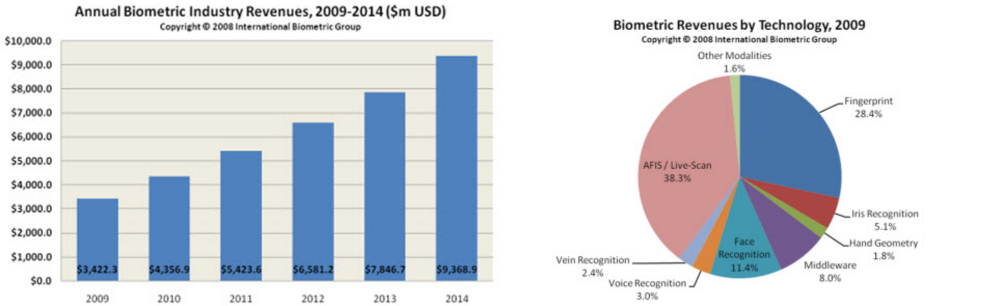
4

Facial Identification Based on Transform Domains for Images and Videos

Carlos M. Travieso, Marcos del Pozo-Baños and Jesús B. Alonso
University of Las Palmas de Gran Canaria, Signals and Communication Department,
Institute for Technological Development and Innovation in Communications,
Campus Universitario de Tafiera, Las Palmas de Gran Canaria,
Spain

1. Introduction

For the last decade, researchers of many fields have pursued the creation of systems capable of human abilities. One of the most admired humans' qualities is the vision sense, something that looks so easy to us, but it has not been fully understood jet. In the scientific literature, face recognition has been extensively studied and; in some cases, successfully simulated.
According to the Biometric International Group, nowadays Biometrics represent not only a main security application, but an expanding business according to Fig. 1a. Besides, as it can be seen in Fig. 1b, facial identification has been pointed out as one of the most important biometric modalities (Biometric International Group 2010).
However, face recognition is not an easy task. Systems can be trained to recognize subjects in a given case. But along the time, characteristics of the scenario (light, face perspective, quality) can change and mislead the system. In fact, the own subject's face varies along the time (glasses, hats, stubble). These are major problems with which face recognition systems have to deal using different techniques.
Since a face can appear in whatever position within a picture, the first step is to place it. However, this is far from the end of the problem, since within that location, a face can present a number of orientations. An approach to solve these problems is to normalize space position; variation of translation, and rotation degree; variation of rotation, by analyzing specific face reference points (Liun & He, 2008).
There are plenty of publications about gender classification, combining different techniques and models trying to increase the state of the art performance. For example, (Chennamma et al., 2010) presented the problem of face or person identification from heavily altered facial images and manipulated faces generated by face transformation software tools available online. They proposed SIFT features for efficient face identification. Their dataset consisted on 100 face images downloaded from http://www.thesmokinggun.com/mugshots, reaching an identification rate up to 92 %. In (Chen-Chung & Shiuan-You Chin, 2010), the RGB images are transformed into the YIQ domain. As a first step, (Chen-Chung & Shiuan-You Chin, 2010) took the Y component and applied wavelet transformation. Then, the binary two dimensional principal components (B2DPC) were extracted. Finally, SVM was used as classifier. On a database of 24 subjects, with 6 samples per user, (Chen-Chung & Shiuan-You Chin, 2010) achieved an average identification rate between 96.37% and 100%.



(a) (b)

Fig. 1. Evolution on Biometric Market and Modalities according to International Biometric Group.

In (Zhou & Sadka, 2010) approach, while diffusion distance is computed over a pair of human face images, shape descriptions of these images were built using Gabor filters consisting of a number of scales and levels. (Zhou & Sadka, 2010) used the Sheffield Database, which is composed of 564 face images from 20 individual persons, mixed race/gender/appearance, along with the MIT-CBCL face recognition database that contains images of ten subjects and 200 images per user. They run experiments comparing the proposed approach against several competing methods and the proposed Gabor diffusion distance plus k-means classification (“GDD-KM”), reaching a success rate over 80%. (Bouchaffra, 2010) presented a machine learning paradigm that extends an HMM state-transition graph to account for local structures as well as their shapes. This projection of a discrete structure onto a Euclidean space is needed in several pattern recognition tasks. In order to compare the COHMMs approach proposed with the standard HMMs, they made a set of experiments using different wavelet filters for feature extraction with HMMs-based face identification. GeorgiaTech, Essex Faces95 and AT&T-ORL Databases were used. Besides, FERET Database was used on evaluation or test. Identification accuracies over 92.2% by CHMM-HMM, and 98.5% by DWT/COHMM were achieved.

In (Kisku et al., 2007), the database and query face images are matched by finding the corresponding feature points using two constraints to deal with false pair assignments and optimal feature sets. BANCA database was used. For this experiment, the Matched Controlled (MC) protocol was followed. Computing the weighted Error Rate prior EER on G1 and G2 for the two methods: gallery image based match constraint, and reduced point based match constraint, the error reached was between 8.52% and 4.29%. In (Akbari et al., 2010), a recognition algorithm based on feature vectors of Legendre moments was introduced as an attempt to solve the single image problem. A subset of 200 images from FERET database and 100 images from AR database were used in their experiments. The results achieved 91% and 89.5% accuracy for AR and FERET, respectively.

In (Khairul & Osamu, 2009) the implementation of moment invariants in an infrared-based face identification system was presented to develop a face identification system in thermal spectrum. A hierarchical minimum distance measurement method for classification was used. The performance of this system is encouraging with 87% of correct identification rate for test to registered image ratio of 2:1 and 84% of correct identification rate for test to registered image ratio of 4:1. Terravic facial IR database was used on this work. (Shreve et al, 2010) presented a method for face identification under adverse conditions by combining regular, frontal face images with facial strain maps using score-level fusion. Strain maps were generated by calculating the central difference method of the optical flow field obtained from each subject's face during the open mouth expression. Extended Yale B database was used on this work, only the P00A+000E+00 image of each of the 38 subjects were used for the gallery and the remaining 2376 images were used as probes to test the accuracy of the identification system. Success rates achieved between 88% and 98% depending of the level of difficulty.

(Sang-Il et al., 2011) presented an approach that can simultaneously handle illumination and pose variations to enhance face recognition rate (Sang-Il et al., 2011). The proposed method consists of three parts which are pose estimation (projects a probe image into a low-dimensional subspace and geometrical distribution of facial components), shadow compensation (first they calculate the light direction) and face identification. This work was implemented under CMU-PIE and Yale B Databases, reaching a success rate between 99.5% and 99.9%. In (Yong et al., 2010) a three solution schemes for LPP (mathematical), using a k-nearest-neighbour as classifier was proposed and checked with ORL, AR, and FERET Databases. The success rates were 90%, 76% and 51%, respectively. (Chaari et al., 2009) developed a face identification system and used the reference algorithms of Eigenfaces and Fisherfaces in order to extract different features describing each identity. They built a database partitioning with clustering methods which split the gallery by bringing together identities with similar features and separating dissimilar features in different bins. Given a facial image that they want to establish its identity, they computed its partitioning feature that they compared to the centre of each bin. The searched identity is potentially in the nearest bin. However, by choosing identities that belong only to this partition, they increased the probability of an error discard of the searched identity. XM2VTS database was used reaching over 99% of success rate.

Techniques introduced in (Kurutach et al., 2010) were composed of two parts. The first one was the detection of facial features by using the concepts of Trace Transform and Fourier transform. Then, in the second part, the Hausdorff distance was employed to measure and determine of similarity between the models and tested images. Finally, their method was evaluated with experiments on the AR, ORL, Yale and XM2VTS face databases. The average of accuracy rate of face recognition was higher than 88%. In (Wen-Sheng et al, 2011), as each image set was represented by a kernel subspace, they formulate a KDT matrix that maximizes the similarities of within-kernel subspaces, and simultaneously minimizes those of between kernel subspace. Yale Face database B, Labeled Faces in the Wild and a self-compiled database. Success rates were achieved between 98.7% and 97.9%.

In (Zana & Cesar, 2006), polar frequencies descriptors were extracted from face images by Fourier–Bessel transform (FBT). Next, the Euclidean distance between all images was computed and each image was then represented by its dissimilarity to the other images. A pseudo-Fisher linear discriminant was built on this dissimilarity space. The performance of discrete Fourier transform (DFT) descriptors and a combination of both feature types was also evaluated. The algorithms were tested on a 40- and 1196-subjects face database (ORL and FERET, respectively). With five images per subject in the training and test datasets, error rate on the ORL database was 3.8, 1.25, and 0.2% for the FBT, DFT, and the combined classifier, respectively, as compared to 2.6% achieved by the best previous algorithm. In (Zhao et al., 2009), feature extraction was carried out on face images respectively through conventional methods of wavelet transform, Fourier transform, DCT, etc. Then, these image transform methods are combined to process the face images. Nearest-neighbour classifiers using Euclidean distance and correlation coefficients used as similarity are adopted to recognize transformed face images. By this method, five face images from a face database (ORL database) were selected as training samples, and the rest as testing samples. Success recognition rate was 97%. When five face images were from Yale face database, the correct recognition rate was up to 94.5%.

The methodology proposed in (Azam et al., 2010) is a hybrid approach to face recognition. DCT was applied to hexagonally converted images for dimensionality reduction and feature extraction. These features were stored in a database for recognition purpose. Artificial Neural Network (ANN) was used for recognition. Experiments and testing were conducted over ORL, Yale and FERET databases. Recognition rates were 92.77% (Yale), 83.31% (FERET), and 98.01% (ORL). In (Chaudhari & Kale, 2010) the process was divided in two steps: 1) Detect the position of pupils in the face image using geometric relation between the face and the eyes and normalizes the orientation of the face image. Normalized and non normalized face images are given to holistic face recognition approach. 2) Select features manually. Then determine the distance between these features in the face image and apply graph isomorphism rule for face recognition. Then apply a Gabor filter on the selected features. This Algorithm takes into account Gabor coefficient as well as Euclidean distance between features for face recognition. Brightness normalized and non normalized face images were given to feature based approach face recognition methods. ORL database was used, reaching over 99.5% for the best model. (Chi Ho Chan & Kittler, 2010) combined sparse representation with a multi-resolution histogram face descriptor to create a powerful representation method for face recognition. The recognition was performed using the nearest neighbour classifier. Yale Face Database B and the extended Yale Face Database B were used, achieving a success rate up to 99.78%.

In this chapter, we present a model to identify subjects from TV video sequences and images from different public databases. This model basically works with three steps: detect faces within the frames, undergo face recognition for each extracted face, and for videos, use information redundancy to increase the recognition rate. The main goal is the use of transform domains in order to reach good results for facial identification. Therefore, this chapter aims to show that our best proposal, applied on face images, can be extended to be used on videos, in particular on TV videos, developing for this purpose a real application. Results presented on the experiment sections show the robustness of our proposal on illumination, size and lighting variations of facial images.. Finally, we have used the inter-frame information in order to improve our approach on its use for video mode. Therefore, the whole system presents a good innovation.

The fist experimental setting is based on images from ORL and Yale databases; in order to determinate the best transform domains under illumination changes and without it. Besides, those databases have been used widely; therefore, we show our good results vs other approaches. We have applied different Transform Domains Feature Extractions, as Discriminative Common Vectors (DCV), Discrete Wavelet Transform (DWT), Independent Component Analysis (ICA), Discrete Cosine Transform (DCT), Linear Discriminant Analysis (LDA), and Principal Components Analysis (PCA). For the supervised classification, we have used Support Vector Machine (SVM), Neural Network (NN) and Euclidean Distance methods. Our proposal adjusts both parameterization and classification steps, and our best approach is finally applied to our TV video database (V-DDBB), composed of 40 videos.

2. Facial pre-processing

The first block is the pre-processing block. It gets the faces samples ready for the forthcoming blocks, reducing the noise and even transforming the original signal in a more readable one. An important property of this block is that it tries to reduce lightning variations among pictures. In this case, samples are first resized to standard dimensions of 20x20 pixels. This ensures that the training time will not reach unviable levels. Then images' grey scale histograms are equalized. Fig. 2 female face (left) shows the effect of apply this process to a given sample.

Fig. 2. Example of Pre-processed image.

Finally, for a new set of experiments, a local normalization function is added at the end of the block (Xiong, 2005). This function is based on a double Gaussian filtering, and makes the local mean and variance uniform along the picture. The effect of this new tool is dramatic and it can be seen in Fig. 2 male face (right).

3. Feature extraction

Once the samples are ready, the feature extractor block transforms them in order to obtain the best suited information for the classification step. Six types of feature extraction techniques are shown in this section.

3.1 Principal Component Analysis (PCA)

Principal components analysis (PCA) is a technique used to reduce multidimensional data sets to lower dimensions for analysis (Banu & Nagaveni, 2009). The applications include exploratory data analysis and generating predictive models. PCA involves the computation of the eigenvalue decomposition or singular value decomposition of a data set, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of scores and loadings. This process applied to face recognition is named blind source separation, where there are fewer sources than input channels.

The blind source separation consists in several sources that are mixed in a system, these mixtures are recorded together and they have to be separated to obtain the estimations of the original sources. The following figure shows the mixing system;

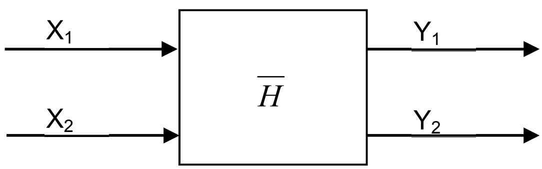


Fig. 3. Two Sources in a two mixtures system.

Generally, there are n source signals statistically independent $s(t)=[s_1(t),...,s_n(t)]$, and m observed mixtures that are linear and instantaneous combinations of the previous signals $x(t)=[x_1(t),...,x_n(t)]$. Beginning with the linear case, the simplest one, we have that the mixtures are:

$$x_i(t)=\sum_{j=1}^{n} h_{ij}\cdot s_j(t) \tag{1}$$

Now, we need to recover *s(t)* from *x(t)*. It is necessary to estimate the inverse matrix of H, where h_{ij} are contained. Once we have this matrix:

$$y(t)=\overline{W}\cdot x(t) \tag{2}$$

where *y(t)* contains the estimations of the original source signals, and is the inverse mixing matrix. Now we have defined the simplest case, it is time to explain the general case that involves convolute mixtures. The process is defined as follows:

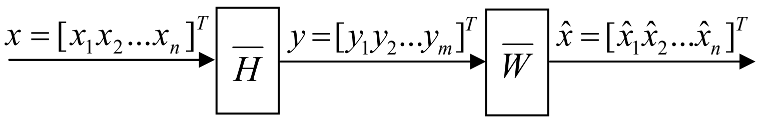


Fig. 4. BSS General Problem.

where $\overline{H}$ is the mixed system;

$$\overline{H}=\begin{bmatrix} h_{11} & ... & h_{1n} \\ ... & ... & ... \\ h_{n1} & ... & h_{nn} \end{bmatrix} \tag{3}$$

The h_{ij} are FIR filters, each one represents a signal transference multi-path function from source, *i*, to sensor, *j*. *i* and *j* represent the number of sources and sensors.

3.2 Discrete Wavelet Transform (DWT)

The Wavelet transform is another preprocessing and feature extraction technique wich can be directly applied to face images. The Discrete Wavelet Transform (DWT) (González & Woods, 2002) is defined as follows:

$$C[j,k]=\sum_{n\in\mathbb{Z}} f[n]\psi_{j,k}[n] \tag{4}$$

where $\psi_{j,k}$ is the transform function:

$$\psi_{j,k}[n]=2^{\frac{-j}{2}}\cdot\psi\left[2^{-j}n-k\right] \tag{5}$$

The application of different mother families on pre-processing (artefacts elimination) and on the feature extraction has got a set of good and discriminate parameters.

3.3 Lineal Discriminate Analysis (LDA)

The objective of LDA is the reduce of sample dimensionality while preserving all the information among possible classes (Huang et al., 2003). As opposed to the components analysis, the discriminate analysis seeks chiefly a projection that separates the data in the way of the quadratic error. Therefore in LDA, a different point of view is taken with respect to PCA.

LDA is a popular mapping technique when the labels of the different classes are known. This is an advantage since it can make use of this previous information, which gives a description of the disposition of the data.

It is obtained the "with dispersion" and "without dispersion" matrixes between classes, maximizing the first one and diminishing the second. This is carried out with the measure of the ratio among the projection of the "with dispersion" matrix determinant between classes and the projection of the "without dispersion" matrix determinant, which does that the projection among separation of classes be maximum.

The "without dispersion" matrix between classes is defined as:

$$S_w = \sum_{i=1}^{c} S_i \tag{6}$$

where c is the number of classes and:

$$S_i = \sum_{\mathrm{x}\in D_i} (\mathrm{x} - \mathrm{m}_i)(\mathrm{x} - \mathrm{m}_i)^T \tag{7}$$

for each set of D_i ,where:

$$\mathrm{m}_i = \frac{1}{n_i} \sum_{\mathrm{x}\in D_i} \mathrm{x} \tag{8}$$

On the other hand, the "with dispersion" matrix between classes is defined as:

$$S_b = \sum_{i=1}^{c} n_i (\mathrm{m}_i - \mathrm{m})(\mathrm{m}_i - \mathrm{m})^T \tag{9}$$

where n_i is the number of samples for each class i and m is the vector of total mean.

As in the case of PCA, where it is not necessary to separate the classes, the total dispersion matrix is defined as:

$$S_t = \sum_{\mathrm{x}\in D} (\mathrm{x} - \mathrm{m})(\mathrm{x} - \mathrm{m})^T \tag{10}$$

where D is the total matrix of data.

Then, in $S_t = S_w + S_b$ from (6) and (9), it is searched a projection, which satisfies:

$$\begin{aligned} W &= \arg\max_{w^*} \frac{|W^{*T} S_b W^*|}{|W^{*T} S_w W^*|} \\ &= [\mathrm{w}_1 \quad \mathrm{w}_2 \quad \cdots \quad \mathrm{w}_d] \end{aligned} \tag{11}$$

where W^* are the auto-values of $S_w^{-1}S_b$.
Considering S_w as nonsingular, the base-vectors of W correspond to the M greater eigen-values of $S_w^{-1}S_w = W^*$. The representation of M-dimensional space is obtained with the projection of the original data on the sub-space W, with M eigen-vectors; where M is lower than n.
This technique provides a tool for the classification that permits to diminish in a considerable way the calculation of the characteristics of the different samples. Moreover, it conserves the information among each class, what provides a greater level of discrimination among the samples.

3.4 Discrete Cosine Transform (DCT)

We have applied Discrete Cosine Transform (DCT) for noise elimination and details of high frequency (González & Woods, 2002). Besides, this transform has a good energy compaction property that produces uncorrelated coefficients, where the base vectors of the DCT depend only on the order of the transformation selected, and not of the statistical properties of the input data.
Another important aspect of the DCT is its capacity to quantify the coefficients utilizing quantification values, which are chosen of visual way. This transformation has had a great acceptance inside the image digital processing, because there is a high correlation among elements for the data of a conventional image.

3.5 Independent Component Analysis (ICA)

The main objective of the blind source separation (BSS) is to obtain, from a number of observations, the different signals that compose these observations. This objective can be reached using either a spatial or a statistical approach. The former is based on a microphone array and depends on the position and separation of them. It also uses the directions of arrival (DOA) from the different audio signals.
On the other hand, the statistical separation supposes that the signals are statistically independent, that they are mixed in a linear way and that it is possible to get the mixtures with the right sensors (Hyvärinen et al., 2001) (Parra, 2002). This technique is the newest and it is in a continuous development. It is used in different fields such as real life applications (Saruwatari et al., 2003) (Saruwatari et al., 2001), natural language processing (Murata et al., 2001), bioinformatics, image processing (Cichocki & Amari, 2001), etc.
Two main types of BSS problem can be differentiate: with linear and the nonlinear signal mixture. The former presents linear mixtures where the data is mixed without echoes or reverberations, while the mixtures of the latter are convolutive and they are not totally independent due to the propagation of the signal through dynamic environments. This more complex case is known as the "Cocktail party problem".
Depending on the mixtures, there are several methods to solve the BSS problem. The first case can be seen as a simplification of the second one.
In this work the statistical approach named Independent Component Analysis (ICA) is studied. ICA comes from the previously introduced PCA (Hyvärinen et al., 2001) (Smith, 2006). The BBS based on ICA is also divided into three groups; the first one are those methods that works in the time domain, the second are those who works in the frequency domain and the last group are those methods that combine frequency and time domain methods.

3.6 Discriminative Common Vectors (DCV)

The DCV approach was first introduced as a solution for the small sample size problem. This tool performs a double transformation on the training set. One transformation relies in the within-class properties, while the other plays in role in the inter-class domain.

In the first step, the DCV approach extracts the common properties of each class, making within-class samples more alike (Cevikalp et al., 2005) (Travieso et al., 2009). Let C denotes the number of classes, N_C the number of samples of each class, and x_n^c a d-dimensional column vector that represents the nth samples of cth class. Assume that the small sample size problem is present, thus $d > M - C$. In this case, the within-class S_W, between-class S_B, and total S_T projection matrixes can be defined as:

$$S_W = \sum_{c=1}^{C}\sum_{n=1}^{N}\left(x_c^n - \mu_c\right)\left(x_c^n - \mu_c\right)^T \tag{12}$$

$$S_B = \sum_{c=1}^{C}\left(\mu_c - \mu\right)\left(\mu_c - \mu\right)^T \tag{13}$$

$$S_T = \sum_{c=1}^{C}\sum_{n=1}^{N}\left(x_c^n - \mu\right)\left(x_c^n - \mu\right)^T \tag{14}$$

where μ is the mean of all samples, and μ_c is the mean of samples in the ith class. In order to obtain a common vector for each class, it is interesting to know that every sample can be decomposed as follow:

$$x_n^c = y_n^c + z_n^c \tag{15}$$

where y_n^c denotes the rank of S_W and z_n^c the null space of S_W .Therefore, the common vectors will be obtained in the null space as:

$$z_n^c = x_n^c - y_n^c \tag{16}$$

and the rank space can be obtained using the eigenvectors corresponding to the no null eigenvalues of S_W as a projecting matrix Q.

$$z_n^c = x_n^c + QQ^T x_n^c \tag{17}$$

It can be proven [8] that this formula drives to one unique vector for each class, the common vector.

$$x_{com}^c = z_n^c\,,\ \forall n \tag{18}$$

And the within-class dispersion matrix of these new common vectors is null:

$$S_W^{com} = 0 \tag{19}$$

In the second transformation, the differences between classes are magnified in order to increase the distance between common vectors, and therefore increase the inter-class dispersion. To achieve this, it is necessary to project into the rank of S_W^{com} , or similarly into the rank space of S_W^{com} (remember equations (14) and (19)). This can be accomplished by an eigenanalysis of S_W^{com}. Using eigenvectors corresponding to no null eigenvalues to form a new projecting matrix W, and apply it to the common vectors:

$$\beta^C = W^T x_{com}^c \tag{20}$$

Because all common vectors are the same for each class, the S_T^{com} function is calculated using only one common vector for each class.
This approach maximizes the DCV criterion:

$$\mathrm{J_{DCV}}\left(W_{op}\right) = \max_{\left|W^T S_W W\right|=0} \left|W^T S_T W\right| \tag{21}$$

4. Classification system

For this work, we have studied three supervised classification systems, based on Euclidean Distance (lineal classification), Neural Network (Bishop, 1991), and Support Vector Machines (Hamel, 2009) (Cristianini & Shawe-Taylor, 2000). With this set of classifiers, we can observe the behaviour of parameterization techniques versus the illumination conditions of a facial image.
The first classifier is a linear classifier; in particular, it is a Euclidean Distance. Its expression is,

$$\left\|x-y\right\|_e = \left[\sum_{i=1}^{D} \left|x_i - y_i\right|^2\right]^{1/2} \tag{22}$$

Another classifier is the Neural Network, in particular the perceptron score. The perceptron of a simple layer establishes its correspondence between classes with a lineal discrimination rule. However, it is possible to define discriminations for not lineally separable classes utilizing multilayer perceptrons, which are networks without refreshing (feed-forward) with one or more layers of nodes between the input layer and exit layer. These additional layers contain hidden neurons or nodes, which are directly connected to the input and output layer.
A neural network multilayer perceptron (NN-MLP) of three layers is shown in the figure 5 (Bishop, 1991), with one hidden layer. Each neuron is associated with a weight and a bias, these weights and biases of the connections of the network will be trained to make their values suitable for the classification between the different classes.
Particularly, the neural network used in the experiments is a Multilayer Perceptron (MLP) Feed-Forward with Back-Propagation training algorithm, and with only one hidden layer.
The third classifier is a SVM light. SVM light is an implementation of Vapnik's Support Vector Machine (Hamel, 2009) (Cristianini & Shawe-Taylor, 2000) for the problems of pattern recognition, regression, and learning a ranking function. The optimization algorithms used in SVM light are described in (Hamel, 2009) (Cristianini & Shawe-Taylor, 2000). The algorithm has scalable memory requirements and can handle problems with many thousands of support vectors efficiently. For this reason is very interesting for this applications, because we are going through thousands of parameters. The program is free for scientific use (Joachims, 1999) (svmlight, 2007).

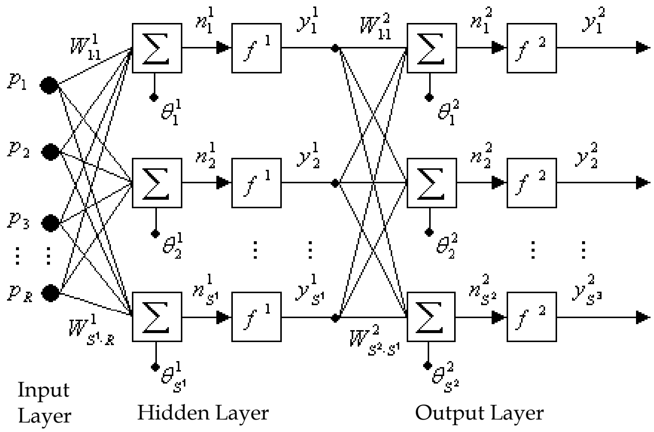


Fig. 5. Multilayer Perceptron.

In Fig. 6, we can see the detection of support vectors and the creation of boundary, one per each class, because this is a bi-class classifier. In our implementation, we have built a multi-classes classification module, from this SVM light.

5. Experimental settings

Our experiments are done on images and videos. The first step has been to study the effect of Transform Domain on facial images, and the best results to study on facial videos. In this work, six feature extractions and three classifiers have been used. All the experiments were done under automatic supervised classification.

5.1 Image and video face databases

Two data databases have been used for the first experiments, allowing to study the effects of illumination conditions on different techniques. These databases are ORL and Yale databases (AT&T Laboratories Cambridge, 2002) (Face Recognition Homepage, 2011). ORL is composed by 40 faces with 10 samples by face, having a total of 400 samples. The images are in grey scale (8 bits), and all the images have a dimension of 92×112 pixels. This database is independent of illumination conditions, because its light focus is fixed. Yale is composed by 15 faces with 11 samples by face, in total, it has 165 images. Again, the database is in grey scale with 8 bits, and the size of the images is 243×320 pixels. Each sample of this database has a different focus of illuminations. Samples from both databases are showed in Fig. 7.

We have also acquired a video database (V-DDBB) composed by 40 videos; 10 for each 4 subjects. This database is divided in two sets regarding the quality of the picture. Videos from the lower resolution V-DDBB present 32 frames per second with dimensions 208x160 and variable data speed between 39 and 46 kbps, while videos from the higher resolution V-DDBB present 29 frames per second with dimensions 320x240 and a constant speed rate of 2000 kbps. Information about length, size, and used can be find in tables 1, 2, 3, and 4, where 'Tr/Test' means that the use of the video can be either training or test depending on the

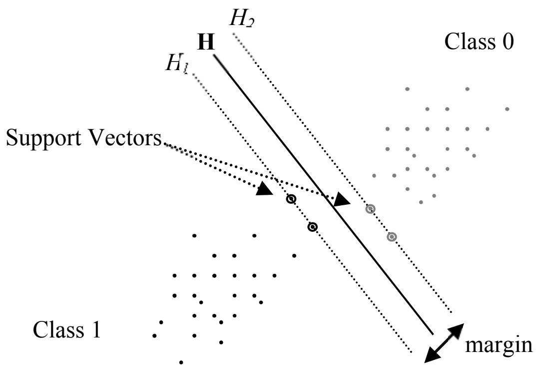


Fig. 6. Separate lineal Hyperplane in SVM.

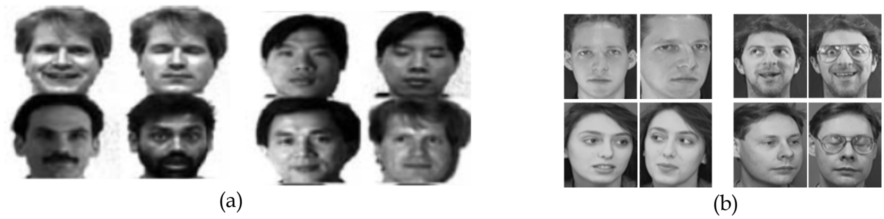


Fig. 7. Samples from both databases. (a) Yale Database. (b) ORL Database.

experiment. In any case, it is important to clarify that videos used for tests were not used for training the system.

Videos were originally captured from regular analogical TV emissions. However, the capture process was not the same for every video. Even though the aspect ratio range of AVI files remains almost constant, picture's resolution goes from a fuzzy image to an acceptable one. Moreover, lightning and background are very changing aspects from both inter-subject and intra-subject V-DDBB since each video came from a different TV program.

SUBJECT 1							
Lower Definition Data Base				Higher Definition Data Base			
Video	Length	Size	Used for	Video	Length	Size	Used for
1	02:35	5.75 MB	Training	1	01:17	12,40 MB	Training
2	01:01	2,51 MB	Training	2	01:41	14,00 MB	Training
3	01:21	3,81 MB	Training	3	01:03	4,80 MB	Training
4	01:43	4,56 MB	Tr/Test	4	02:45	28,80 MB	Tr/Test
5	01:40	4,43 MB	Test	5	01:21	12,40 MB	Test

Table 1. V-DDBB information for subject 1.

SUBJECT 2							
Lower Definition Data Base				Higher Definition Data Base			
Video	Length	Size	Used for	Video	Length	Size	Used for
1	01:35	4,09 MB	Training	1	01:03	5,51 MB	Training
2	01:42	4,06 MB	Training	2	00:58	4,45 MB	Training
3	02:00	4,70 MB	Training	3	01:06	4,49 MB	Training
4	02:07	4,95 MB	Tr/Test	4	01:40	12,0 MB	Tr/Test
5	01:27	3,47 MB	Test	5	01:40	14,0 MB	Test

Table 2. V-DDBB information for subject 2.

SUBJECT 3							
Lower Definition Data Base				Higher Definition Data Base			
Video	Length	Size	Used for	Video	Length	Size	Used for
1	01:33	3,87 MB	Training	1	02:42	14,5 MB	Training
2	03:28	10,2 MB	Training	2	02:30	16,0 MB	Training
3	02:10	5,53 MB	Training	3	10:03	40,3 MB	Training
4	02:39	7,26 MB	Tr/Test	4	01:24	10,5 MB	Tr/Test
5	03:01	9,27 MB	Test	5	00:46	2,38 MB	Test

Table 3. V-DDBB information for subject 3.

SUBJECT 4							
Lower Definition Data Base				Higher Definition Data Base			
Video	Length	Size	Used for	Video	Length	Size	Used for
1	01:54	4,59 MB	Training	1	01:37	10,3 MB	Training
2	01:01	2,58 MB	Training	2	00:59	6,52 MB	Training
3	02:10	5,38 MB	Training	3	02:26	12,1 MB	Training
4	02:20	5,46 MB	Tr/Test	4	00:20	2,65 MB	Tr/Test
5	01:00	2,58 MB	Test	5	00:56	5,05 MB	Test

Table 4. V-DDBB information for subject 4.

Focus was a major issue too as subjects were not always presented in the same way. Thus our system has to be tuned to recognize characters only from close up pictures. On Fig. 8 different examples with lower and higher definition are shown.

For the video database, OpenCV library were used to build up the Face Extractor Block (FEB), a library of programming functions mainly aimed at real time computer vision. For face detection, it uses Haar-like features to encode the contrasts exhibited by a human face and their spacial relationships. Basically, a classifier is first trained and subsequently applied to a region of interest (Bradski, 2011).

(a) (b)

Fig. 8. (a) Pictures from two different subjects of the higher quality Video Database. (b) Pictures from two different subjects of the lower quality Video Database.

The FEB receives an AVI XVID MPEG-4 video file and a sample rate parameter (SR) as inputs. It checks frames for faces every SR seconds. If any face is found, the FEB extracts and saves it as a JPEG picture. The number of the studied frame sequence is saved for the future time analysis. In order to delimit the quality of our future face database we imposed an aspect ratio of the extracted face and a minimum face size.

5.2 Experiments on facial images

The goal of this present work is to study and search a good identification system for different illumination conditions. Therefore, we have tested all parameterization tools with different classification techniques. In particular, our classification system has been tested with three different methods, Euclidean Distance (lineal classification), Neural Network (Bishop, 1991), and Support Vector Machines (Hamel, 2009) (Cristianini & Shawe-Taylor, 2000).

These classification systems have been used with supervised classification; and therefore, we have developed two modes in this process. The first mode is the training mode, where the system is trained with 50% of the database in use, while the remainder is used during the test mode. Our experiments have been repeated 10 times, and therefore, our results are showed as mean and variance.

Table 5 shows the results achieved for each classifier with different parameterization systems. It can be observed that the best system is that using SVM, for all parameterizations and both databases. Therefore, the best classifier with independency of illumination conditions is a SVM, based on Radial Basis Function (RBF) kernel. However, the same cannot be said for the parameterization technique, as there is not one dominant tool along different scenarios (see table 5). Besides, from this table 5 we can observe that in general the obtained results are quite robust.

About the computational time, we have studied the mean value with the best classifier. With the train mode, it has been obtained 30 seconds and with test mode, 0.5 seconds. Those computational times have been reached with MATLAB language (Matlab, 2011), which is an interpretative programming language. On future works, we expect to decrease the computational times by 5 to 7 times by migrating to C++ programming language.

Type of parameters	ORL Database			Yale Database		
	EDC	NN	SVM	EDC	NN	SVM
DWT - Haar	95,30% ±2,84	86,50% ±14,95	97,20% ±0,96	97,73% ±3,58	< 50%	98,67% ±1,58
DWT – Bior	96,80% ±1,23	< 50%	98,90% ±0,79	98,27% ±9,17	< 50%	99,07% ±1,60
LDA	95,00% ±6,83	90,90%±5,10	96,10% ±3,16	98,80% ±1,36	94,13% ±17,55	99,20% ±0,87
PCA	95,90% ±3,04	86,40%±11,82	96,15% ±2,89	98,13% ±4,43	89,60% ±10,59	98,80% ±2,15
DCT	94,90% ±3,04	95,55% ±1,69	96,96% ±1,19	97,87% ±3,64	78,67% ±19,40	99,07% ±1,60
ICA	87,85% ±6,95	< 50%	92,50% ±4.94	97,60% ±12,82	70,67% ±19,27	99,33% ±0,89
DCV	95,47% ±1,23	< 50%	98,21% ± 1,67	95,34% ±4,93	88,25% ±7,32	98,12% ±2,93

Table 5. Success Rates for ORL and Yale Databases.

From the Table 5, it is observed that DWT gives good success for image without illumination changes using SVM (ORL Database). For data with illumination changes (Yale Database), we can see ICA gives the better modelling. Therefore, those two techniques will be used on the next section.

5.3 Experiments on facial videos

As it is specified in tables 1, 2, 3, and 4, different videos were used for either training or test. A number of models with different setups were created in order to find the best configuration. Results for each experiment are presented in the following tables. Some of these videos present public events, therefore doing the pool of subjects immeasurable.

Both for ICA and DWT experiments different set ups were used, varying the training / test ratio. Moreover, for ICA experiments (see Table 6) the number of principal components used were also a parameter of configuration; ranging from 5 to 40. Higher classification rates of 76,87 % and 83,12 % were obtained for lower and higher resolution databases respectively.

For DWT experiments (see Table 7) the number of iterations ranged from 1 to 3. On the other hand, both 'bior 4.4' and 'haar' filters were used. In this case, higher classification rates of 86,25% and 98,75 % were obtained for lower and higher DDBBs respectively.

In order to further understand the results, a few more experiments were done reducing the registered subjects from 4 to 3 and 2. Here, it was shown that the identification rate increases while the number of subjects decreases. However, and exception occurs for DWT experiments and higher quality DDBB. This exception appears due to a lower performance of the system when particular subjects were modelled to be detected. In other words, the system appeared to have problems differencing between two specific subjects. When these two subjects were presented in the experiment, the system performs extremely badly, and therefore decreasing the mean identification rate.

ICA				
Principal Components	Lower Resolution		Higher Resolution	
	80 for training / 20 for test	60 for training / 40 for test	80 for training / 20 for test	60 for training / 40 for test
5	5 % ± 0	30,62 % ± 0	41 % ± 4,76	58,12 % ± 0
10	47,08 % ± 4	51,25 % ± 0	33,33 % ± 3,14	62,29 % ± 0.36
20	58,75 % ± 0	46,25 % ± 0	25,83 % ± 5,05	64,37 % ± 0
30	57,5 % ± 0	64,37 % ± 0	17,5 % ± 0	83,12 % ± 0
40	58,75 % ± 0	76,87 % ± 0	15 % ± 0	79,37 % ± 0
Subjects	Using the best combination		Using the best combination	
3 subjects	77,08 % ± 4,34		84,56 % ± 1,99	
2 subjects	80 % ± 15,41		89,68 % ± 13,2	

Table 6. Identification results for ICA experiments and lower and higher Resolution V-DDBB.

DWT				
Using 'bior 4.4'				
Principal Iterations / Coefficients	Lower Resolution		Higher Resolution	
	Lower Resolution		Higher Resolution	
	80 for training / 20 for test	60 for training / 40 for test	80 for training / 20 for test	60 for training / 40 for test
1 / 1764	85 % ± 0	75,62 % ± 0	97,5 % ± 0	76,25 % ± 0
2 / 638	76,25 % ± 0	66,25 % ± 0	98,75 % ± 0	75,62 % ± 0
3 / 285	45 % ± 0	40,625 % ± 0	93,75 % ± 0	71,87 % ± 0
Using 'haar'				
Iterations / Coefficients	Lower Resolution		Higher Resolution	
	80 for training / 20 for test	60 for training / 40 for test	80 for training / 20 for test	60 for training / 40 for test
1 / 1440	86,25 % ± 0	81,25 % ± 0	93,75 % ± 0	74,37 % ± 0
2 / 368	86,25 % ± 0	80,62 % ± 0	90 % ± 0	76,25 % ± 0
3 / 96	81,25 % ± 0	76,25 % ± 0	86,25 % ± 0	74,37 % ± 0
Subjects	Using the best combination		Using the best combination	
3 subjects	87,08 % ± 10,3		95,41 % ± 5,16	
2 subjects	93,12 % ± 6,88		98,12 % ± 3,75	

Table 7. Identification results for DWT experiments and lower and higher Resolution V-DDBB.

Finally, Testing Process, these are not the final identification rates of the whole system. The results coming out of the SVM section are processed in the TU. After applying the TU's double condition the identification rate is dramatically increased. The new probability of success is $P(X\geq2)$; two or more recognitions in 3 seconds.

$$P(X\geq2) = 1 - [P(X=0) + P(X=1)] \tag{23}$$

For our case, with 6 faces analyzed in 3 seconds, (4) can be expressed as:

$$P(X\geq2) = 1 - [(1 - p)6 + 6(1 - p)5p] \tag{24}$$

Here, 'p' is the probability of success in one attempt (percentages shown in Tables 6 and 7). Applying (5) to the results obtained with best combinations of ICA and DWT experiments with both lower and higher resolution DDBBs, really encouraging performances are obtained (see Table 8).

SYSTEM'S PERFORMANCE	
Lower Resolution	
DWT 'haar': 1 Iteration / 1440 Coefficients 80 for training / 20 for test	99,97 %
ICA: 40 Principal Components 60 for training / 40 for test	99,68 %
Higher Resolution	
DWT 'bior 4.4': 2 Iterations / 838 Coefficients 80 for training / 20 for test	100 %
ICA: 30 Principal Components 60 for training / 40 for test	99,93 %

Table 8. Identification results for the whole system using best configuration.

6. Discussions and conclusions

A number of experiments have been done using different databases and configurations. In general terms, the results show very high standard deviations. This fact points out that results are highly dependent on the set of samples used for training and test. However, this may not be due system's instability, but due the experimental procedure. This refers to the fact that the same number of samples has been used to build up both positive and negative classes in an unbalance problem, where different classes need different number of training samples.

For experiments based on image databases, we have obtained a classification system, which can be used for arbitrary illumination conditions. We have searched a classifier with a good efficacy for fixed and dynamic illumination conditions, though the parameterization has to be different for reaching a better success rate. In particular, ICA has been used for a database with arbitrary illumination conditions, and DWT- Bior has been used for a database with fixed illumination conditions. The results are upper 98.9% for the studied cases.

For experiments based on videos databases, we have used the DWT and SVM classifier; we have created a system implemented in Matlab which is able to detect a subject in a video sequence with an accuracy of 100%. The major errors detected here were maximum delay around 2 seconds. Even though for our V-DDBB subjects were always detected, more tests with a wider DDBB is needed before came to a conclusion in this aspect, and ensure that the system performance in perfect. For example, as it has been said before, the FEB does not perform perfectly, which means that the system does not always have 6 pictures of the subject's face in 3 seconds. Obviously, this plays against system's accuracy rate.

However, computing time has turned up to be a handicap of the resulted system. With and actual processing time of 5 times the length of the video, more research is needed in order to speed it up. One solution could be sharp the Face Extractor Block in order to increase its accuracy, reducing the number of false face founds. Tuning Training and Testing Blocks are always an interesting point, and along with the Face Extractor Block improvement the number of analyzed faces per second could be reduced, and therefore the whole processing time will be reduced to without decrease the system's accuracy.

Finally, processing time will be shorted again once the Matlab code has been translated to C++ and run as a normal application.

7. Acknowledgment

This work has been partially supported by "Cátedra Telefónica ULPGC 2009-10", and by the Spanish Government under funds from MCINN, with the research project reference "TEC2009-14123-C04-01".

8. References

Akbari, R.; Bahaghighat, M.K. & Mohammadi, J. (2010). "Legendre moments for face identification based on single image per person", *2nd International Conference on Signal Processing Systems (ICSPS)*, Vo. 1, pp. 248-252, 5-7 July 2010

AT&T Laboratories Cambridge (January 2002). ORL Face Database, 14-03-2011, Available from http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Azam, M.; Anjum, M.A. & Javed, M.Y. (2010). "Discrete cosine transform (DCT) based face recognition in hexagonal images," *2nd International Conference on Computer and Automation Engineering* (ICCAE), Vo. 2, pp. 474-479, 26-28 Feb. 2010

Banu, R.V. & Nagaveni, N. (2009), "Preservation of Data Privacy Using PCA Based Transformation", *International Conference on Advances in Recent Technologies in Communication and Computing*, ARTCom '09., pp. 439 – 443

Biometric International Group, (January 2011), 14.03.2011, Available from http://www.ibgweb.com/

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press.

Bouchaffra, D. (2010). "Conformation-Based Hidden Markov Models: Application to Human Face Identification", *IEEE Transactions on Neural Networks*, vol.21, no.4, (April 2010), pp.595-608

Bradski G., (March 2011), OpenCVWiki, 18-03-2011, Available from http://opencv.willowgarage.com

Cevikalp, H., Neamtu, M., Wilkes, M. & Barkana, A. (2005). "Discriminative common vectors for face recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, No.1 pp. 4-13, (Jan. 2005)

Chaari, A.; Lelandais, S. & Ben Ahmed, M. (2009). "A Pruning Approach Improving Face Identification Systems," *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009*, pp.85-90, 2-4 Sept. 2009

Chaudhari, S.T. & Kale, A. (2010), "Face Normalization: Enhancing Face Recognition," *3rd International Conference on Emerging Trends in Engineering and Technology* (ICETET), pp. 520-525, 19-21 Nov. 2010

Chen-Chung, Liu & Shiuan-You, Chin, (2010). "A face identification algorithm using support vector machine based on binary two dimensional principal component analysis," *International Conference on Frontier Computing. Theory, Technologies and Applications*, (2010 IET), pp. 193-198, 4-6 Aug. 2010

Chennamma, H.R.; Rangarajan, Lalitha & Veerabhadrappa (2010). "Face Identification from Manipulated Facial Images Using SIFT," *3rd International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pp. 192-195, 19-21 Nov. 2010

Chi Ho, Chan; Kittler, J. (2005). "Sparse representation of (Multiscale) histograms for face recognition robust to registration and illumination problems," *17th IEEE International Conference on Image Processing* (ICIP), pp. 2441-2444, 26-29 Sept. 2010

Cichocki, A., Amari, S. I.. Adaptive Blind Signal and Image Processing. John Wiley & Sons, 2002. In oroceedings EUROSPEECH2001, 2603 – 2606, 2001.

Cristianini, N. & Shawe-Taylor, J. (2000). *An introduction to support vector machines*, Cambridge University Press, 2000

Face Recognition Homepage, (March 2011). Yale Database, 14-03-2011, Available from http://www.face-rec.org/databases/

González R. & Woods R. (2002). *Digital Image Processing*, Ed. Prentice Hall

Hamel, L. H. (2009). *Knowledge Discovery with Support Vector Machines*, Wiley & Sons

Huang J.; Yuen, P.C.; Wen-Sheng Chen & Lai, J.H. (2003). "Component-based LDA method for face recognition with one training sample ", *IEEE International Workshop on Analysis and Modeling, of Faces and Gestures*, pp. 120 – 126

Hyvärinen, A. ; Karhunen, J. & Oja, E. (2001). I*ndependent Component Analysis*. John Wiley & Sons

Joachims, T. (1999). "Making large-Scale SVM Learning Practica"l. *Advances in Kernel Methods* - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 25-07-2007, Available from : http://svmlight.joachims.org/

Khairul, H.A. & Osamu, O. (2009). "Infrared-based face identification system via Multi-Thermal moment invariants distribution", *3rd International Conference on Signals, Circuits and Systems (SCS)*, pp.1-5, 6-8 Nov. 2009

Kisku, D.R.; Rattani, A.; Grosso, E. & Tistarelli, M. (2007). "Face Identification by SIFT-based Complete Graph Topology," *IEEE Workshop on Automatic Identification Advanced Technologies*, pp.63-68, 7-8 June 2007

Kurutach, Werasak; Fooprateepsiri, Rerkchai & Phoomvuthisarn, Suronapee (2010). "A Highly Robust Approach Face Recognition Using Hausdorff-Trace Transformation ", Neural Information Processing. Models and Applications, *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg ; Vo. 6444, pp. 549-556, 2010

Liun, R. & He, K. (2008). "Video Face Recognition on Invariable Moment". *International Conference on Embedded Software and Systems Symposia '08*, pp : 502 – 506, 29-31 July 2008

MathWorks, (January 2011), MATLAB, 14.03.2011, Available from http://www.mathworks.com/products/matlab/

Murata, N. ; Ikeda, S. & Ziehe, A. (2001). "An approach to blind source separation based on temporal structure of speech signals". *Neurocomputational*, Vo. 41, pp. 1–24

Parra, L.(2002). *Tutorial on Blind Source Separation and Independent Component Analysis*. Adaptive Image & Signal Processing Group, Sarnoff Corporation

Sang-Il, Choi ; Chong-Ho, Choi & Nojun Kwak, "Face recognition based on 2D images under illumination and pose variations", *Pattern Recognition Letters*, Vo. 32, No. 4, pp. 561-571, (March 2011)

Saruwatari, H. ; Kawamura T. & Shikano, K. (2001). "Blind Source Separation for speech based on fast convergence algorithm with ICA and beamforming. 7th European Conference on Speech Communication and Technology ", In EUROSPEECH-2001, pp. 2603-2606, September 3-7, 2001

Saruwatari, H. ; Sawai, K. ; Lee A. ; Kawamura T. ; Sakata M. & Shikano, K. (2003). "Speech enhancement and recognition in car environment using Blind Source Separation and subband elimination processing ", *International Workshop on Independent Component Analysis and Signal Separation*, pp. 367 – 37

Shreve, M.; Manohar, V.; Goldgof, D. & Sarkar, S. (2010), "Face recognition under camouflage and adverse illumination", *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pp.1-6, 27-29 Sept. 2010

Smith, L. I. (2006), "A tutorial on Principal Component Analysis", 01-12-2006, Available from :
http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf,

Travieso, C.M. ; Botella, P. ; Alonso, J.B. & Ferrer M.A. (2009). "Discriminative common vector for face identification". 4*3rd Annual 2009 International Carnahan Conference on Security Technology, 2009*, pp.134-138, 5-8 Oct. 2009.

Wen-Sheng, Chu ; Ju-Chin, Chen & Jenn-Jier, James Lien (2006). "Kernel discriminant transformation for image set-based face recognition" , *Pattern Recognition*, In Press, Corrected Proof, Available online 16 February 2011

Xiong. G. (January 2005), The "localnormalize", 14.03.2011, Available from http://www.mathworks.com/matlabcentral/fileexchange/8303-local-normalization

Yong, Xu ; Aini, Zhong ; Jian, Yang & Zhang, David (2010). "LPP solution schemes for use with face recognition", *Pattern Recognition*, Vo.43, No. 12, (December 2010), pp 4165-4176

Yossi, Zana & Roberto M. Cesar, Jr. (2006). "Face recognition based on polar frequency features." *ACM Trans. Appl. Percept.* 3, 1 (January 2006), pp. 62-82

Zhao Lihong; Zhang Cheng; Zhang Xili; Song Ying; Zhu Yushi (2009), "Face Recognition Based on Image Transformation", WRI Global Congress on *Intelligent Systems*, 2009. GCIS '09., vol.4, pp. 418-421, 19-21 May 2009

Zhou, H. & Sadka, A. H. (2010). "Combining Perceptual Features With Diffusion Distance for Face Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Review*, No. 99, (junio 2010) pp. 1-12