Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels
UNIVERSITAT POLITÈCNICA DE CATALUNYA

FEATURE EXTRACTION TECHNIQUES FOR FACE IDENTIFICATION

TITLE: Feature extraction techniques for face identification

MASTER DEGREE: Master in Science in Telecommunication Engineering & Management

AUTHOR: Javier Granado Font

DIRECTOR: Josep Ramon Morros Rubió

DATE: July, 2012

Title: Feature extraction for face identification

Author: Javier Granado Font

Director: Josep Ramon Morros Rubió

Date: July 2012

Overview

For face recognition, it is very important determining which features of the faces will be used in the classification process.

The identification based on appearance uses the pixels of the corresponding image to extract the features. Using the pixels directly is not very efficient due to the high dimensionality of the resulting features which results in a poor discriminative capability between different persons and in increased computational complexity.

Implementing any kind of data transform could be a good strategy for reducing the dimensionality of the data and increasing the discriminator capability.

Using PCA or DCT transforms it is possible to implement systems with a good rate of recognition if the number of recognizable persons is low.

In this project, it has been investigated others features extraction techniques, especially the ones based on Local Binary Patterns.

ÍNDEX

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. Context

Automatic face analysis which includes for example face detection, face recognition and facial expression recognition has become a very active topic in computer vision research. A key issue in face analysis is finding efficient descriptors. Different methods such as Principal Component Analysis (PCA) [4], Discrete Cosines Transform (DCT) [11] have been widely studied and also local descriptors as Local Binary Pattern (LBP) have gained attention due to their robustness to challenges such as pose and illumination changes.

A recognition system can have two different approaches: feature-based and appearance-based. Feature-based methods look for relevant features in faces such as eyes, nose or mouth, being able to determine the identity of a given individual by comparing feature distances and positions with other candidates. However, in appearance-based methods, faces are considered to be groups of pixels that respond to a certain pattern which can be extracted and used for comparing between faces. Using the pixels directly is not very efficient due to the high dimensionality of the resulting features which results in a poor discriminative capability between different persons and in increased computational complexity.

Moreover, a recognition system can be designed in two different ways, as identification or as verification system. The identification one searches among all the created models which are the most similar with the entered image. Also, it can find unknown persons. However, a verification system gives you the information about if the entered image is the person that the user thinks or not. In this project, a feature-based identification system has been implemented.

In face recognition, determining which features of the faces will be used in the classification process is very important For extracting these features two different techniques have been used in this project: Discrete Cosines Transform (DCT) and Local Binary Patterns (LBP). Also different approaches combining these two techniques have been proved.

The Local Binary Patterns (LBP) operator is one of the best performing texture descriptors. It has proven to be highly discriminative and it has great advantages, namely its invariance to monotonic gray level changes and computational efficiency. The idea of using LBP for face description is motivated by the fact that faces can be seen as a composition of micro-patterns which are well described by such operator.

The Discrete Cosine Transform (DCT) is used as a means of feature extraction for later classifying faces. The DCT is computed for an input image containing a face, and only a small subset of the coefficients is maintained as a feature vector. This feature vector may be conceived of as a representing point in a high-dimensional “face” space. Classification is usually based on a simple Euclidean distance measure.

For both techniques, the image has been divided into a set of small rectangular regions. If DCT has been used, some coefficients for each region have been extracted and then concatenated for creating the feature vector. If LBP has been used, histograms have been extracted for each region and then concatenated into a single, spatially and enhanced feature vector. This feature vector can be formed using different resolution levels, that is, diving the image into small, intermediate and high regions for creating the feature vector concatenating the obtained features of each region. With these two different approaches, it has been possible to determine which information is more relevant, the local or the global one.

Once the feature vectors are created, next step is classification. In this project, a system has been developed in order to classify each new image with its corresponding model. A model has to be understood as a set of feature vectors formed for each individual. Each feature vector has been created from different images of the same individual. So, once the models of all individuals have been created and a new image enters to the system, the classification step is carried out. For doing that, several classifiers can be used. The one used has been “k Nearest Neighbors” (kNN). kNN is a method for classifying objects based on closest training examples in the feature space. Hence, when a new image (called test image) enters, the system extracts its relevant features for creating its feature vector and then obtain which is the model closer than the test image using kNN.

Thanks to good results obtained with LBP technique, other approaches combing both techniques, LBP and DCT, have been implemented. In the majority of the situations LBP presents better results than DCT. However, the reason of combining both techniques is to reduce dimensionality. LBP offers very good results but depending on the application, LBP can have a high computational cost. Hence, DCT can help to reduce this due to its low dimensionality.
Another reason of using both techniques is that it has been supposed that DCT can have better global information of the image than the one obtained with LBP. So, in this approach, the local information will be obtained through LBP and the global one with DCT.

Finally, a comparison between all approaches has been done taking into account aspects as processing time and rate of right results.

1.2. Structure of the dissertation

This project has been divided in six different parts.

In chapter one, the aim of the project has been introduced. In chapter two, the state of the art has been presented explaining the methods and techniques used in this project and others which has not been used but they are important because they belong to the literature. In chapter three, It has been explained the different necessary steps for implementing a feature-based face identification system and in chapter four, the implementation carried out in this project is explained differentiating what has been developed in each step of the system. In chapter five, all the experimental results and a comparison between different approaches have been explained. Finally, in chapter six, the conclusions are presented.

CHAPTER 2. BASIC TECHNIQUES AND STATE OF THE ART

In this chapter, it will be explained the basic techniques used in this project taking into account all the elements which form part of it.

A face recognition system can be understood as a set of processes which are applied to a facial image to recognize its identity in an automatic way. In order to achieve it, the following processes are necessary:

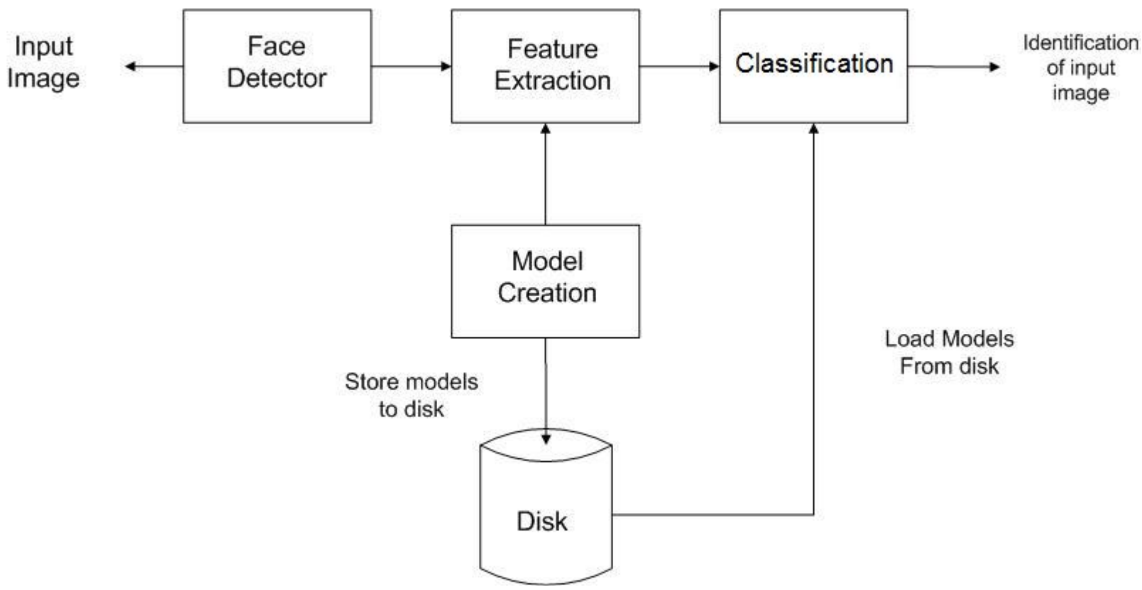

Figure 2.1- ***Block diagram representing the performance of a face recognition system***

Firstly, the face detection is done. In this project, this block has not been developed so it has used the face detector provided by the library OpenCV [23] which is based on the Viola-Jones algorithm.

Once the faces are extracted from the image, the next step is extracting the best features to create a feature vector. A model of each person has to be formed. A model is a set of feature vectors related to a specific person. This block is the one which has been studied in this project so it will have a specific attention in the following sections.

The third block is used to carry out the classification process. Given a test image, this block compares its features with the ones provided by all the existing models and it decides which model is the most closer than this image.

It is important to know that there are two different methods for building a face identification system: feature-based and appearance-based. This project will be focus on building a feature-based face identification system. Feature-based methods look for relevant features in faces, such as eyes, nose or mouth, and the identity of a given individual is found by comparing feature distances and positions with other candidates. The process of extracting face features, comparing them with other faces and finding the closest one might seem apparently simple but it can be difficult due to several problems can occurred. For example, the system can have problems to find these facial features in very small images. In these cases, no recognition can be done.

In appearance-based methods, faces are considered to be groups of pixels that respond to a certain pattern which can be extracted and used for comparing between faces. In the following section, all the blocks will be explained with more details.

2.1. Face detection

A face detection system extracts all the faces which are in an image. The face detector used in this project is the one provided by the library openCV which is based on the algorithm developed by Viola-Jones.

The Viola–Jones object detection algorithm [21] provides competitive object detection rates in real-time. Viola and Jones introduced a face detector based on cascades of very simple and aggressive classifiers which look for very specific face features, discarding backgrounds and other objects in the image.

This technique uses Adaptative Boosting (also known as AdaBoost) which is a machine learning algorithm and inherits the idea of weak Haar classifiers from Freund and Schapire's first general approach in [26]. The main idea is that, if well chosen, the union of weak classifiers leads to a very powerful classifier.

So, the first classifiers discard very strong features as backgrounds and shape of a face while the other classifiers scan the image for eyes and nose position, auto-scaling their scanning area from a minimum to a maximum face size. Scaling and strictness are adaptable parameters of the face detector and should be carefully used because the overall face identification system could underperform.

Cascade of classifiers are obtained after a very broad "face/no face" machine learning process. These face detectors also output a joint confidence parameter for all the involved weak classifiers that could prevent the recognition system from classifying wrong faces.

This implementation will work with a face detector in OpenCV, using frontal and left profile faces cascades. The face detection module will perform as follows:

1. The detector scans the image for frontal faces from minimum to maximum sizes. The implemented can bear one individual at a time.
2. If no frontal face is found, the detector scans for left profile faces form minimum to maximum sizes.
3. If no left profile face is found, the detector vertically flips the input image and scans for right profile faces from minimum to maximum sizes
4. Finally, if the detector has found a face, its coordinates will be outputted as a bounding box. IN other case, an empty bounding box is sent.

2.2. Face Features Extraction

Feature extraction is a process that consists on finding a set quantitative features that better define the input data in order to accomplish a goal, in this case, the classification of the faces into set of predefined classes.

Classification of this kind of data may be very costly but, fortunately, not all the features are equally important for classification purposes. Dimensionality reduction refers to the process of data simplification that allows keeping a small subset of features or coefficients. Note that, choosing a dimensionality reduction technique and the right number of kept coefficients is not trivial, since performance and results depend on it.

Many feature extraction and dimensionality reduction techniques can be found in the literature as for example PCA [4], LDA[27], DCT or LBP. The ones used in this project have been Discrete Cosine Transform [11], Local Binary Patterns [1] and different combination of both. However, thanks to the modular structure of the system, any other feature extraction technique could be used just doing few modifications.

All of these techniques follow the same process:

- Face images are resized, cropped or masked.
- Images are projected using different transforms.
- Some coefficients from the transformed images are selected for the features vectors.

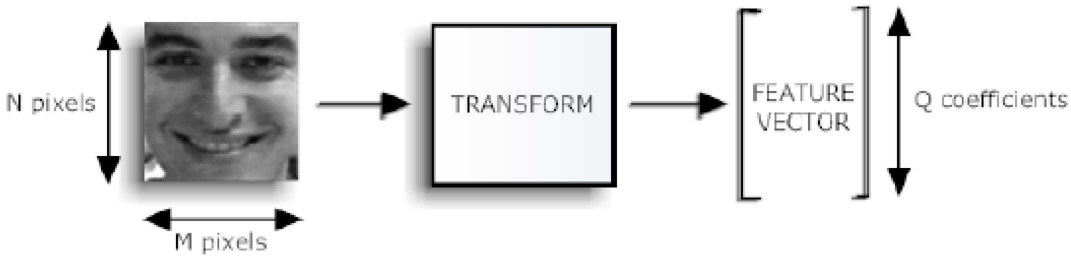

Figure 2.2- Representation of the transformation process

2.2.1. Principal Component Analysis

The Principal Component Analysis (PCA) [4] is one of the most successful techniques that have been used in image recognition and compression. The purpose of PCA is to reduce the large dimensionality of the data space into a smaller dimensionality of feature space. The PCA can be used for doing prediction, removing redundancy, extracting features, etc. So, it has been a good technique for face recognition.

The main idea of using PCA for face recognition is to express the large 1-D vector of pixels constructed from 2-D facial image into the compact principal components of the feature space. This is called eigenspace projection.

Eigenspace is calculated by identifying the eigenvectors of the covariance matrix derived from a set of facial images (vectors).

PCA is a technique which requires a previous training. In the training phase, feature vectors have to be extracted for each image. In order to extract PCA features, the image has to be converted into a pixel vector. Let suppose that p images will be used and each image has a pixel resolution of MxN (M rows, N columns). Each of the M rows will be concatenated into a single vector called v. The dimension of this vector will be $k = MxN$.

In order to apply PCA to the training set, a training data matrix A has to be formed. This matrix will contain the same number of rows as training vectors. Thus, the dimensionality of A will be $p * k$. First, the covariance matrix of A has to be calculated and it will be called C_A. Then, the eigenvalues and their corresponding eigenvectors of C_A should be computed. There will be k eigenvalues and eigenvector pairs where each eigenvector is of size k. Then, the eigenvalues has to be sorted in decreasing order and the biggest d eigenvalues and eigenvector pairs have to be selected. Another transformation matrix B has to be formed by simply putting the selected eigenvectors as columns. This matrix will be used to compute features vectors with small dimensionality called W. These vectors are from the original ones. The computation of W is simply by:

$$W = B^T * v^T \tag{0}$$

Note that, each column of B corresponds to an eigenvector which is the length k. This is equal to $M * N$ which is the resolution of input images. Thus, each eigenvector can be converted to an image by reversing the concatenation operation. These converted eigenvector images are called eigenfaces since they are similar to human faces.

Figure 2.3- Ten different eigenfaces extract of PCA

Once all the W vectors are obtained using the d eigenvectors, the image of a person can be reconstructed. If all the k eigenvectors are used instead of d when forming B, the reconstructed image will be the same as the original one. However, since one of the aim of using PCA is dimensionality reduction and d << k, the reconstructed image will be an approximation of the original one.

Figure 1.10 shows the reconstructed images of two persons using different number of eigenvectors. Note that, if more eigenvectors are used, the reconstructed image is more similar to the original one.

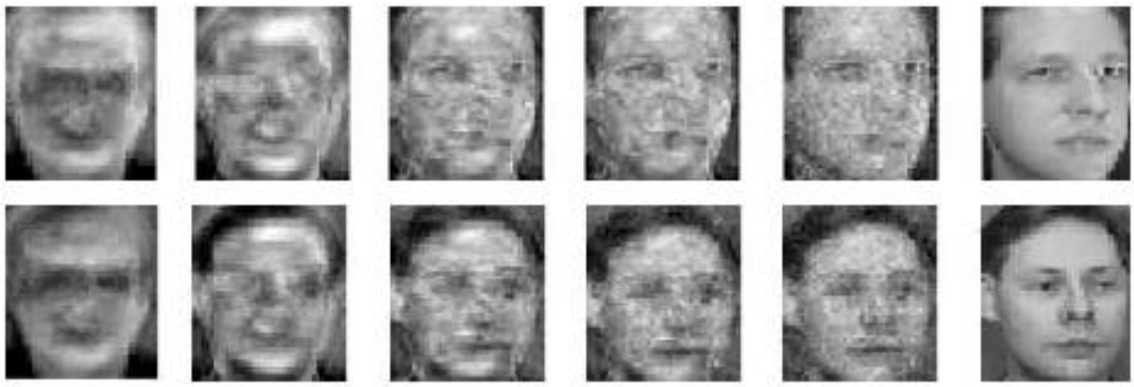

Figure 2.4- PCA reconstructions of the two subjects. Starting from the first row, the most important 5, 10, 100, 150, 700, and 1584 eigenvectors are used to reconstruct a face image. The resolution of the original face images is 44 x 36 = 1584.

2.2.2. Discrete Cosine Transform

In this project, different techniques have been used. One of them is based on the Discrete Cosine Transform (DCT) [11]. This transform is a particular case of the Fourier Transform. There are a lot of definitions of the DCT but the DCT-II is widely used in image processing and also in this project.

Given an NxN grayscale image, the two-dimensional DCT-II transform is defined as:

$$X(i,j) = \frac{2}{N} k_i k_j \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I(x,y) \cos\left(\frac{\pi(2x+1)i}{2N}\right) \cos\left(\frac{\pi(2y+1)j}{2N}\right) \qquad [1]$$

$$\boldsymbol{k_i, k_j} = \begin{cases} \dfrac{\mathbf{1}}{\sqrt{\mathbf{2}}}, & \boldsymbol{i, j = 0} \\ \mathbf{1}, & \boldsymbol{otherwise} \end{cases}$$

The resulting matrix contains NxN DCT coefficients which are related to one of the independent NxN DCT basis. Figure 2.5 shows the $8x8$ DCT invariant basis for an $8x8$ image.

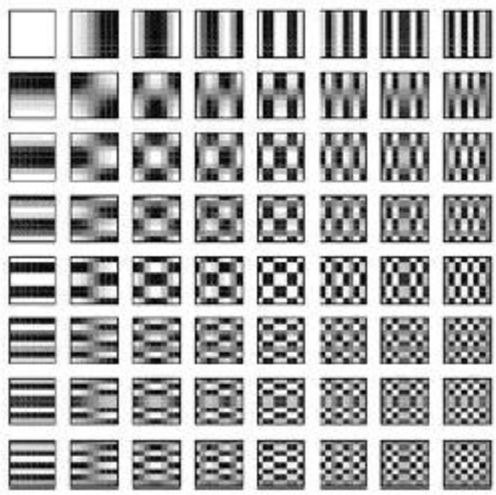

Figure 2.5- Independent basis related to DCT transform

The DCT transform has de the property of compacting the energy of the image so the most relevant coefficients for describing an image are located in the upper left. This can be seen in the following figure.

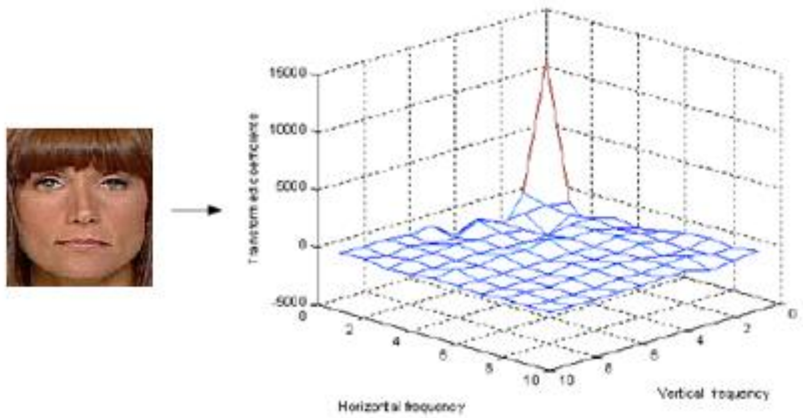

Figure 2.6- 3D graphic representing the concentration of energy in $X(0,0)$

An NxN image has a total of NxN transformed coefficients but a face recognition system can be designed using only a small subset. This property will reduce computational costs and memory usage in the face identification, classification and face model creation stages.

When all the DCT coefficients are computed, they are usually scanned in a zigzag order and it starts from $X(0,0)$.

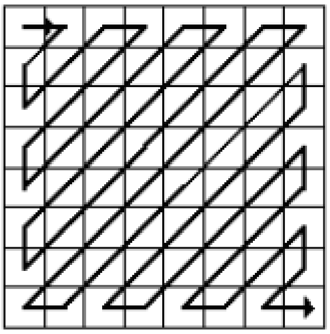

Figure 2.7- Zigzag scann for DCT coefficients

2.2.3. Local Binary Patterns

The Local Binary Pattern (LBP) [1] operator is a non-parametric 3x3 kernel which summarizes the local spatial structure of an image. It was firstly introduced by Ojala et al. [22] who showed the high discriminative power of this operator for texture classification.

At a given pixel position (x_c, y_c), LBP is defined as an ordered set of binary comparisons of pixel intensities between the center pixel and its eight surrounding pixels.

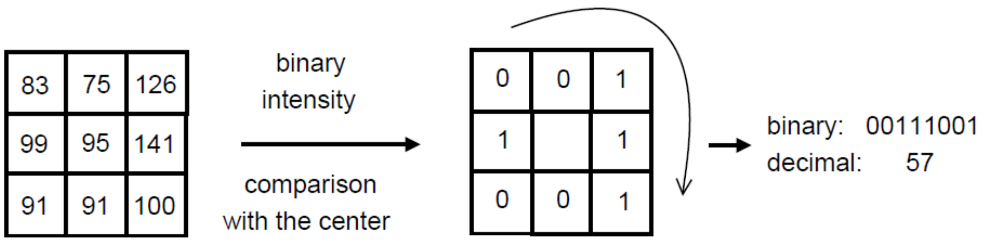

Figure 2.8- The basic LBP operator

The decimal form of the resulting 8-bit word (LBP code) can be expressed as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \qquad [2]$$

Where i_c corresponds to the grey value of the center pixel (x_c, y_c), i_n into the grey values of the 8 surrounding pixels, and function s(x) is defined as:

$$s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \qquad [3]$$

Note that each bit of the LBP code has the same significance level and two successive bit values may have a totally different meaning. By definition, the LBP operator is unaffected by any monotonic gray-scale transformation which preserves the pixel intensity order in a local neighborhood.

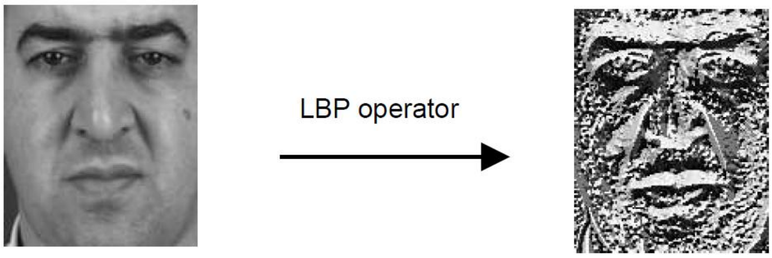

Figure 2.9- Left: original image. Right: LBP image

Later, Ojala et al. [22] made two extensions of their original LBP operator. The first extended the LBP operator to a circular neighborhood of different radius size. Their $LBP_{P,R}$ notation refers to P equally spaced pixels on a circle of radius R.

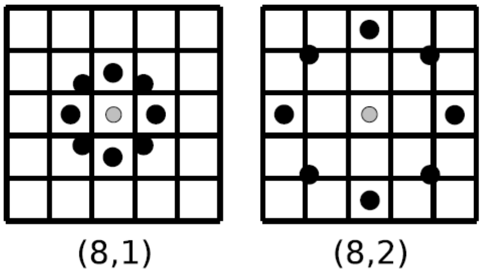

Figure 2.10- The circular (8,1) and (8,2) neighborhoods. The pixel values are bilinearly interpolated whenever the sampling points is not in the center of a pixel.

The second defined the so-called uniform patterns. So, an LBP is uniform if it contains at most two bitwise 0 to 1 or 1 to 0 transitions. “11111111”, “00000110” or “10000111” are considered uniform patterns too. Uniformity is an important concept in LBP methodology because it represents primitive structural information such as lines, edges or corners.

Due to its texture discriminative property and its very low computational cost, LBP is becoming very popular in pattern recognition. In fact, it is applied for instance to face detection, face recognition, image retrieval, motion detection or visual inspection.

2.3. Classification

Once all the necessary information is extracted and the models (for each person) have been created, it is necessary to have a system which is able to classify each new image with its corresponding model. This process is known as classification.

There are several classifiers as K Nearest Neighbors[17], Support Vector Machine[5,9], Parzen [28], etc. For carrying out a correct classification, normally more images are needed in the training stage.

The classifiers and distances used in this project will be explained in this chapter.

2.3.1. K Nearest Neighbors

The main idea of K Nearest Neighbor [17] classifier is based on a simple concept. Let P be a feature vector of length N and let $Q_i = \{q_{i0}, \cdots, q_{iN-1}\}$ be a matrix with N_i training vectors of the ith model on each row. The distances between P and the available feature vectors for every one of the N_q models are calculated and stored in a distance vector D.

The k smallest distance in D are selected, that is, the nearest feature vectors to P. Every selected neighbor will correspond to a face model or class and its identity will be registered as a vote. The winner class will be decided taking the most voted rule.

kNN is robust enough in two critical situations, despite its simplicity. Let us assume that a test vector belonging to a given class B must be classified and it is far enough from other existing classes as it is shown in figure 1.11

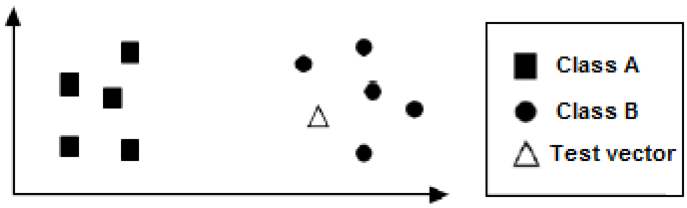

Figure 2.11- Image ilustrating two different classes and a test class which has to be classified.

Figure 1.12 and table 1.1 show step by step the results of the classification process.

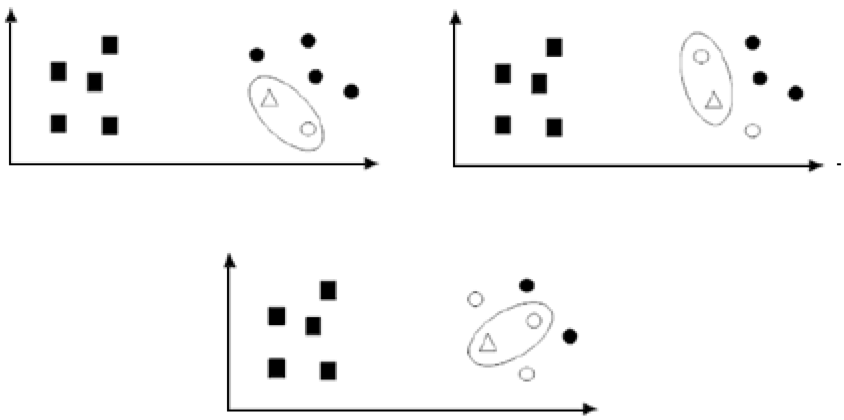

Figure 2.12- Figures showing the election of the kNN if its neighbours are k=1, k=2 and k=3

		Class A		Class B				
k-th N N	Decision	Winner	Accum	Winner	Accum	Match	Partial	Winner class
1	B	No	0	Yes	1	Yes	B	B
2	B	No	0	Yes	2	Yes	B	
3	B	No	0	yes	3	yes	B	

Table 2.1. Instant classification results and votes given for each class

As k increases, the threshold distance to a near neighbor also increases or remains as in the previous k. In this example, class B will continue to be selected up to $k = 5$ due to the proximity of the rest vector to the class B cluster. Note that, even choosing $k = 6$, as a parameter for kNN, the classified class will be class B, winning by 5 votes to 1.

Let us assume now that a second test vector belonging to a new class B must be classified but class A is very close to class B.

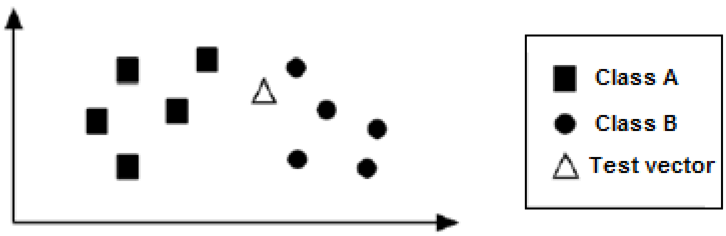

Figure 2.13- ***Image ilustrating two different classes and a test class which has to be classified***

Figure 1.14 and table 1.2 show step by step the results of the classification process.

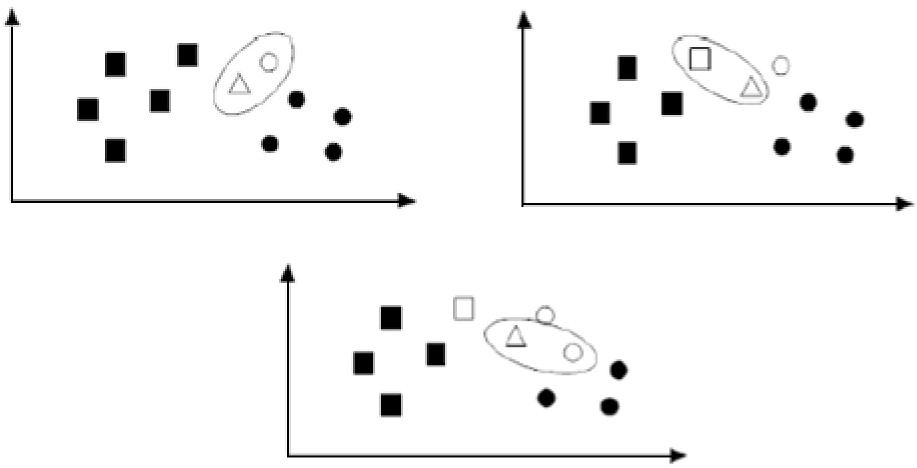

Figure 2.14- ***Figures showing the election of the kNN if its neighbours are k=1, k=2 and k=3***

		Class A		Class B				
k-th N	**Decision**	**Winner**	**Accum**	**Winner**	**Accum**	**Match**	**Partial**	**Winner class**
1	B	No	0	Yes	1	Yes	B	
2	A	Yes	1	No	1	No	A/B	B
3	B	No	1	yes	2	yes	B	

Table 2.2. ***Instant classification results and votes given for each class.***

If a test vector is not situated in a space region under the clear influence of any of the existing classes makes more difficult its classification. If $k = 2$ (see Figure 1.14) both classes has been selected once so the decision will be randomly taken. This kind of situations, were random decisions have to be taken, should be avoided by selecting only odd values of k.

In these two situations, the feature vector has been correctly classified but when having overlapped regions it is risky to classify using majority voting.

2.3.2. Support Vector Machine

Support Vector Machine (SVM) [5,9] is a binary classification method that finds the optimal linear decision surface based on the concept of structural risk minimization. The decision surface is a weighted combination of elements of the training set. These elements are called support vectors and characterize the boundary between the two classes.

Assuming linearly separable data, the goal of maximum margin classification is to separate the two classes by a hyperplane such the distance to the support vector is maximized. This hyperplane is called the Optimal Separating Hyperplane (OSH).

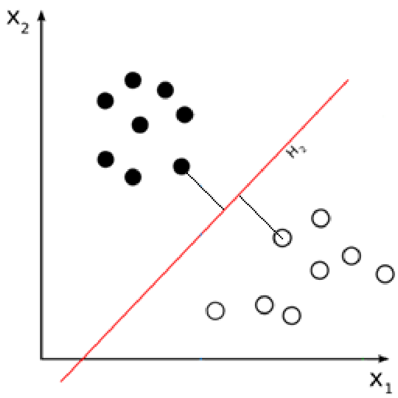

Figure 2.15- Figure representing a hyper-plane which separates two differentiated classes.

Classification of new data x is performed by computing the sign of the distance (d) from x to the hyperplane. Hence, the farther away a point is from the decision surface, for example the larger |d|, the more reliable the classification result. All of this can be extended to the case of nonlinear separating surfaces.

The original optimal hyperplane algorithm was a linear classifier. However, in 1992, Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick (originally proposed by Aizerman et al. [24]) to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space is high dimensional. Thus though the classifier is a hyperplane in the high-dimensional feature space, it may be nonlinear in the original input space.

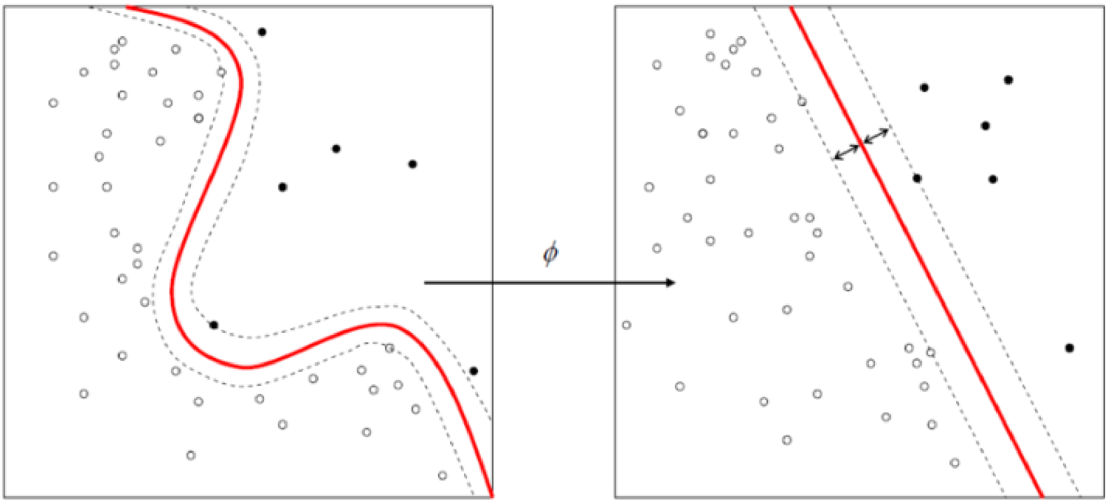

Figure 2.16- ***Kernel machines are used to compute a non-linearly seperable functions into a higher dimension linearly separable function.***

SVM can be extended to multi-class classification [5,9] using. There are two basic strategies for solving q-class problems with SVM:

i) One-vs-all approach: In this approach, q SVMs are trained. Each of the SVMs separates a single class from all remaining classes.
ii) Pair-wise approach. In this approach, $q(q-1)/2$ machines are trained. Each SVM separates a pair of classes. The pair-wise classifier is arranged in trees where each tree node represents an SVM.

Regarding the training effort, the **one-vs-all** approach is preferable since only q SVMs have to be trained compared to $q(q-1)/2$ SVMs in the par-wise approach. The run-time complexity of the two strategies is similar. The one-vs-all requires the evaluation of q, but the other requires the valuation of $(q-1)$ SVMs. However, since the number of classes in face recognition can be rather large, the one-vs-all approach is normally used because the number of SVMs is linear with the number of classes.

2.3.3. Distances metrics

In order to correctly classify the test image, it is necessary to compare the different features vectors. For doing that, three different distances have been used in this project.

Euclidean Distance

In mathematics, the Euclidean distance is the ordinary distance between two points that one would measure with a ruler.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad [4]$$

This distance is widely used in 2D metrics but can also be highly useful in n-dimensional vectors (for example, feature vectors) due to its low computational cost.

Manhattan Distance

The name of this distance is given by a simple geometrical interpretation: the distance between two points in a Manhattan-like structure is the minimum number of blocks that a person should cross.

$$d(x,y) = \sum_{i=1}^{n}|x_i - y_i| \quad [5]$$

Chi-square Statistics Distance

The Chi-square distance is used in histogram intersection because it is shown that it performs better than the Euclidean one.

$$x^2(S,M) = \sum_i \frac{(S_i - M_i)^2}{(S_i + M_i)} \quad [6]$$

where S, M are the histograms.

Unlike histogram values, since features that are extracted from DCT coefficients can also have negative coefficients, absolute values of these features are used while calculating the Chi-square statistics distance.

Weighted Chi-square Statistics Distance

When the image is divided into regions, it can be expected that some of the regions contain more useful information than other in terms on distinguishing between people. For example, eyes seem to be an important region in human face recognition.

To take advantage of this, a weight can be set for each region based on the importance of information it contains. For example, the weighted x^2 statistics becomes

$$x_w^2(S, M) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{(S_{i,j} + M_{i,j})} \qquad [7]$$

where S, M are the histograms and w are the weights.

2.4. Normalization

The normalization carried out in this project consists on a rotation when the face image is not completely straight.

2.4.1. Active Shape Models

The Active Shape Model algorithm (first introduced by Cootes et al. [10]) is a fast and robust method of matching a set of points controlled by a shape model to a new image.

Figure 2.17- Relevant point extracted from ASM

A shape model defines an allowable set of shapes. The shape models used in this project have a fixed number of points and a matrix formula which specifies the relationship between the points. The goal of the shape model is to convert the shape suggested by the models to an allowable face shape.

The allowable shape deformations are learnt from manually labelled training set to produce a linear shape model with the following form:

$$x = \bar{x} + P_S * b_S \qquad [8]$$

where $\bar{x}$ is the mean shape, P_S is a set of orthogonal modes of variation and b_S is a set of shape parameters.

The shape parameters for the model, along with parameters defining the global pose (the position, orientation and scale) define the position of the model points in an image. To improving the fit of the points to an image, first the region of the image around each current model point is examined in order to find the best nearby match and then the parameters to best fit the model to the new found points are updated. This is repeated until convergence, see 1.16.

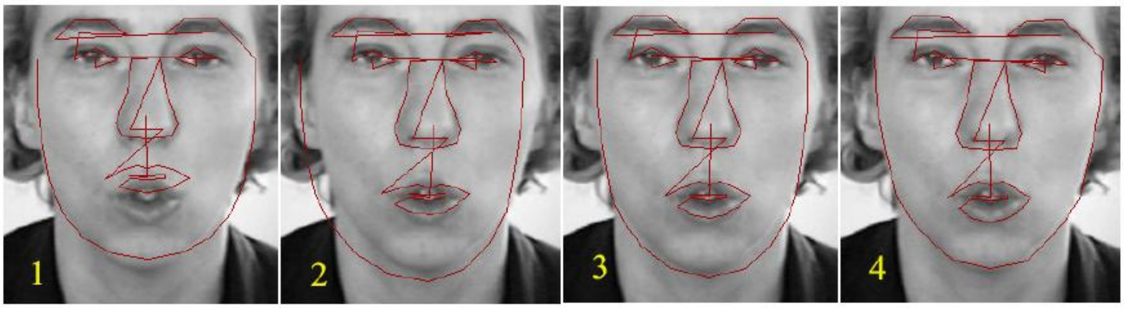

Figure 2.18- Different steps carried out by ASM until convergence.

Active Shape Models are used in order to find important facial points such eyes, nose or mouth. Normalization and rotation of the image can be done thanks to the location of these points.

2.4.2. Illumination compensation

Variable illumination is one of the most important problems in face recognition. The main reason is the fact that illumination is one of the most significant factors that alters the perception of faces [12,13]. Lighting conditions change largely between indoor and outdoor environments but also within indoor environments. Thus, a direct lighting source can produce strong shadows that accentuate or diminish certain facial features. Moreover, extreme lighting can produce too dark or too bright images which can disturb the recognition process.

There are several methods for carrying out illumination compensation. In this project no methods has been used due to the main used technique for extracting the features (LBP) is presented as a well discriminator of illumination changes. However, a brief description of three methods commonly used for illumination compensation is explained.

Histogram Equalization

The histogram equalization has been a widely used image processing technique for speech enhancement, which has the property of increasing the global contrast of an image while simultaneously compensating for the illumination conditions present at the image acquisition stage. It represents a useful pre-processing task which can provide an enhanced face image, improving the robustness of face recognition algorithms operating under different illumination conditions.

The histogram manipulation [15] which automatically minimizes the contrast in areas too light or too dark of an image, consist on a non-linear transformation that it considers the accumulative distribution of the original image to generate a resulting image whose histogram is approximately uniform.

On the ideal case, the contrast of an image would be optimized if all the 256 intensity levels were equally used. This is not possible due to the discrete nature of the digital data of the image. However, an approximation can be achieved by dispersing peaks in the histogram of the image, leaving intact the lower parts. This process is achieved through a transformation function that has a high inclination where the original histogram has a peak and a low inclination in the rest of the histogram.

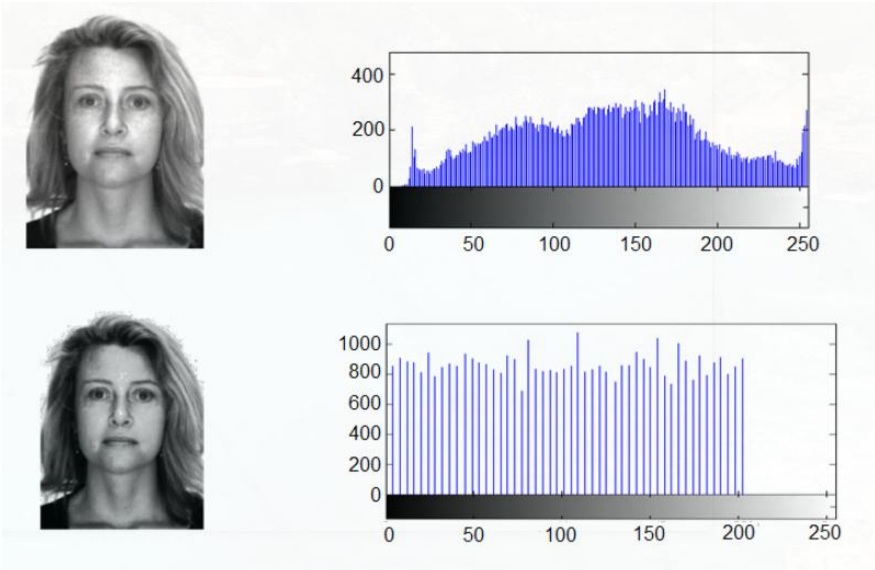

Figure 2.19- Difference between original histogram and equalizated histogram.

LogAbout

The logAbout method proposed by Liu et al. [13] can be implemented by applying a high pass filter followed by a logarithm transformation described by:

$$g(x,y) = a + \frac{\ln(f(x,y)+1)}{b*\ln(c)} \qquad [9]$$

where $f(x,y)$ is the original image, a, b and c are parameters which control the location and shape of the logarithm distribution.

Wavelets

In this method the image is decomposed in high (sub-bands LH, HL and HH of figure 3.20) and low frequencies (LL2). The, histogram equalization is applied on the approximation coefficients (low frequencies), and at the same time the detail are enhanced (high frequency) by multiplying each element of the detail coefficient matrix by a scale factor (>1) [13].

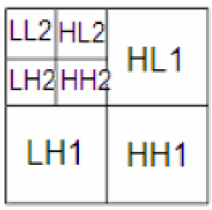

Figure 2.20- Multi-resolution structure

The image is reconstructed from its approximation coefficients and details coefficients in all the three directions by using the inverse wavelet transform, resulting the normalized image.

Homomorphic Filter

In this method the original image is split vertically in two halves, generating two sub-images from the original one (figure 2.21). Afterwards, the filter is applied in each sub-image and the resultant sub-images are combined to form the whole image. Then, the original image is dived horizontally and the same procedure us applied [13]. the two resultant images are grouped together in order to obtain the output image, given by:

$$I_{HMMOD}(x,y) = \frac{1}{2}\,[I_{HMV}(x,y) + \ 0.75 * \ I_{HMH}(x,y)] \qquad [10]$$

where $I_{HMV}(x,y)$ is the vertically divided image after the application of the homomorficfilter and $I_{HMH}(x,y)$ is the horizontally dived image.

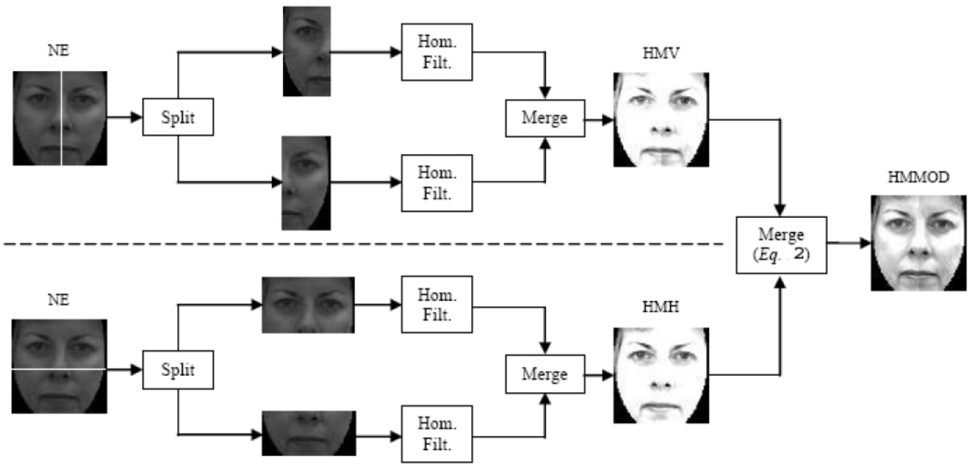

Figure 2.21- Homomorphic filter variation

CHAPTER 3. DESIGN OF THE SYSTEM

In this chapter, the system used for carrying out the recognition of people will be described. In general, a face recognition system is based on the following modules.

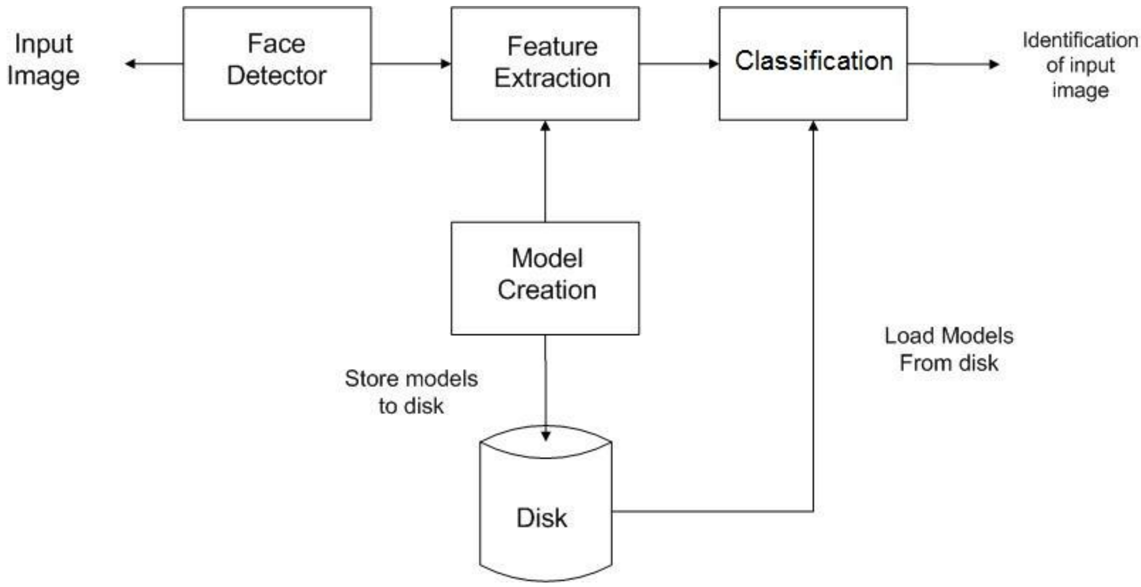

Figure 3.1- Block diagram representing the performance of a face recognition system

Image input: It is the image or set of image which are going to enter to the system for carry out the recognition. In our project, a txt file which the path of several images of one person is accepted too.

Face detection: Once the images are entered to the system, the faces have to be detected. In this project, the Viola-Jones [25] technique has been used because it is very fast and it achieves high detection rates. However, certain number of false positive images will be inevitably given.

Features extraction: Once the face is detected, it is necessary to extract some features which will be used for carry out the recognition of the person.

Model creation: in this module, the models of the people have been created. Each of these models represents a single individual in multiple poses or faces orientations, so the face identification module can look for the closest data vectors when a new face is detected. Note that, the system proposed in this project is a system with a predefined number of individuals and no new models will be created or deleted during its execution.

Face identification: This module is responsible to recognize the individual who has just entered in the system. It will extract the same features used in the moment of models creation, and it will compare these features with the models. Finally, the model closer than the entered image will be the selected as a winner.

3.1. Face detection

This block is necessary to carry out the detection of the faces. Note that, the system only works correctly with face images. So, for creating the models and for testing an image, face images will be used.

The performance of this block is the following. An image containing the person is entered. Then, the face detector used in the block detects the different faces of the image. Finally, it outs the coordinates of the face.

3.2. Training stage

As it has been described before, for creating a recognition system, it is necessary to create a model for each person. A model has to be understood as a set of features related to one person. For creating a model, several images of the same person are used. For each image, a feature vector is created and saved. So, a model is a set of feature vectors related to a person. Note that, a model also contains the name of the person.

In the creation of the models, different techniques have been tested. The most relevant have been, the one based on DCT and the one based on LBP. Then, different experiments have been carried out merging both techniques or proving different assumptions. This will be explained in chapter 4.

The following image represents the performance of this block:

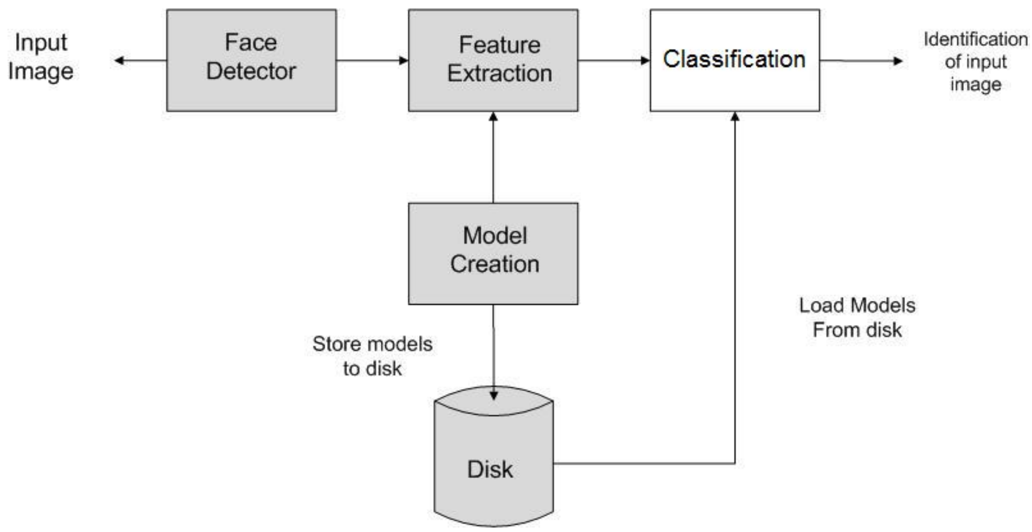

Figure 3.2- Functional block diagram representing the training stage in a face recognition system

It is necessary to enter a set of face image of the person who has to be examined and the name of the person. With these parameters, the block is able to create the model.
Before extracting the features, faces are detected and then, it is possible to extract the features for each image. Note that, depending on the technique

employed, the number of features can change. For example, if DCT is used, the number of features will be more o less 60 but if LBP is used, the number of features can be 3160 or even more.

Finally, the models for each person are created saving the name which has been entered as a parameter.

3.3. Classification Stage

After creating face models for every individual, the system is ready to start identifying individuals.

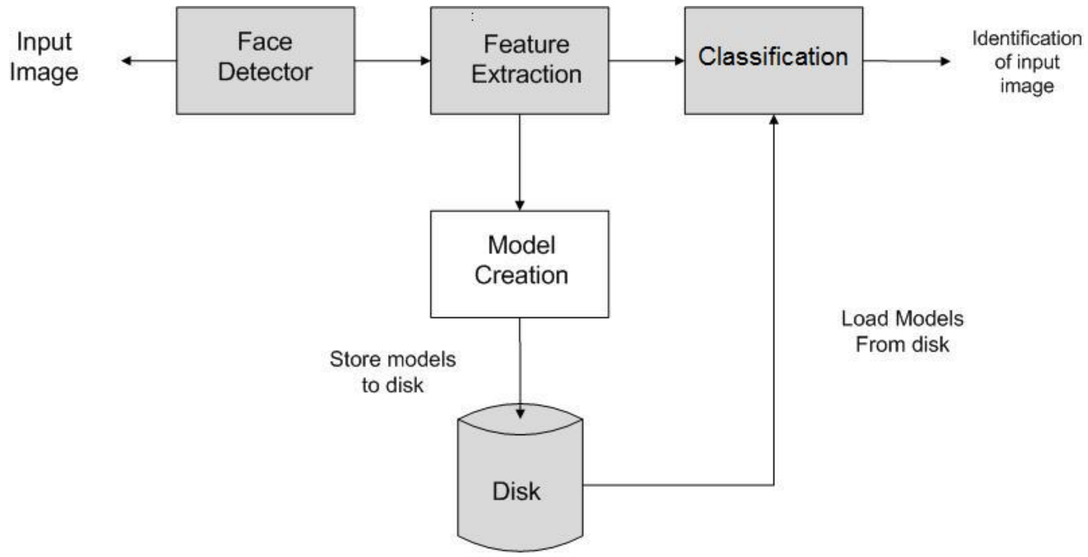

Figure 3.3- Functional block diagram representing the testing stage in a face recognition system

With no information given previously, the system must be able to detect the faces. So these faces can be classified by comparing their transformed coefficients with those stored in the existing face models.
So, once the face is found in the image, then the same feature extraction technique has to be carried out in order to create a feature vector called tested feature vector.

When the classifier, using for example K Nearest Neighbors, finds a feature vector in a face model that minimizes the distance to the tested feature vector, the identity of the chosen face model is displayed.

CHAPTER 4. TECHNIQUES FOR FACE FEATURES EXTRACTION

In this chapter, the main techniques for features extraction used in this project will be explained. Also, the different approaches carried out and the combination of the different techniques will be detailed.

The dissimilarity measures used are also very important, so it will be described the most representative ones.

4.1. Discrete Cosine Transform

The proposed feature extraction based on DCT method used in this project has two different approaches.

In the first one, the image is not dividing into blocks. The DCT is applied for the entire image and then some coefficients are kept in order to represent the entire image. Local information is not given in the final vector but it has a lower computational cost.

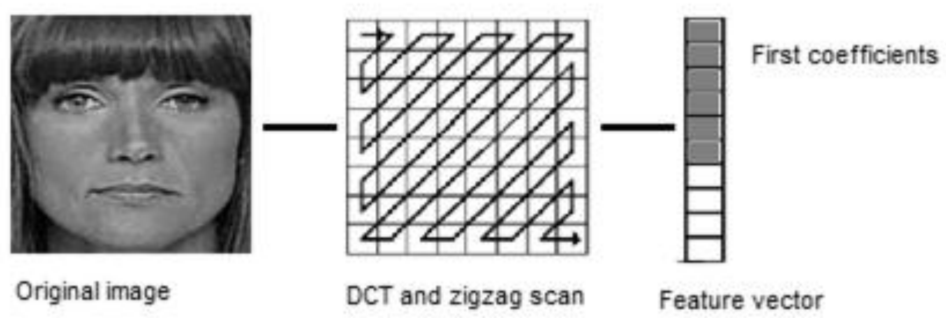

Figure 4.1- Feature extraction representation for a given image using DCT

The second one consists on dividing the image into LxL (single resolution) blocks and applying the DCT for each block. Note that, the blocks are not of size L, but the image is divided into LxL blocks. With this approach, local information is extracted of the face image and this provides a number of advantages. Firstly, the use of local features allows spatial information about the image to be retained. Secondly, for an appropriately selected block size, illumination affecting the block can be assumed quasi-stationary. Consequently, the resulting recognition system can be designed for improved performance in environments where variation in illumination occurs across images.

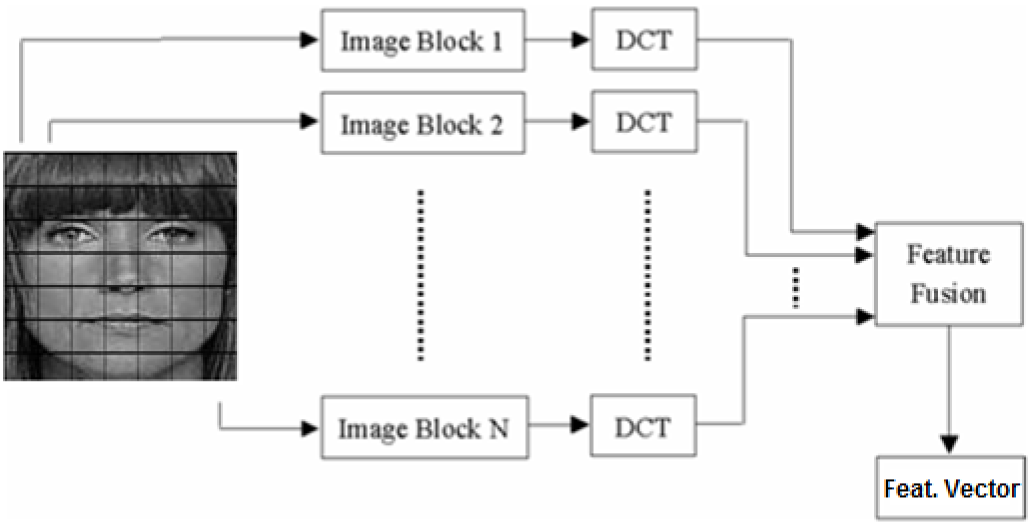

Figure 4.2- Feature extraction diagram using DCT block based

The DCT-based feature extraction for local appearance face recognition can be summarized as follows: A detected and normalized face image is divided into blocks of different size. Then, the DCT is applied on each block. The obtained DCT coefficients are ordered using zig-zag scanning. From the ordered coefficients, M of them are selected and then normalized resulting in an M-dimensional local feature vector. These extracted local features are then concatenated to represent the entire face image.

4.2. Local Binary Patterns

4.2.1. Introduction

In this project it has been tested a face recognition system based on LBP representation of the face. The procedure consists on using the texture descriptor to build several descriptions of the face and combining them into a global description. One of the reasons to follow this approach is it seems to be more robust in against variations in pose or illumination.

The individual sample image is divided into T non-overlapping local regions of same size and texture descriptors are extracted from each of the regions independently. These descriptors are the histograms calculated for each block. Finally, they have to be concatenated to form a global descriptor of the face.

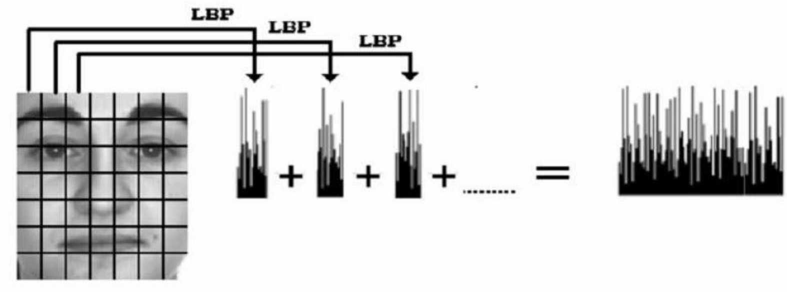

Figure 4. 3- Histogram extraction representation for face recognition using local binary patterns

The basic histogram can be extended into a spatially enhanced histogram which encodes both the appearance and the spatial relations of facial regions. As the m facial regions $R_0, R_1, \ldots, R_{m-1}$ have been determined, a histogram is computed independently within each of the m regions. The resulting m histograms are combined so the spatially enhanced histogram is created. The spatially enhanced histogram has size of $m\ x\ n$ where n is the length of a single LBP histogram.

4.2.2. Approaches

In this project, two different approaches have been used in the creation of the enhanced histogram: The “Single Resolution” and the ”Multi-Resolution”.

In the Single Resolution approach, the image is dived in LxL blocks. For each block, it has been calculated the LBP histograms and the enhanced histogram is created concatenating these histograms. In this approach only one level of resolution is given so the enhanced histogram has only local information of the image. The main advantage is that the computational cost is lower due to less LBP histograms have to be calculated. Hence, the length of the enhanced histogram is lower.

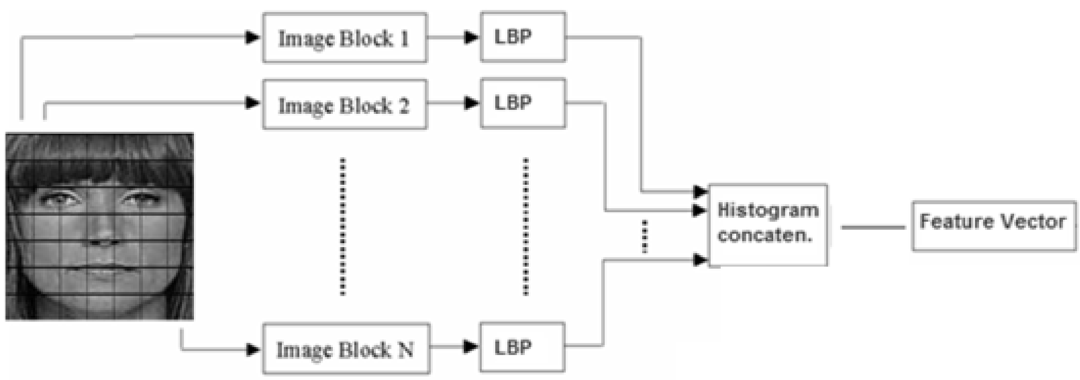

Figure 4.4- Block diagram representing the feature extraction performance using LBP in single resolution approach.

In the Multi-Resolution (MR) approach the image is first, divided in LxL blocks, then in $L_a x\, L_a$ blocks and finally in $L_b x\, L_b$ blocks, being L_a and L_b smaller number than L. For each block, a LBP histogram has to be calculated and the concatenation of all the histograms becomes the enhanced histogram. So, in the spatially enhanced histogram, it effectively has a description of the face on three different levels of locality: The LBP labels for the histogram contain information about the patters on a pixel-level (7x7 blocks used in this project), the labels are summed over a small region to produce information on a regional level (5x5 blocks used in this project) and the regional histograms are concatenated to build a global description of the face (3x3 blocks used in this project).

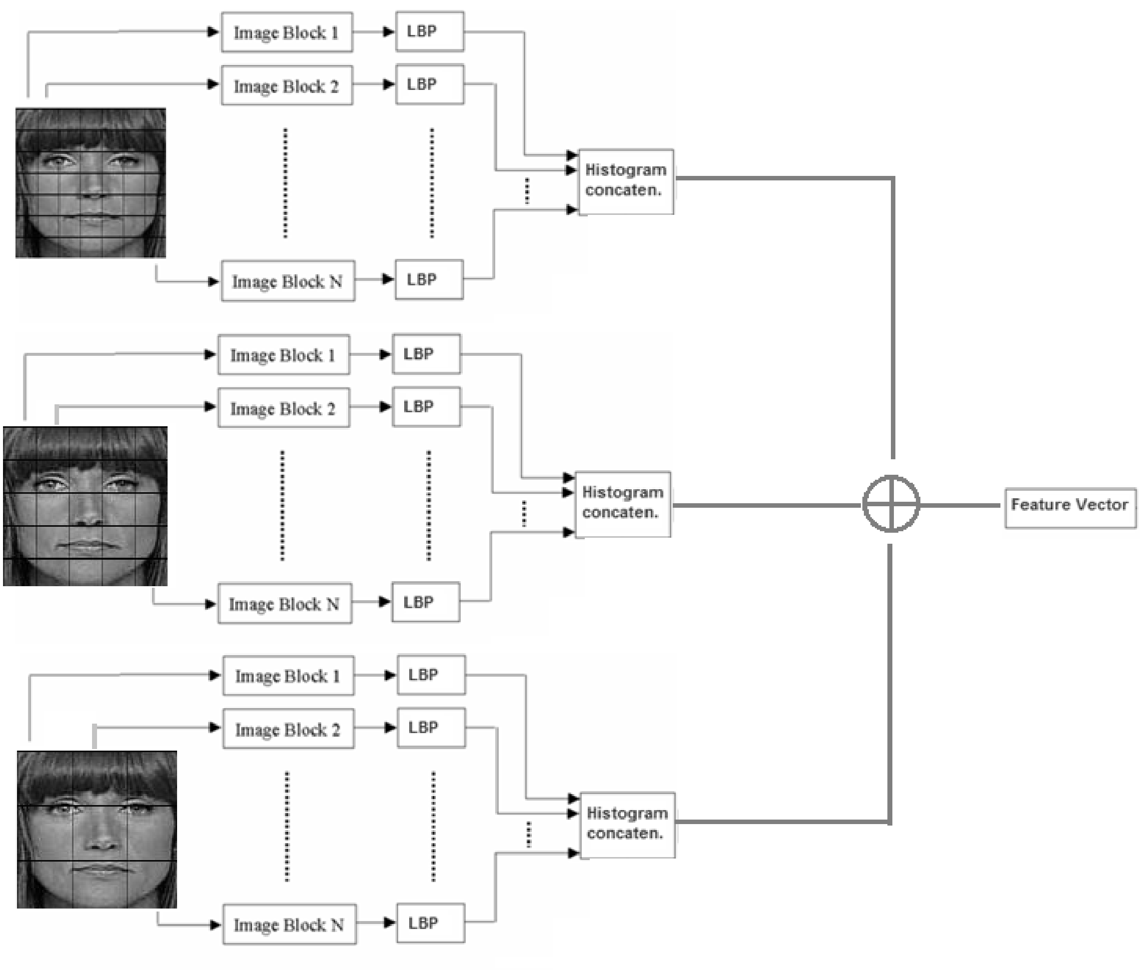

Figure 4.5- Block diagram representing the feature extraction performance using LBP in multiple resolution approach.

4.3. DCT over Local Binary Patterns

This combination of both techniques consists on firstly dividing the image into blocks and Local Binary Patterns are extracted for each block. As LBP histograms have a large length, the Discrete Cosine Transform will be applied in the created feature vector from the LBP. So the dimensionality of the resulting vector will be reduced in order to improve computationally costs.

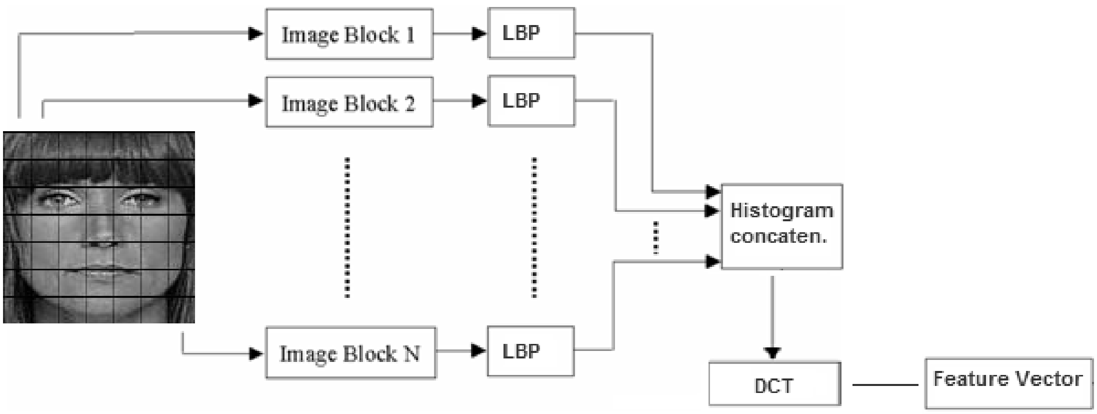

Figure 4.6- ***Block diagram representing the feature extraction performance using LBP and the dimensionality reduction using DCT..***

4.4. DCT and Local Binary Pattern

This combination of both techniques consists on firstly, Discrete Cosine Transform is conducted on an input image. As always, a few DCT coefficients on the left top corner are chosen as global features. At the same time, the image is divided into several blocks. Local Binary Pattern is conducted on each block and then LBP histograms are accepted as local features.

Finally, all the features are concatenated having both global and local features of the face image.

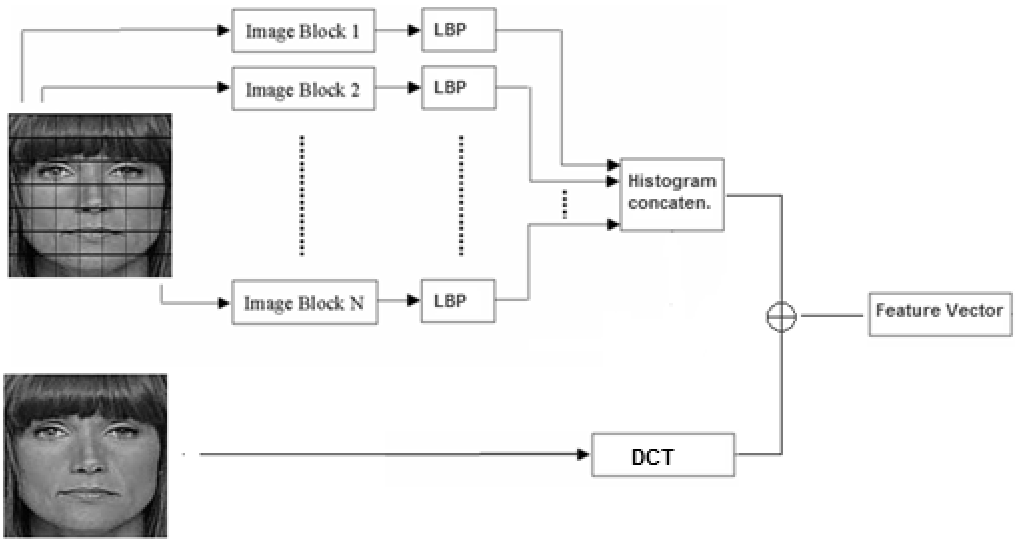

Figure 4.7- Block diagram representing the feature extraction performance using LBP and DCT techniques.

4.5. Weighted Chi-Square Distance

4.5.1. Introduction

The idea of a spatially enhanced histogram can be exploited further when defining the distance measure. A property of the proposed face descriptor method is that each element in the enhanced histogram corresponds to a certain small area of the face.

Based on psychophysical findings, which indicates that some facial features (such eyes, mouth or nose) play more important roles in human face recognition than others features [3], it can be expected that in this method some facial regions contribute more than others.

For this reason, one of the distances used in this project is Weighted Chi-square distance. With this distance, it is possible to give different weights to each region depending on its importance.

4.5.2. Obtaining weights

The calculation of the weights for each face region has been a goal in this project. In this point, it will be explained how it has been done.

A simple procedure was adopted to find the weights. In this procedure, a training set was classified using the same window and the same LBP parameters.
To find the weights of each block, the following steps have been carried out:

1. A face photo collection of an individual is collected.
2. For each face image, only the block which is being evaluated will be compared with the same block of all the training vectors.
3. The result obtained will be stored.
4. This procedure has to be done several times with different individuals.
5. Finally, the right results for each block will be counted and a probability measure will be calculated.

The results obtained have not improved the results obtained when weights have not been used. For this reason, it has been deduce that the technique used for finding the weights have not been right. In fact, the figure 4.8 does not show the expected weights if it is based on psychophysical findings where eyes are a very import region. Note that, the white colour represents the highest weight and the black, the lowest one.

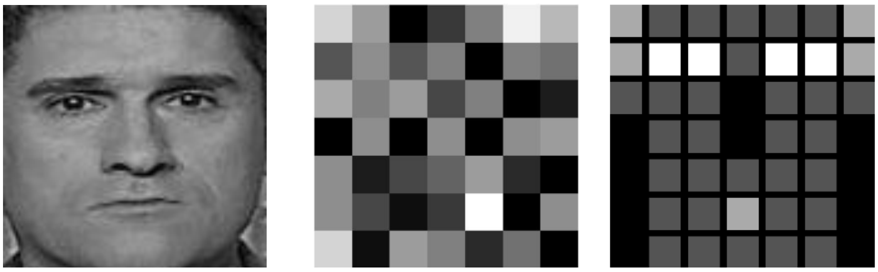

Figure 4.8- On the left, original image. In the center, the obtained weights for each block and on the right, the expected weights.

The weights has not been the ones expected. As it can be seen in figure 4.8, the most relevant regions of a face have not the higher weights and the weights related to the background of the image should be lower. The figure 4.8 (the last one) shows how the weights should be.

4.6. Normalization

4.6.1. Active Shape Models

Active shape Models (ASM) have been used for carrying out a normalization step. Thanks to ASM, the cropped faces provided by the detector can be rotated if they are originally turned.

Also, ASM has been used in order to find some relevant points of the face image for drawing an ellipse which perfectly adjust to the face image.

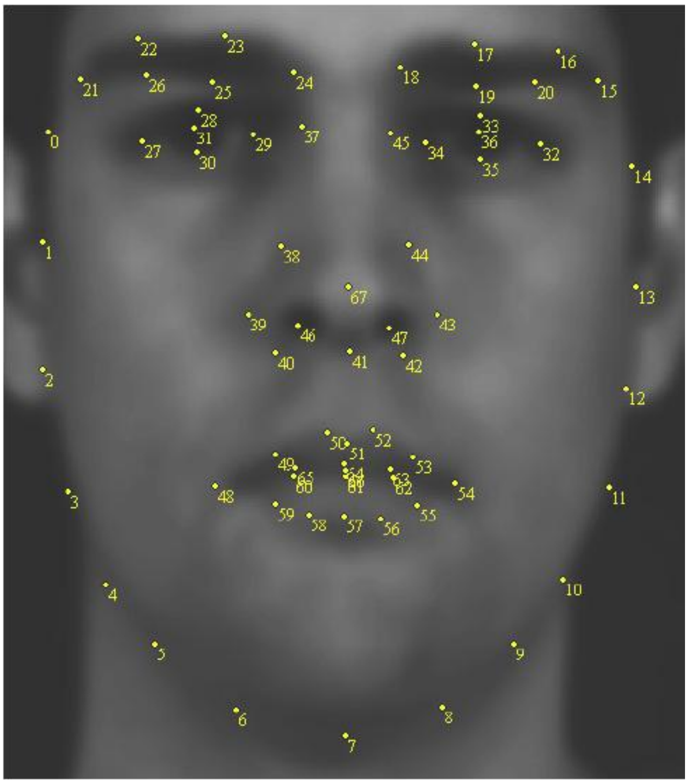

Figure 4.9- Obtained face relevant points after using Active Shape Models

4.6.2. Removing background

This approach consists on applying a mask over an image in order to discriminate some regions which can be irrelevant.

In fact, it has been thought that the regions which are related to the background of the image are not relevant and also can contribute to wrong results. So, an ellipse mask has been used in order to ignore these regions. The reason of using an ellipse mask is because with this form it is possible to adjust to the form of any face.

For adjusting the mask, the Active Shape Models have been used. Knowing some relevant points of the image, like the eyes, mouth or nose coordinates it is possible to create an ellipse which encloses the face image. For drawing the ellipse, its two diameters are needed. To find these diameters, several points of the face have been used.

In figure 4.10, Points number 1 and number 13 have been used in order to find the small diameter and points number 67 and number 7 have been used in order to find the big radius. Finally, the ellipse has been cropped in order to adjust the ellipse with the face.

Figure 4.10- On the left, the original image. On the right, the original image with an ellipse mask applied.

4.6.3. Rotation of the image

Another normalization step developed in this project has been the rotation of the images originally turned. For carrying out this rotation, ASM have been used again.

Thanks to ASM, relevant points of the original image can be compared with some relevant points of a reference image and creating a rotation matrix. These relevant points are the ones related to eyes (points number 27, 29, 32, 45, etc. See figure 4.11).

Finally, this rotation matrix is applied to the original image for achieving the rotation.

Figure 4.11. On the left, the original image. On the right, the original image after doing a rotation using ASM.

CHAPTER 5. EXPERIMENTAL RESULTS

In this chapter, all the tests and the results obtained will be presented. Note that all the tests have been proved with several sets of image.

The most important difference between these collections is that the Yale's one has abrupt illumination changes in its images. This is an important point to be considered due to improving the face recognition ratio of this kind of images is one of the main goals of this project.

All the images of theses collections are frontal faces. The images are automatically cropped and normalized to 105x105 pixels. This size has been chosen because the images have to be multiple of 7, 5 and 3 simultaneously for carrying out the tests implemented in this project. Moreover, as it has been studied in [17], image sizes from 32x32 pixels gives excellent results and decrease computational complexity.

5.1. Face Databases

For carrying out all the tests, it has been used several face Databases. In a face recognition system there a lot of parameters that have to be adjusted. So, it is necessary to use databases providing well-known, high resolution faces.

The number of coefficients kept in the transformation, the best parameters of the transformation used, combination of different techniques for developing the recognition system, etc will be set up using different types of databases.

In this section, five databases and their main characteristics will be introduced. TN and CP databases can be considered the best scenarios while Yale database describes a scenario with abrupt changes of illumination.

TN

The TN databases consist on a hundred JPG images from six individuals. These individuals are TV presenters so all the images are frontal and they do not have abrupt changes of illumination. The sizes of the images are 720 x 576 pixels and they are color-scale.

Figure 5.1- Images representing one of the individuals of the TN collection

CP

The CP databases consist on a hundred JPG images from six individuals. These individuals are politician and they are in the parliament. The images do not have abrupt changes of illumination but they can present different poses. The sizes of the images are 720 x 576 pixels and they are color-scale.

Figure 5.2- Images representing one of the individuals of the CP collection

GT

The GT databases consist on a fifteenth JPG images from fifty individuals. These images can present different poses of the individuals and also, some of them can bring glasses. Note that, this collection do not have abrupt changes of illumination. The sizes of the images are 640 x 480 pixels and they are color-scale.

Figure 5.3- Images representing different individuals of the GT collection.

ATT

This directory contains a set of faces taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge, UK.

There are 10 different images of 40 distinct subjects. For some of the subjects, the images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images are taken against a dark homogeneous background and the subjects are in up-right, frontal position (with tolerance for some side movement). The sizes of the images are 92 x 112 pixels and they are gray-scale.

Figure 5.4- Images representing different individuals of the ATT collection.

YALE

The Yale Face Database was created in Yale University. The database provides sixteenth frontal GIF images from forty individuals in several facial expressions and lighting variations. The size of images is 168 x 192 pixels, 8-bit grayscale.

Each individual undergo a set of 9 different poses and 64 illumination variations (containing abrupt changes of illumination). Faces are 8-bit, grayscale and sized 168x192 pixels

Figure 5.5- Images representing a individual of the Yale collection with different illumination changes.

5.2. Introduction to the tests

Several tests have been carried out in this project. The two used main techniques in this project have been DCT and LBP. With these two techniques, several combinations of them have been proven following three main approaches.

- The first approach consist on applying the used technique directly to the original image.
- The second one, called “Single Resolution” (SR), consist on diving the image in L^2 blocks and applying the technique to each block. In this project, when single-resolution approach is used, the image have been divided in 7^2 squared blocks.
- The last one, called “Multi Resolution” (MR), consist on diving the image first in L^2, then in L_a^2 and finally in L_b^2 and applying the techniques in all the different blocks. Note that L_a and L_b are smaller than L. With this approach, different levels of resolutions are achieved. In this project, when MR approach is used the image is divided in 7^2, 5^2 and 3^2 blocks.

5.3. Best Parameters for DCT

One of the used techniques in this project has been the DCT. This technique consist on apply the Discrete Cosine Transform to the image and then, save few coefficients for creating the model.

As it has been explained in the chapter 4, for the DCT technique, two different approaches have been used. The first one consists on applying the DCT to the entire image and extract some coefficients and the second one consists on dividing the image in 7x7 blocks and apply the DCT to each block. For each one, few coefficients are extracted. Finally, all of these coefficients are concatenated to create the model.

5.3.1. DCT over all image

When DCT is applied, few coefficients are extracted. In fact, the rule determines that the coefficients saved have to represent the 95% of the image energy. In our system, this is achieved using around 70 coefficients. However, several tests have been done taking different numbers of coefficients.

Independently of the number of coefficients, with this approach only the global information of the image is taken into consideration.

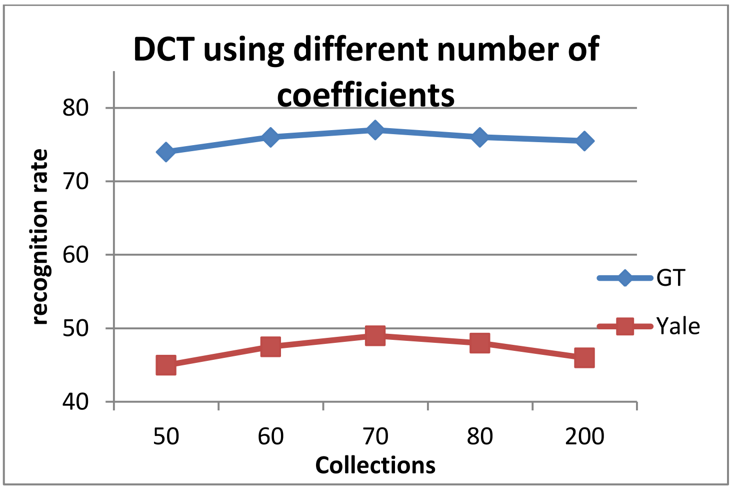

Figure 5.6- Graphic representing the recognition rate using DCT over all image and taking different coefficients. The collection used are GT and Yale.

As it can be seen, the best results are obtained when 60 or 70 coefficients are taken. Then, the results deteriorate if the number of coefficients is augmented. This is due to the important information for face recognition is given by low frequencies. High frequencies are related to details and even to noise. So, if the number of coefficients is very high, high frequencies are also taken into account. Hence, noise is taken into consideration, too.

5.3.2. DCT over blocks

In this approach, the image is divided into 7x7 blocks. For each block, the DCT is calculated and some coefficients are taken using the 95% rule.

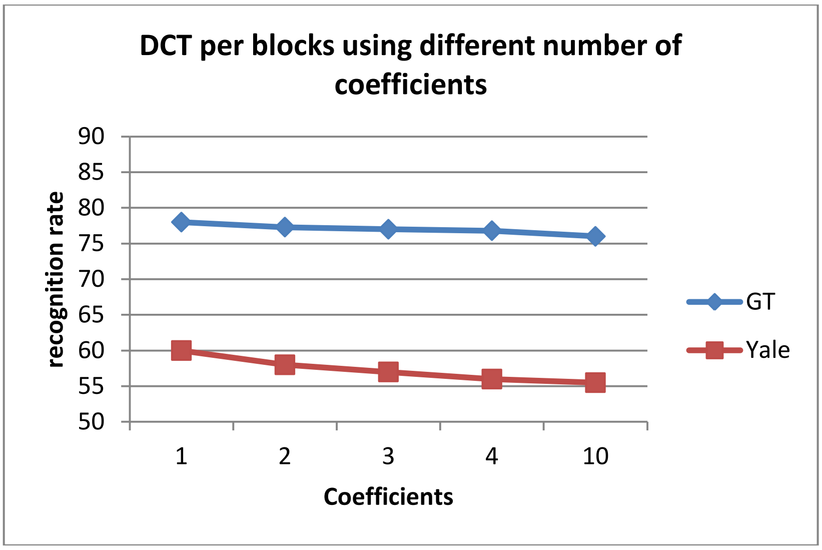

Figure 5.7- ***Graphic representing the recognition rate using DCT block based and taking different coefficients. The collections used are GT and Yale.***

Note that the best results have been achieved when only one coefficient is saved for each block. Also, this approach presents better results than the one presented in the previous section 5.1.1. However, the DCT calculated per blocks increases the computational cost and the improvement achieved is not much better in the majority of the collections. For this reason, in some cases the DCT calculated over all image may be a better option.

5.4. Local Binary Patterns

5.4.1. Influence the best LBP parameters

The Local Binary Pattern (LBP) technique has different parameters which can substantially modify the results. These parameters are: The number of neighbors which are taken in order to calculate the value of the evaluated pixel, the radius which is taken for searching these neighbors, the number of bits for the histogram calculated and uniform patterns (see chapter 2). Note that, it has been assumed that all of these parameters are independent and they will be studied separately.

So, a LBP image with 8 neighbors, a radius of 2 and the uniform patterns used, the notation used in this project is the following: **LBP 8|2|T**

As it is explained in [1], the best performance is achieved with LBP 8|2|T. However, in this project different test have been carried out in order to validate this assumption. The results are specified in the following table:

	CP		TN		GT		ATT		YALE	
	SR	**MR**	**SR**	**MR**	**SR**	**MR**	**SR**	**MR**	**SR**	**MR**
Chi Distance; Nbins = 64										
LPB 8|1|F	97,64%	96,83%	100%	100%	87,11%	88,4%	96%	98%	71,33%	62,4%
LPB 8|1|T	99,20%	96,85%	100%	100%	87,85%	88,73%	96,5%	98%	71,7%	62,18%
LPB 8|2|F	99.21%	98,43%	99,6	99,60%	88,32%	90,65%	98%	98%	76,4%	69,7%
LPB 8|2|T	99,21%	**98,41%**	**99,60**	**99,60%**	87,38%	91,12%	97%	95,5%	76,1%	**69,1%**
LPB 16|1|F	78,74%	80,31%	94,20	95,42%	77,00%	80,00%	91%	95%	30,95%	28,28%
LPB 16|1|T	79,37%	80,95	92,50	95,38%	76,21%	82,77%	90,5	94%	30,58	26,30%
LPB 16|2|F	88,98%	90,55%	97,52	99,60%	75,76%	81,70%	92,5%	94,5	37,73%	34,34%
LPB 16|2|T	88,98%	90,48%	98,34	99,17%	78,00%	81,70%	92,5%	95%	37,18%	34,06%

Table 5. 1- Table with the obtained results for the different LBP parameters. The green row shows the best parameters chosen.

As it can be seen, the best results are obtained when the radius is 2 and the neighbors are 8. However, there are not big differences when a radius of 1 or 2 is used or when uniform patterns are used.

For this reason, the best chosen parameters have been neighbors 8, radius 2 and uniform patterns not used due to its computational cost is lower. This is marked in green in table 5.1.

5.4.2. Influence of the best block approach

All of these tests have proved using two different approaches. As it has been explained in chapter 4. In the first one, the LBP image is dived only in 7x7 blocks and histograms are calculated for each block. This approach allows to know the local information of the image but not the global one. The other approach consists on dividing the image in 7x7, then in 5x5 blocks and finally in 3x3 blocks. For each block, a histogram is calculated. Finally, all the histograms are concatenated so the model has both global and local information.

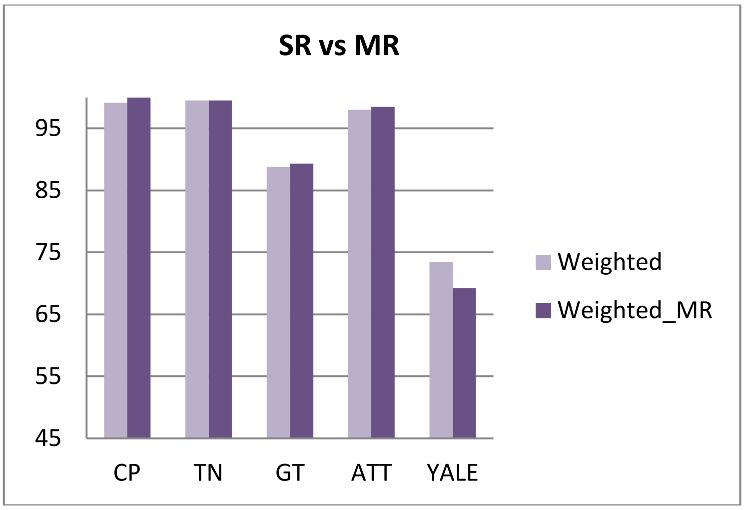

Figure 5.8- Graphic comparing the obtained results between single resolution and multilpe resolution

After doing the tests, it can be seen that the results have improved in the majority of the collections when the multi-resolution approach have been implemented. However, in the Yale's collection (the one which has abrupt illumination changes) the results have deteriorated. This can be because this collection has images with a very abrupt illumination changes so the biggest blocks are not small enough for considering that the illumination affecting the blocks is quasi-stationary.

5.4.3. Study of the computational complexity between SR and MR

It has been proved that the multi-resolution approach (MR) improved the results in almost all the image collections. However, it is important to know if this improvement provokes that the testing time greatly increases or not.

The following picture shows the time relation between the single resolution and multi-resolution approaches.

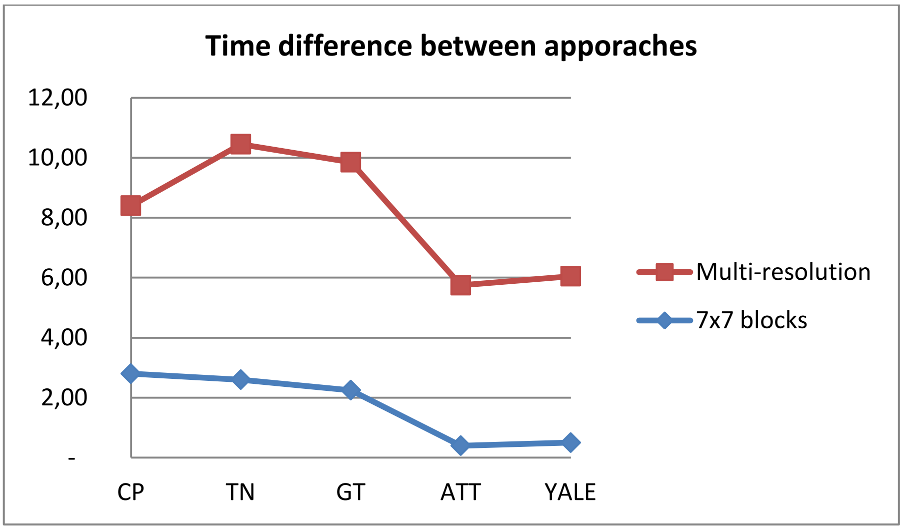

Figure 5.9- Graphic representing the computational complexity between the two approaches (SR and MR). The SR approach has been used with 7x7 blocks.

In conlusion, the best chosen approach has been the one which divided the image only in LxL blocks (SR). The reason is that both approaches present almost the same results but the multi-resultion approach increases significatively the computational cost.

5.4.4. Influence of the number of bins

Another important parameter used in the LBP technique is the number of bins of the histogram calculated for each block.

Different tests have been done in order to figure out which is the best number of bins. The results are specified in the following table:

	CP		TN		GT		ATT		YALE	
	7x7	MR	7x7	MR	7x7	MR	7x7	MR	7x7	MR
LBP 8|2|F [7x7 blocks – Chi Distance]										
Nbins 32	99,06	98,43	99,47	99,6	84,02	90,65	97,50	98,00	75,00	69,70
Nbins 64	99,21	100,00	99,6	99,44	88,32	91,50	98,00	98,00	76,40	70,70
Nbins 128	99,07	100,00	99,44	99,46	89,44	91,50	97,50	98,00	78,02	71,25
Nbins 256	99,4%	100%	99,5%	99,46%	89%	91,49%	97,5%	98%	77,84%	71,25%

Table 5.2- Table representing the obtained results using different number of bins

There are not big differences when different number of bins is used. However, in some collections the recognition rate has been augmented when the number of bins has augmented. But when the number of bins is greater than 128, in the majority of the cases, the results have not been improved more.

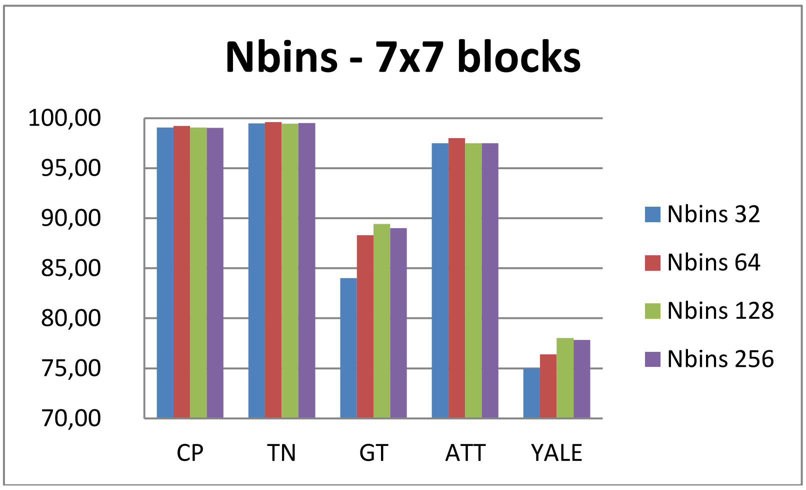

Figure 5.10- Graphic representing the recognition rates when different number of bins has been used for all the collections.

5.4.5. Computational complexity among different number of bins

It is important to evaluate if the time testing increases when the number of bins increases too.

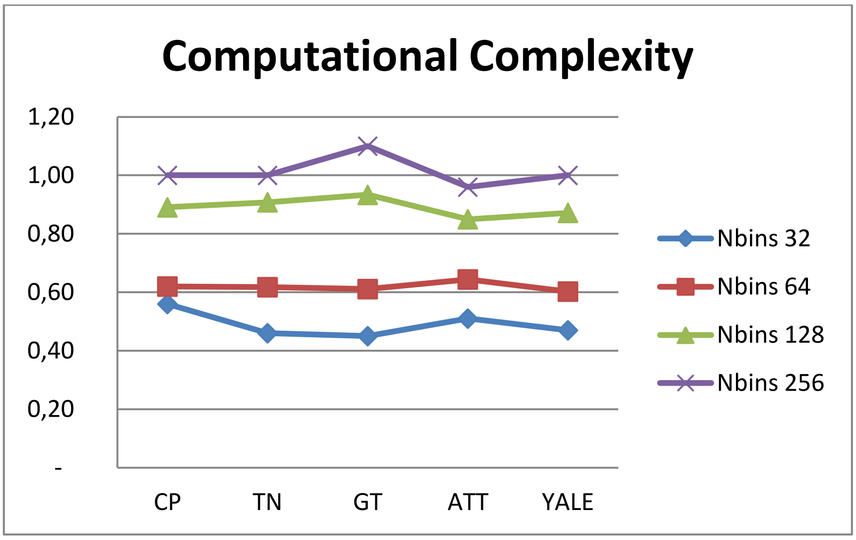

Figure 5. 11- Computational complexity when different number of bins is used.

As it shows the figure 5.11, the computational complexity always increase when then number of bins increseas. This is logic due to the the bigger is the number of bins, the bigger is the size of the enhanced histogram calculated. Hence, in the testing stage, the comparison between feature vectors needs more time.

Finally, the best number of bins used has been 64 because there is almost no difference with greater numbers but the computational charge is lower.

5.4.6. LPB over all image

In the literature, the LBP always is calculated per blocks. However, in this project, it has been proved if the LBP in the entire image presents good results. The main disadvantage of LBP block based is its high dimensionality which can be translated in high computational cost. With this approach, the dimensionality of the resulting feature vector will be reduced a lot. Hence, the computational cost will be also reduced.

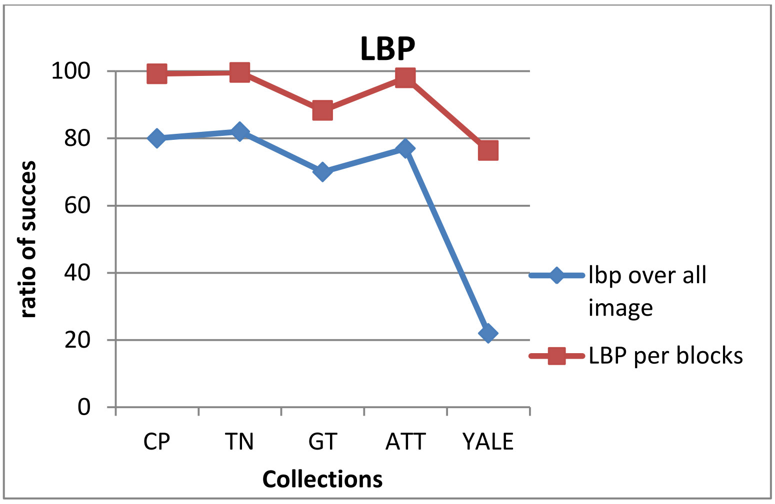

Figure 5.12- Recognition rates when LBP is used over all image or when it is used in blocks.

The obtained results have decreased in all the collections around a 15-20%, in comparison with LBP block based, but in the Yale's collection the reduction is much greater. Note that, the Yale's collection is the one with abrupt illumination changes. The reason of the results in this collection have greatly decreased is that in this approach there is not small blocks where stationary can be provided and hence, the illumination affecting the block cannot be assumed quasi-stationary.

5.5. Influence of the distance

5.5.1. When LBP is used

These tests have been repeated using difference distances. As it has been explained chapter 2, the distances used in this project have been: Euclidean Distance, Chi-Square Distance and Weighted Chi-Square Distance.

	CP		TN		GT		ATT		YALE	
	7x7	MR	7x7	MR	7x7	MR	7x7	MR	7x7	MR
LPB 8|2|T; Nbins = 64										
Weighted	99,16%	**100%**	99,52%	99,52%	88,3%	89,36%	98%	98,5%	73,44%	69,23%
Chi	99,21%	**98,43%**	**99,60%**	99,60%	88,32%	90,65%	98%	98%	76,40%	69,70%
Euclidean	99,21%	98,43%	99,17%	**100%**	88,05%	87,07%	95%	97,5%	59,52%	48,07%

Table 5. 3. Table representing the obtained results for the different images when LBP is applied per blocks.

The differences between the Weighted Chi-Square Distance and the Chi-Square Distance are not relevant. In fact, the most of the time, they achieved the same results except in the collection of abrupt illumination changes where the Weighted Chi-Square Distance obtains worse results.
However, in this kind of collection the results have been greatly improved when the Chi or Euclidean distance have been used.

The main difference is that the Euclidean distance almost never improves the results. Actually, when a collection with abrupt illumination changes is used, the ratio of success decreases a 17% approximately.

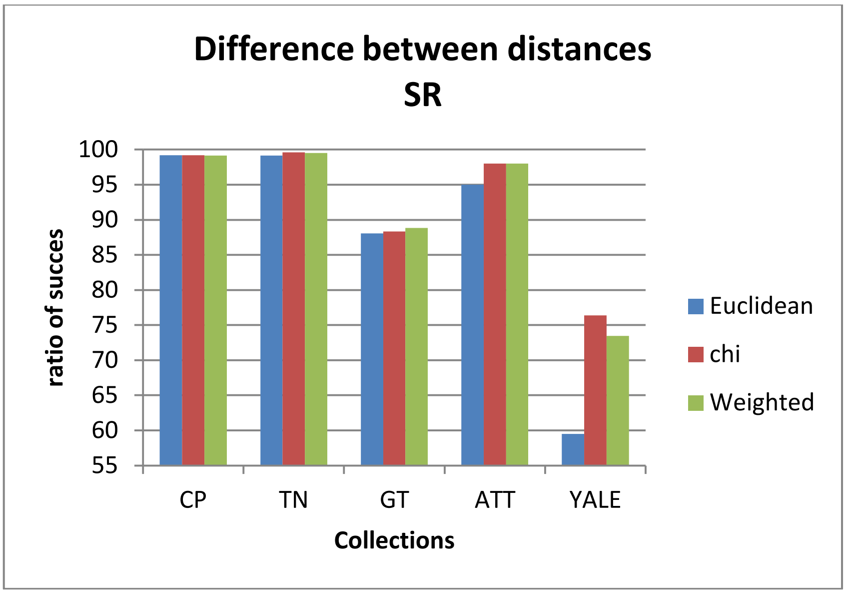

Figure 5.13- Recognition rates obtained for each used distance when LBP is applied per blocks following the SR approach.

These tests have been also proved with the multi-resolution approach and as it can be seen in the figure 5.14, there are not big difference with the 7x7 approach.

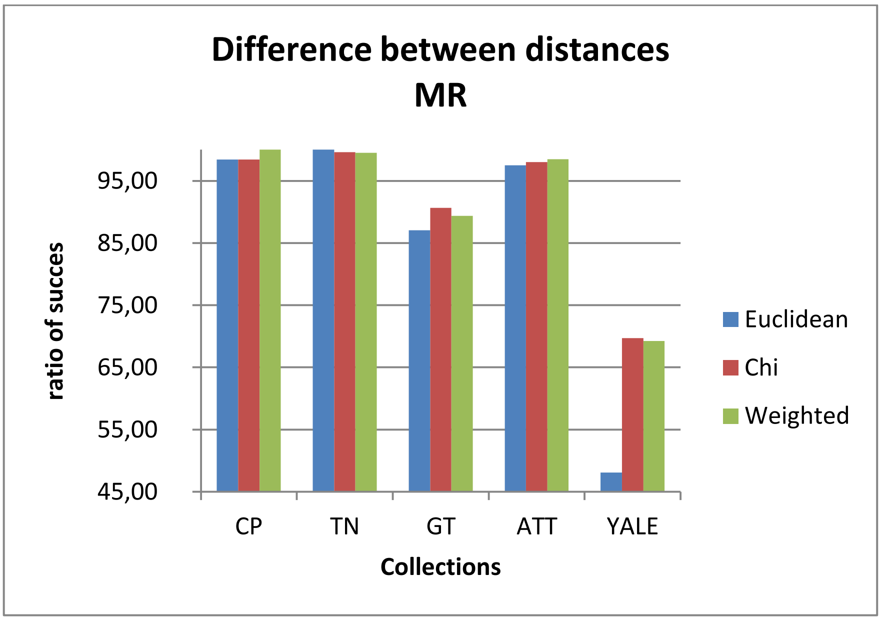

Figure 5.14- Recognition rates obtained for each used distance when LBP is applied per blocks following the MR approach.

In conclusion, the best distance when LBP is used is the Chi Square distance because it obtains almost the same results than the Weighted Chi-Square one but its computational complexity is lower.

5.5.2. When DCT is used

When DCT is used, the obtained results are very different. In this case, when Euclidean distance is used, the results are good but if Chi-squared distance is used, the results are very bad. The figure 5.15 shows the difference between the two distances.

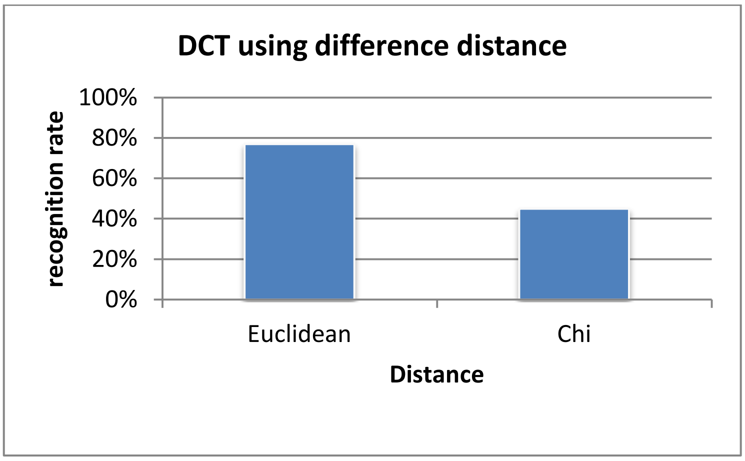

Figure 5.15- - Recognition rates obtained for each used distance when DCT is over all image.

When Chi-squared distance is used, the ratio decreases around a 35%. So, this distance cannot be used for DCT. Due to these results, the Weighted Chi-squared distance has not been proved because it is obvious that it cannot improve the results in relation to the Euclidean.

5.5.3. Computational complexity among all the distances

It has been proved which distance is faster in order to evaluate which is the one best one depending on the situation.

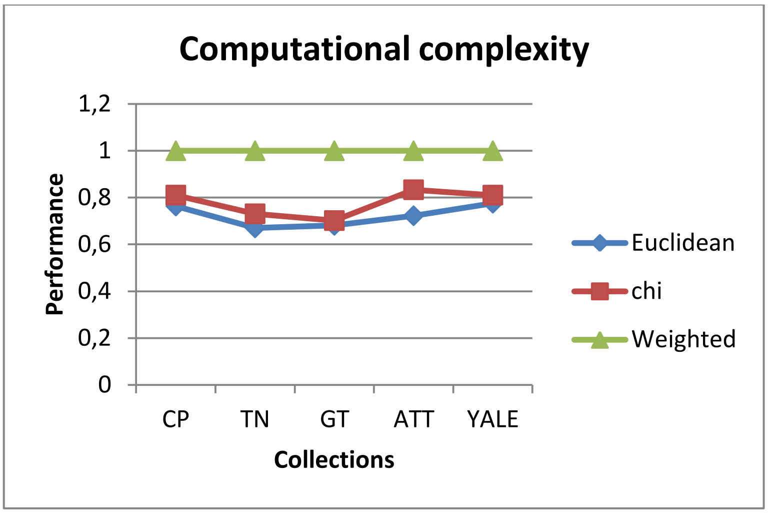

Figure 5.16- Computational complexity between distances.

The figure 5.16, shows that Euclidean and Chi-Square distance do not present big differences, although the Euclidean one has lower computational cost. However, the Weighted Chi-Square distance increases significantly the computational cost. This is due to it has to read a file from the disk which contains all the weights for each block.

5.6. LBP and DCT

In this approach, firstly LBP has been computed for then extracting the enhanced histograms. As the image is divided in 7x7 blocks, only local information is provided by the histograms.

For having global information, DCT has been applied to the entire image and some coefficients have been saved and added in the feature vector. So, finally the feature vector is composed by several LBP features and some DCT coefficients.

This approach has been thought due to previous approaches where global information was extracted has not presented good results. The following figure shows the results when Euclidean or Chi.squared distance have been used.

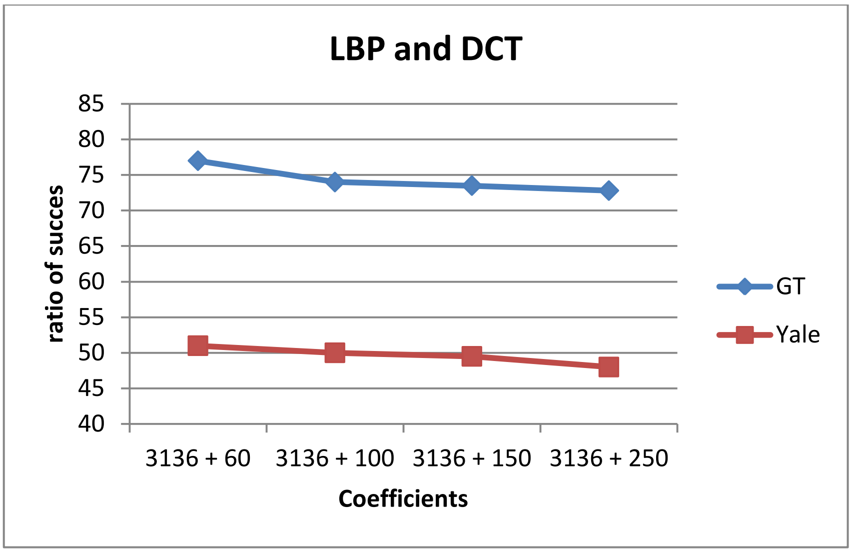

Figure 5.17- Recognition rate obtained when LBP and DCT approach is used. The image collections used is GT and Yale.

As it shows the figure 5.17, although this approach does not obtain bad results, it does not improve the obtained results when only LBP is used and even the results decreases a little bit. Moreover, the computational complexity increases because DCT is calculated.

This can be due to the using of Chi-squared distance. As it has been seen in section 5.4.2, when this distance has been used in DCT, the results have been drastically decreased. A solution can be a fusion of the results using different distances depending on the used technique in the extraction of the features. For example, Euclidean distance can be used in the features extracted from DCT and Chi-Square distance for the ones extracted from LBP. Finally, different weights should be calculated for both techniques.

5.7. DCT over LBP

In this approach, firstly LBP has been computed in order to create he enhanced feature vector composed by the histograms of each block and then, DCT is carried out.

The goal of applying DCT over the enhanced feature vector is for reducing dimensionality. As it has been explained, LBP is very robust in front of illumination changes but its computational complexity is very high due to its high dimensionality.

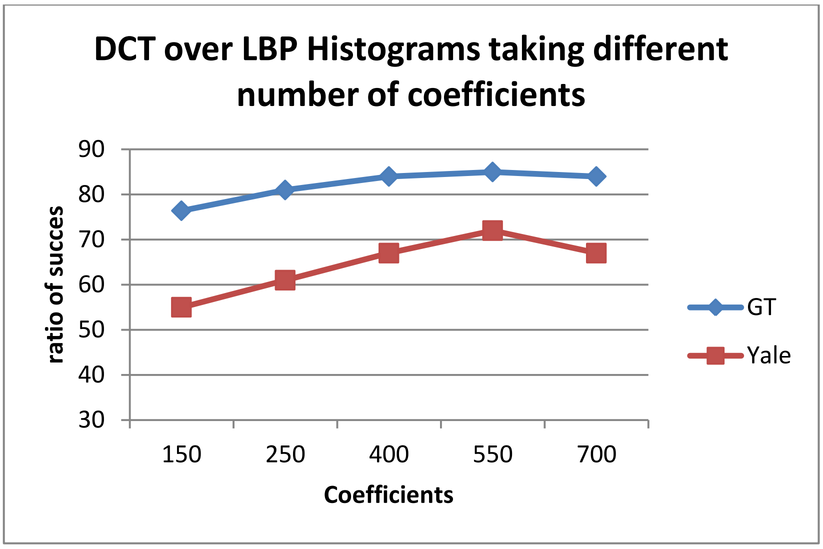

Figure 5.18- Recognition rate obtained with the DCT over LBP approach taking different number of coefficients. The GT collection is used for this test.

The obtained results are good enough. It is certain that the results have deteriorated if they are compared with the normal LBP approach, but they achieved ratios of 85% in normal collection and 70% in collections with illumination changes. Moreover, the computational complexity greatly decreases due to the dimensionality of the feature vector is 49 times lower.

5.8. Removing background

As it has been explained in chapter 4, an ellipse mask has been created using ASM in order to eliminate de background. Firstly, it has been thought, that the background does not contribute to the improvement of the results because it does not contain information of the face.

This approach has been proven using the Chi-squared distance. However, the obtained results were worse than the obtained without mask. The explanation is the following: When the image is divided into blocks, there can be completely black blocks. So, all the bins of LBP histogram of this region are ‘0’. In the classification step, the test vector is compared with all the train vector. If some histograms of these vectors are ‘0’ when distance is computed, the value of the distance is not real because the histogram has been forced to be ‘0’ due to it belongs to a region of the background.

For this reason, a second distance has been developed in order to ignore the calculation of the distance when there is a completely black block (both in the train or test vector). The results are shown in the figure 5.19.

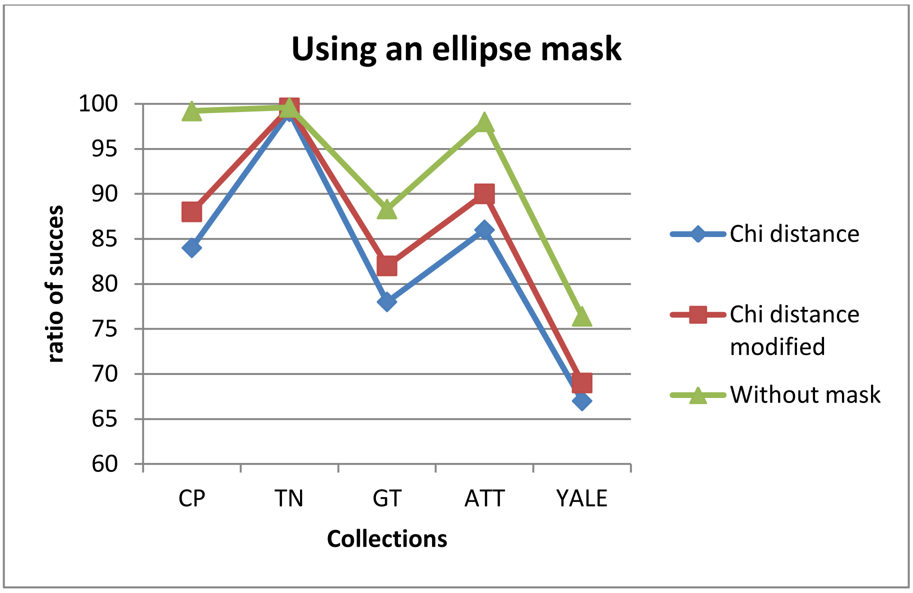

Figure 5.19- Comparison between using or not an ellipse mask able to remove the background. The chi distance modified is the one used for discriminate the completely black blocks.

Using an ellipse mask the results have not improved. However, the modified Chi-squared distance presents better results than the normal Chi-squared distance. Hence, this demonstrates that the distance within a black zones do not have to be calculated.

5.9. Computational complexity between techniques

The performance of the different used techniques and approaches has been calculated in order to choose which approach is the best one. In fact, the performance of DCT, LBP single resolution, LBP multi-resolution, LBP with DCT and DCT over LBP have been computed.

In this case, two different times have been calculated. The first one is the projection time, that is, the necessary time when the transformation is done. The results are shown in the figure 5.20.

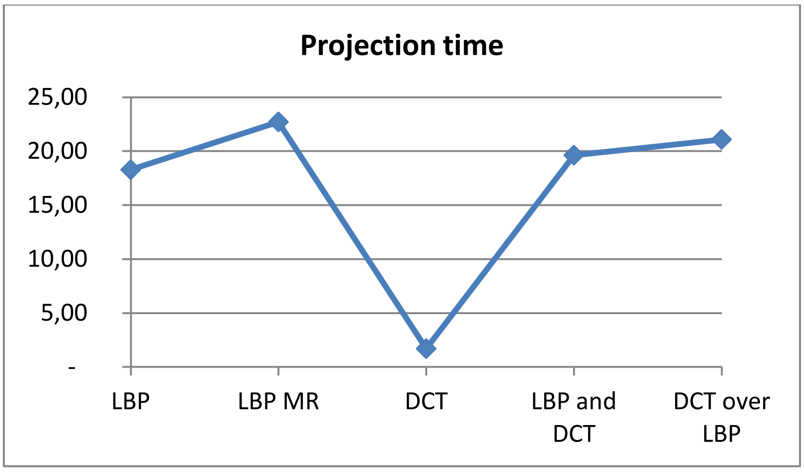

Figure 5.20. Projection time between [ms] the most important techniques and approaches used in this project.

The techniques which requires less time for doing its transformation is the DCT. As it has been explained in this project, LBP is a technique which offers better results but which needs more computational complexity. However, this time is not the most relevant.

For this reason, it has been calculated the classification time, too. It is the necessary time when a feature test vector is being compared with the feature train vectors in order to be classified.

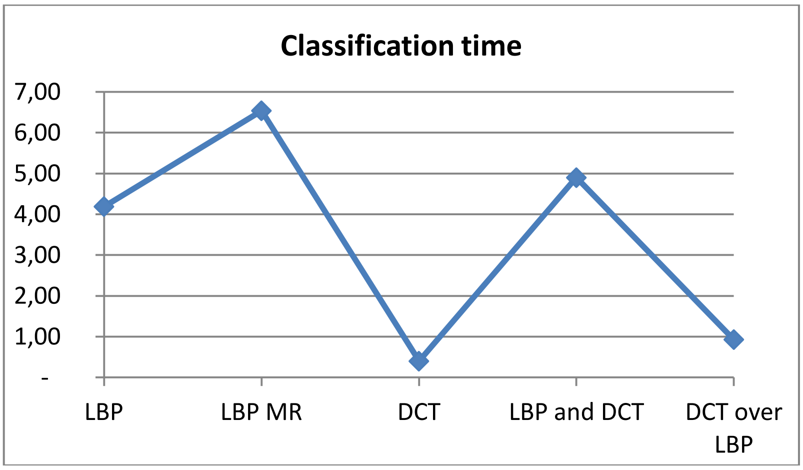

Figure 5.21- Classification time [ms] between the most important techniques and approaches used in this project.

Again, the lowest times are when DCT is used and in contraposition, when LBP is done, the time significantly increases. Note that, the time between LBP approaches is quite different. If 7x7 block approach is used the time is quite lower than when multi-resolution approach is done. That is an important point for be considered.

CHAPTER 6. Conclusions

In this project, a face identification feature-based system has been tested focusing on feature extraction. To do this extraction, two techniques have been tested: Discrete Cosine Transform (DCT) and Local Binary Patterns (LBP). DCT is a common used technique which is very fast and also achieves great results. LBP offers better results but its increase the computational complexity which can be a problem.

As it has been explained in chapter 3, different approaches and a combination of both techniques have been developed. DCT has been applied in two different ways. The first one has consisted on applying DCT over all the input gray image. The results obtained fluctuate between 80-90% in collection without illuminations problems. However, they decrease until 50% when a collection with abrupt illumination changes is used. The second approach using only DCT consists on dividing the image into blocks. It has been thought that with this division, if the blocks are small enough, the illumination will remain constant within the block. The obtained results do not present important changes in the majority of the collections except of the one with illumination changes where the results have achieved a 65%. This demonstrates that the illumination can be considered constant if the blocks are small enough. Moreover, it is important to remark, that the DCT approach spends little computational complexity which is a very important point to be considered.

One of the main goals of this project is improving the results in collections with abrupt illumination changes. For this reason, LPB has been used. In fact, LBP is a texture descriptor which is very robust in front of illumination and poses changes. With LBP, two different approaches have been used. The first one consists on dividing the image in 7x7 blocks. For each block, a histogram has been calculated and finally a feature vector has been created concatenating all the block histograms. In this approach, only local information is taken into account. In the second approach called Multi-resolution, the image has been divided into 7x7blocks, then into 5x5 blocks and finally into 3x3 blocks. The same procedure has been carried out, so a feature vector has been created concatenating the histograms of all the blocks. The difference is that in this approach, the feature vector contains both local and global information. The obtained results comparing both approaches do not present important differences in the majority of the collections. In fact, with the Multi-resolution approach, the results have improved the obtained results with the other approach a 0.75% in average. However, in the collection with abrupt illumination changes, the Multi-resolution approach decreases a 7 or 8%. And also, the computational complexity in the classification step is 1.5 times greater. For these reasons, the best approach which has been considered is the first one.

DCT and LBP techniques have been compared, too. LBP technique achieves better results in all the collections and specially, in the collection with abrupt illumination changes, where the results have increased a 25%. However, the main disadvantage of LBP is it computational complexity. In relation with DCT, its computational complexity is almost 10 times greater. This is a problem if recognition has to be done in real time.

For this reason, we propose to create the feature vector by applying DCT to the vector resulting of the concatenation of the block histograms on the LBP image. Thanks to DCT, the dimensionality of the vector is drastically reduced. In fact, it will be forty lower. Dimensionality is strictly related with computational complexity.

The obtained results are very good. It is certain that they have been reduced a little bit (3-5%) but the computational complexity has become 4.5 times lower, achieving similar computational complexity as when DCT is used.

Moreover, as the Multi-resolution approach has been discarded due to its high computational cost, a combination of DCT and LBP has been tried. The goal is using the local information provided by LBP and also, the global information provided by DCT. The obtained results are not the ones expected. In all the collections, the results have decreased around a 10% if it is compared with LBP and around a 2% if they are compared with DCT. In addition, the computational complexity has obviously increased because DCT and LBP have been calculated. For this reason this approach has been discarded.

Finally, another approach has been developed. In this approach, the used technique does not matter. The goal is erasing the background of the face image because it can influence in bad results. So using the Active Shapes Models, an ellipse mask is created. This mask is adjusted to the face of the image and LBP is calculated taking into account only the regions different to ‘0’ in the mask. Another time the results are not the ones expected. In fact, they are reduced around a 15% in the majority of the collections.

In conclusion, LBP is a robust technique when collection with illumination and pose changes are used but it requires a high computational complexity. For this, reason, the best approach could be the one which apply DCT over the feature vector created from LBP because it achieves very good results and its computational complexity is quiet low (similar to DCT).

BIBLIOGRAFY

[1] Timon Ahonen, Abdenour Hadid and Matti Pietikäinen. *Face Description with Local Binary Patterns: Application to Face Recognition*. Infotech Oulu. Finland (2004)

[2] Hazim Ekenel, Rainer Stiefelhagen. *Local Appearance based Face Recognition using Discrete Cosine Transform.* Germany (2005).

[3] Xiaoyu Wang, Tony X.Han, Shuicheng Yan. *An HOG-LBP Human Detector with Partial Occlusion Handling*. University of Missuri. Columbia (2008)

[4] Matthew A. Turk and Alex P.Pentland. *Face Recognition Using Eigenfaces*. Vision and modelling group, The media laboratory. Massachusetts (1991)

[5] Ignas Kukenys and Brendan McCane. *Support Vector Machines for Human Face Detection.* New Zeland (2008)

[6] Hazum K. Ekenel, Mika Fischer, Erkin Tekeli, Rainer Stiefelhagen and Aytül Erçil. *Local Binary Pattern Domain Local Appearance Face Recognition*. Germany (2008).

[7] Caifeng Shan, Shaogang Gong and Peter W. McOwan. *Robust Facial Expression Recognition Using Local Binary Patterns.* London, UK (2005).

[8] Sébastien Marcel, Yann Rodriguez and Guillaume Heusch. *On the Recent Use of Local Binary Paterns For Face Authentication.* Switzerland (2006).

[9] Bernd Heisel, Purdy Ho, Tomaso Poggio. *Face Recognition with Support Vector Machines: Global versus Component-based Approach.* Massachusetts (2006).

[10] David Cristinacce and Tim Cootes. *Boosted Regression Active Shape Models.* Manchester, UK (2007)

[11] Aman R. Chadha, Pallavi P. Vaidya, M. Mani Roja. *Face Recognition Using Discrete Cosine Transform for Global and Local Features.* Mumbai, India (2008).

[12] Javier Ruiz-de-Solar and Julio Quinteros. *Illumination Compensation and Normalization in Eigenspace-based Face Recognition: A comparative study of different pre-processing apporaches.* Chile (2008).

[13] Michelle M. Mendonça, Juliana G. Denipote, Ricardo A.S. Fernandes, María Stela V.Paiva. *Illumination Normalization Methods for Face Recognition.* Sâo Paulo (2007).

[14] Kelsey Ramirez-Gutierrez, Daniel Cruz-Perez, Jesus Olivares-MErcado. *A Face Recognition Algorithm using Eigenphases and Histogram Equalization.* Proceedings of the IEE: issue 1, vol.5, 2011.

[15] Mireia Farrús, Pascual Ejarque, Andrey Temko, Javier Hernando. *Histogram Equalization in SVM Multimodel Person Verification.* Barcelona, Catalonia (2009).

[16] W.Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld. *Face Recognition: A litterature Survey.* Maryland (2003).

[17] Fix, E. and Hodges, J. L. (1951) Discriminatory analysis – nonparametric discrimination: Consistency properties. Tech. Rep. 4, Project no. 21-29-004, USAF School of Aviation Medicine, Randolph Field, Texas

[18] Pawan Sinha, Benjamin Balas, Yuri Ostrovsky and Richard Russel. *Face Recognition by Humans: Nineteen results all computer vision researchers should know about.* Proceedings of the IEE: vol. 94, No. 11, November 2006.

[19] Hazim Ekenel, Rainer Stiefelhagen. *Local Appearance based Face Recognition using Discrete Cosine Transform.* Germany (2005).

[20] Amir Omidvarnia. *PCA based Face Recognition System.* Australia*. October 2007*

[21] Paul Viola and Michael Jones. *Robust Real-time Object Detection.* Vancouver, Canada. July 2001.

[22] T.Ojala, M.Pietikinen and D.Harwood. *A comparative study of texure measures with classification based on features distribution.* Pattern Recognition, vol.29, nº.1, 1996.

[23] OpenCV library. http://opencv.willowgarage.com/wiki/

[24] Aizerman, Mark A.; Braverman, Emmanuel M.; and Rozonoer, Lev I. (1964). "Theoretical foundations of the potential function method in pattern recognition learning". *Automation and Remote Control* 25: 821–837.

[25] Viola, Jones: Robust Real-time Object Detection, IJCV 2001

[26] Yoav Freund and Robert E. Schapire. A short introduction to boosting. In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence ,pages 1401_1406. Morgan Kaufmann, 1999.

[27] W. Zhao, R. Chellappa, and P.J. Phillips. *Subspace linear discriminant analysis for face recognition*. *Technical report.* 1999.

[28] Emanuel Parzen. *On estimation of a probability density function and mode. The Annals of Mathematical Statistics*, 33(3):1065_1076, 1962.