 112 pixels.

After the feature vector is estimated using the DGT or the DWT it is applied to a GMM for obtaining the model of the face under analysis. This is achieved introducing the estimated vectors to the Gaussian mixture model (GMM) to obtain the weights, the mean and variance as described in [6], [7], where for training we assume that the 108 elements obtained in the feature extraction stage are divided in partial features vector of L elements as follows

$$S = \{X_0, X_1, X_2, \ldots, X_T, X_{T+1}, X_{T+2}, X_{T+3}, \ldots, X_{L-1}\} \tag{11}$$

Subsequently, form a group of vectors in L segments with T features vectors, Xt, each one in the following way:

$$S_0 = \{X_0, X_1, X_2, X_t, X_{t+1}, \ldots, X_T\} \tag{12}$$

$$S_k = \{X_k, X_{k+1}, X_{k+2}, \ldots, X_t, X_{t+1}, \ldots, X_{T+k}\} \tag{13}$$

In this work the system uses T=12 in order that each 12 feature vectors of the GMM parameters being updated. In the table 1 and 3 the face rejection results using the DGT DWT respectively are shown, the Fig. 2 and 4 show the false acceptation rate using DGT and DWT respectively.

Table 1. False rejection error using DGT.

# of faces of training	# Gaussian Mixtures	Low threshold %	Half threshold %	high threshold %
	9	12	8	5.1
1 face	16	18.5	12.9	8.7
	32	12.4	9.2	6.8
	9	12	7.7	4.7
3 faces	16	13.25	9.1	5.9
	32	8.66	6.48	4.6
	9	9.3	5.9	4
5 faces	16	10.9	7	4.4
	32	5.9	4.6	3.9

Table 2. False acceptation error using DGT.

# of faces of training	# Gaussian Mixtures	Low threshold %	Half threshold %	high threshold %
	9	11.57	6.71	3.14
1 face	16	18.5	12.35	7.7
	32	11.8	8.34	5.62
	9	11.4	6.5	3.13
3 face	16	12.5	8	4.5
	32	7.42	5	2.9
	9	8.7	4.5	2.2
5 face	16	10.2	5.8	2.9
	32	4.4	2.8	2

Table 3. False rejection error using DWT.

# of faces of training	# Gaussian Mixtures	Low threshold %	Half threshold %	high threshold %
	12	4.55	4	3.41
1 face	16	5.53	4.2	3.3
	32	9.8	7.9	6.1
	12	1.5	2.18	2.26
3 faces	16	8.8	7.32	6.08
	32	16.7	14.8	13.3
	12	6.7	5.9	5.3
5 faces	16	3.6	3.2	2.9
	32	3.4	3	2.9

Table 4. False acceptation error using DWT.

# of faces of training	# Gaussian Mixtures	Low threshold %	Half threshold %	high threshold %
	12	3.82	2.63	1.74
1 face	16	4.54	3	1.9
	32	9.2	7.11	5.1
	12	0.66	0.56	0.43
3 face	16	8	6.45	5.11
	32	16.3	14.3	12.7
	12	5.8	4.3	3.34
5 face	16	1.31	0.9	0.44
	32	1.45	1	0.8

In the verification stage a threshold is used which depends of the face under analysis. Top improve the verification performance this threshold may be divided in three categories: low, half and high thresholds. Two variants more were introduced for evaluation: one is the numbers of faces used for the training and the second one is the numbers of Gaussians mixtures to be used and these variants are applied in the two used techniques

Simulation results show that proposed algorithm performs fairly well in comparison with other previously proposed methods [1], [2], [6], even with faces that present an appreciable rotation, as happens in the ORL database. We can also see that there is not a much difference between using DWT and DGT.

4 Conclusions

This paper presented two face verification algorithms in which the DGT or DWT are used for feature extraction and the GMM to perform the verification task. Evaluation results obtained were very different with each one of the variants proposals, from 18.5% in the worst case, until 3.9% in the best case, using DGT, and 16.7% in the worst case, until 2.9% in the best case, using DWT. These results are very satisfactory if we consider that the database used is composed by 30 persons. This quantity of face is very similar to any database in a real application. In the case of accepting a person with a false identity we have a percentage of error of 18.5% at worst case and a 2% in the best one using the DGT; while using the DWT we have a 16.3% cases at worst case; and a 0.44% in the best one using DWT. In summary, can observe that the system performance becomes better when more faces are used for training the GMM and it has a larger number of mixtures. This is valid for false acceptation error as well as for false rejection error.

Acknowledgements

We thanks the National Science and Technology Council and to the National Polytechnic Institute for the financial support during the realization of this research.

References

1. P. Reid: BIOMETRICS for Networks Security, New Jersey: Prentice Hall, 2004, pp 3-7.
2. R. Chellappa, C. Wilson, and S. Sirohey: Human and Machine Recognition of Faces: A Survey, Proc. IEEE, vol. 83, no. 5, pp. 705-740, 1995.
3. A. Shashua: Geometry and Photometry in 3D Visual Recognition, PhD thesis, Massachusetts Institute of Technology, 1992.
4. Robert J. Baron: Mechanisms of human facial recognition. International Journal of Man-Machine Studies, pp: 137-178, 1981.
5. Dunn, D., Higgins, W. E.: Optimal Gabor Filters for Texture Segmentation, IEEE Trans. Image Proc., Vol. 4, No. 7, Jul. 1995.
6. Jin Y. Kim, Dae Y. Ko, Seung Y. Na.: Implementation and Enhancement of GMM Face Recognition Systems Using Flatness Measure, IEEE Robot and Human Interactive Communication, September 2004.
7. Douglas A. Reynolds, Richard C. Rose: Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models, IEEE Trans. Speech and audio Proc., Vol. 3, No. 1. Jan. 1995.

8. Busso C., Narayanan S. S.: Interrelation Between Speech and Facial Gestures in Emotional Utterances: A Single Subject Study, IEEE Transactions on Audio, Speech, and Language Processing, Vol. PP 2007.
9. Yan, S., Xu, D., Tang, X.: Face Verification With Balanced Thresholds, Transactions on Image Processing, IEEE, Vol. 16, No. 1, Pag. 262- 268, Jan 2007.
10. Kar, Soumitra, Hiremath, Swati, Joshi, Dilip G.: Chadda, Vinod.K.; Bajpai, Apurva, "A Multi-Algorithmic Face Recognition System", International Conference on Advanced Computing and Communications, Pag. 321 – 326, December 2006.