A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF for security and face recognition

Book · April 2018

4 authors:

Ayman Alfalou
Yncréa Ouest

Maher Jridi
Institut Supérieur de l'Electronique et du Numérique

Yassin Kortli
University of Bretagne Occidentale (UBO)

Mohamed Atri
King Khalid University

A Comparative Study of CFs, LBP, HOG, SIFT, SURF, and BRIEF for security and face recognition

Yassin Kortli *†, Maher Jridi†, Ayman Al Falou†, and Mohamed Atri *
* Electronic and Micro-electronic Laboratory
Faculty of Sciences of Monastir, University of Monastir, 5000 Monastir, Tunisia
† Équipe Vision, Laboratoire L@bISEN
20 rue Cuirassé Bretagne CS 42807, 29228 Brest Cedex 2, France
yassin.kortli@isen-ouest.yncrea.fr , maher.jridi@isen.fr, mohamed.atri@fsm.rnu.tn, ayman.al-falou@isen.fr

Abstract: In the last two decades, face recognition has become an interesting research, and used in different disciplines such as computer vision, image processing, pattern recognition, and machine learning. We find several methods local, subspaces and correlation filters of face recognition. The purpose of this project is to establish a comparative study between six approaches, one based on the local feature methods and the second one based on the correlation methods. However, the comparative study between these methods is carried with distance measurements and Peak to Correlation Energy methods. In addition, the evaluation techniques are the yardstick to present the performance and efficiency for each face recognition system. These techniques allows to interpret, compare, and distinguish the different factors such as images, illumination, location and characteristics of subjects. In this chapter, we review, discuss, and compare some common face recognition algorithms. Our objective of this work is to demonstrate the effectiveness and feasibility of the best methods for face recognition in terms of design, implementation, and application.

1. Introduction

In recent years, developing Surveillance Systems (SS) for security has been one of the most active research fields in most applications. These systems used to adjust, enhance, and improve the security. One of these systems, face recognition system plays an efficient and very important tool for several applications including access military, civilian, control, and system security can be employed for optical surveillance systems [1][2]. In addition, the face recognition system is usually used to adjust, enhance, and automate the identification process for security and safety. This system can be used to assist and help police to understand any suspect behavior and to launch adequate measure to tackle that.

Face recognition based on vision systems is almost performed by two processes: the detection of the face and then the recognition. Face detection method uses many algorithms for detect and locate a face in an image obtained by the system. Besides, the face recognition process includes the feature extract from the unknown face in order to compare with all template faces database to decide the identity of this unknown face (‘acceptance or rejection’). Moreover, a face recognition system present three major processing modules [3]: preprocessing and detection module, feature extraction module, and a recognition module. The recognition (or decision) is done by matching the features stored in the system datasets with the features extracted from the face unknown.

The face recognition is a task that humans perform naturally and effortlessly in their daily lives. The high availability of powerful and inexpensive computers and embedded computing systems has generated enormous interest in the automatic processing of digital images and videos in many applications, including biometric identification, monitoring, human-machine interaction and multimedia data management. The recognition of the face can be used without direct contact. It can be used anywhere and used in various applications. However, it suffers from various changes of the same face and the large size of the image. In literature, it can be found that some techniques are used exclusively for the detection process, and some others are dedicated to the recognition process with high robustness and discrimination [4]. For the past two decades, various face recognition methods have been proposed to reduce the amount of calculation and improve the recognition rate. These proposed methods could be categorized into three significant categories:

- ***Local feature approaches***
- ***Subspace learning approaches***
- ***Correlation filters approaches***

Local feature methods are mainly proposed for face recognition and all dhow best and high performance in this task. Many methods implemented for representing the face such as Gabor filters[5], Local Ternary Patterns (LTP) [5], Local Binary Patterns (LBP) [6], Histogram Of Oriented Gradient (HOG) [7], Scale Invariant Feature Transform (SIFT)[8], Binary Robust Independent Elementary Features (BRIEF) [9], and Speeded-Up Robust Features (SURF) [10], etc.

Unlike the local approaches, Subspace learning approaches or global approaches are an another approaches uses for face recognition. These approaches include Fisher linear discriminant analysis (LDA) [11], principal components analysis (PCA) [12], and locality preserving projections (LPP) [13]. Subspace learning approaches are very methods to obtain discriminative subspace. However, these approaches generally ignore local details. This characteristic makes the Subspace learning approaches do not use much for face recognition.

Correlation approaches like Vander Lugt Correlator (VLC), and Joint Transform Correlator (JTC) [2], [14], and [15] are used for face recognition applications. These methods many used due to their ability of instantly detecting, and their discrimination and estimating target objects.

However, these techniques have many problems to surmount. For example, the same person face images can look profoundly different due to changes in the head orientation, illumination conditions, and expression [2]. It is more difficult for any recognition system to obtain a correct decision; because of lighting conditions and unexpected noise interference [15]. To build a reliable system that can deal with all of these problems, we test and evaluate some face recognition techniques in order to choose the best technique regarding computation time and algorithmic complicity.

The rest of the paper is organized as follows: In Section 2, we overview face recognition techniques in order to select some of them. Section 3, is dedicated to the description and implementation of the many face recognition systems. Section 4, presents experimental setup and discusses obtained results obtained under various conditions. The last section, concludes the work and discusses future research directions.

2. Related Work

In recent years, face recognition system in real-time using monocular vision has become a reputed area of studies in the security surveillance systems. In the literature related to computer vision, feature extraction allows the detection and classification of faces. This is possible by the use of symmetry and edges on the features to detect faces. Two main classes are applied to identify faces: detection (extraction) and recognition (identification) modules [15]. The choice made during one step can influence the other step. Detection module after preprocessing step consists in obtaining the face region for each image. Recognition module comprises of description and extraction the feature vectors of the face region obtained in the first step, and then compare with all template faces database. In this section, we explain the overview of many methods related to face detection and recognition.

However, there are many methods for face detection, and recognition tasks rely on describing the local features of an image, the location of the main points or points of interest in each image and specify the region pixels feature around the point of interest.

In addition, local feature methods and correlation methods can divide into two categories: local appearance-based methods and interest point based methods. The first methods includes feature analysis (e.g., Local Binary Patterns (LBP), Histogram Of Oriented Gradient (HOG) and direct correlation methods). These methods divided the face into blocks (or regions) for extraction the local features or directly extract the local features using correlation. The second methods are composed of points of interest methods, contours or edges methods, corners methods. These methods are detected and extracted the point of interest of an image and extract the local features localized on these points (e.g., Scale Invariant Feature Transform (SIFT), Binary Robust Independent Elementary Features (BRIEF), and Speeded-Up Robust Features (SURF)).

2.1. Local feature-based methods

It is a geometrical method, it is also called the method to features, to local characteristic, or analytic. In these methods, the face is represented by a set of characteristic vectors of low dimensions, rather than by a single large dimension vector. Local methods are concerned with the critical points of the face such as nose, mouth, eyes; what it will generate more details. The advantage of these local methods is that they take into account the particularity of the face as a natural form to be recognized and a reduced number of parameters. Also, these methods describe local features by local orientation information, histograms, geometric properties, and correlation. The most well-known methods in these approaches are Histogram of Oriented Gradients (HOG) [7], Local Binary Patterns (LBP) [16], and [16], and correlation filters [2].

Local feature descriptors have present good performance in some application in pattern recognition. Other advantages of these descriptors are robust to scaling, rotation, and translations.

- ***Histogram Of Oriented Gradients (HOG):*** the HOG descriptor has been one of the best process accepted to describe image local texture and it is one of the best features of shape and edge information. The HOG descriptor can be describe the shape of the face by distribution of edge direction or light intensity gradient. The process of this technique done by sharing the whole face image into cells (small region or area), a histogram of pixel edge direction or direction gradients is generated of each cell, and finally, the histograms of the whole cells are combined to extract the feature of the face image. It is displays great success in face recognition [7].
- ***Local Binary Patterns (LBP):*** LBP is very great general texture descriptor used for feature extraction from image, and has been widely used in a lot of applications, such as face recognition, facial expression recognition, texture segmentation, and texture classification. The most important advantages of this descriptor are its low computational complexity for describing local scene regions (good texture structure description), its invariance to monotonic gray-scale value changes and convenient multi-scale extension. At each pixel, the LBP descriptor presented as binary comparisons of pixel intensities between its eight surrounding pixels and the

center pixel. Also, this descriptor has been successfully used in pattern recognition [16]. In this work, for HOG and LBP descriptors, the matching task can be done by distance measures such as Chi-squared distance and histogram intersection. These distances are used to obtain the correspondence between two images [17]. The HOG and LBP descriptor inherits the advantages of Chi-squared distance, which well handles the unbalanced nature of HOG and LBP histograms [18].

- ***Correlation Filters (CFs) :*** The correlation filters used for pattern recognition applications due to their ability of instantly detecting and their discrimination capability and estimating target objects. The correlation filters are based on an optical configuration called 4f-setup. This configuration is done by three planes, the input plane, the Fourier plane and the correlation plane. A convergent lens separates each two planes. The first lens performs the Fourier transform of the target image. The second lens, after the multiplication of the correlation filter (get from the reference image) by the spectrum of the target image (correlation filter * spectrum of target image), performs the inverse Fourier transform to get the correlation result, characterized by the central correlation peak. The matching process is used to classify the face image when the features are selected and extracted. A wide variety of matching algorithms use in pattern recognition. However, to get this matching process as reliable and fast, it is necessary to establish an efficient technique. To evaluate the correspondence of the correlation, the Peak to Correlation Energy (PCE) is used decide the degree of the similarity between the reference and the target images [1].

2.2.Interest-point based methods

Many methods based point of interest rely on describing the features of an image, the location of the main points or points of interest and describe the region pixels feature around the point of interest. Scale Invariant Feature Transform (SIFT)[8], Binary Robust Independent Elementary Features (BRIEF) [9], and Speeded-Up Robust Features (SURF) [10] are methods based on the detection of interest points to describe the feature of an image.

- ***Scale Invariant Feature Transform (SIFT):*** SIFT is one of the most popular descriptor based point of interest that uses the intensity without color information. For this purpose, the Difference of Gaussian (DoG) is used to identify interest points in image region which are invariant to orientation, scale, illumination, and zoom. This descriptor has been widely used in pattern recognition and classification systems such as face recognition, visual mapping, image stitching. However, when the scale or the size of the image of dataset increases significantly, the disadvantage of the SIFT descriptor is their high computational cost[8].
- ***Speeded-Up Robust Features (SURF):*** The SURF descriptor deduced from the SIFT descriptor, even outperforms or approximates SIFT descriptor on robustness, distinctiveness, and repeatability, and also can be compared and calculated more quickly [19]. In this descriptor, the scale changes by altering the size of the box filter instead of changing the image size. Therefore, it significantly minimize the number of operations for the simple box convolutions. It is used mainly for classification, 3D reconstruction, image registration, face recognition, and extraction points of interest. Also, to describe the local feature, the SURF descriptor it uses a 64-dimensional feature vector, while the SIFT uses 128 that increases the processing speed. Besides, the SIFT descriptor is more adapted to describe faces affected by illumination deformations, scaling, translation, rotation [3]. In matching case, a Euclidean-distance based nearest-neighbor matching can be used for matching vector features. In addition, the best nearest-neighbor technique

depends on the descriptor used for the extraction of the feature vector. Also, the distance ratio between two neighborhoods must be required to be less than a predefined threshold, in order to eliminate the candidates which are considered as noise. However, a Euclidean-distance based nearest-neighbor matching can be used for matching vector features. In addition, the best nearest-neighbor technique depends on the descriptor used for the extraction of the feature vector. Also, the distance ratio between two neighborhoods must be required to be less than a predefined threshold, in order to eliminate the candidates which are considered as noise. In the literature, in the high-dimensional features, the fast library for approximate nearest neighbors (FLANN) and the randomized k-d forest are the most efficient techniques have been used for matching. In the literature, in the high-dimensional features, the fast library for approximate nearest neighbors (FLANN) and the randomized k-d forest are the most efficient techniques have been used for matching. In this work, for SIFT and SURF descriptors, the fast library for approximate nearest neighbors (FLANN) is used to obtain the correspondence between two images [20].

- ***Binary Robust Independent Elementary Features (BRIEF):*** BRIEF is a binary descriptor that is simple and fast to compute. This descriptor is based on the differences between the pixel intensity that is similar to the family of binary descriptors such as BRISK and FREAK in terms of evaluation. To reduce noise, the BRIEF descriptor smooths the image patches. After that, the differences between the pixel intensity is used to represent the descriptor. This descriptor has achieved best performance and accuracy in pattern recognition. Moreover, for binary features, for example, BRIEF or FAST, the FLANN, and randomized k-d forest techniques are not adequate. Then, the Hamming distance is used to compare the binary features in a matching case. The Hamming distance instead of Euclidean distance computed by performing the sum of XOR between two binary descriptors followed by a bit count on the result [20].

Generally, for the descriptors based on interest points, the choice of associated image descriptors and the properties of the underlying interest points have affected the performance of matching tasks. In this study, six of the most frequently-used features, HOG, LBP, VLC, SIFT, SURF, and BRIEF are implemented and compared.

3. Methods implementation

In this section, we describe our implement methods for face recognition. We apply six algorithms of feature extraction for the three approaches: the VLC for the correlation approaches and the LBP, HOG, SIFT, SURF, and BRIEF for the local approaches. For each image obtained from PHPID datasets, preprocessing step is applied to enhance the feature extraction in the image. After the extraction phase, we obtain feature vectors for each target and reference images. Next, we use a metric parameter to compare between these images (target and reference images). Finally, our system determines the decision, depending on the comparison scores. The flowchart of our proposed system is presented in *Figure 3.1*.

To local face deformations in terms of insensitive and discriminative matching, the face structure or content needs to be described within a suitable descriptor. This descriptor must be indicated by the main orientations and their red squares of a set of interest points extracted from a face. There are many descriptors of facial features in the literature. In the following parts, we discussed the most common descriptors. Experimental results, for each descriptor, for face description under different lighting conditions are present below.

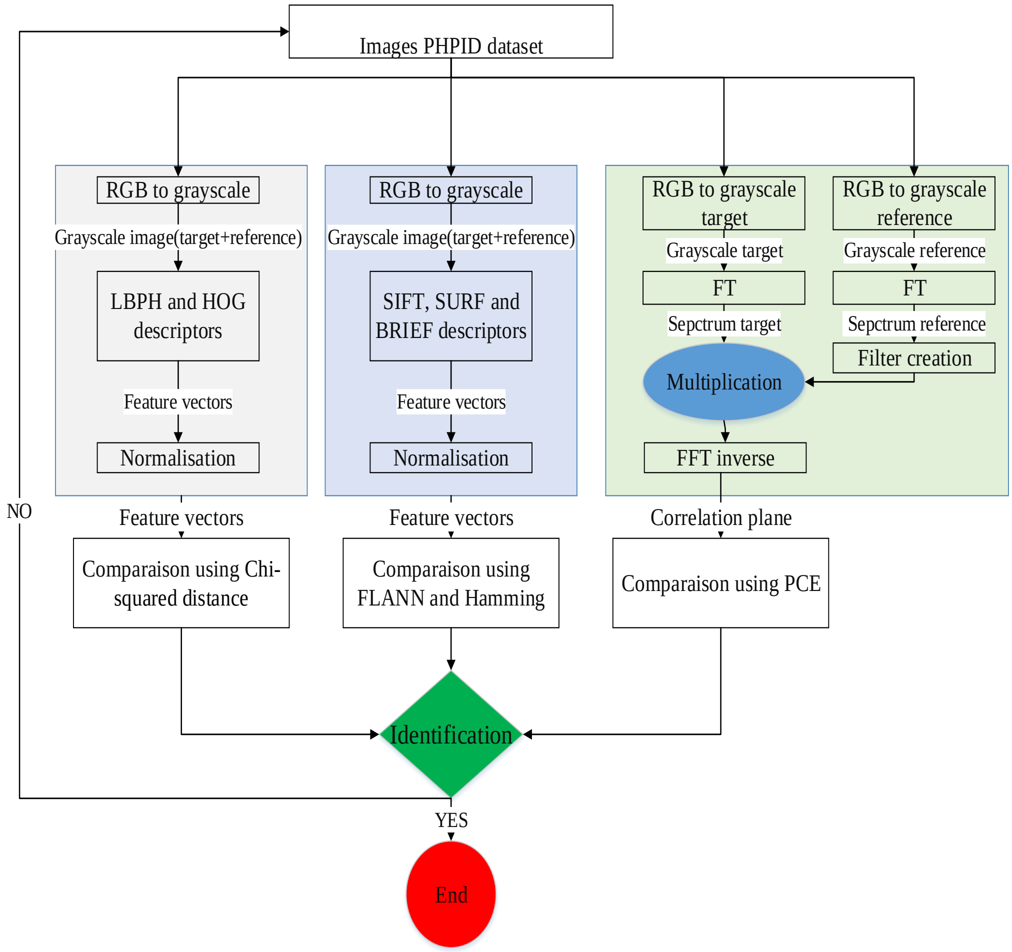

Figure 3.1: Flow diagram of the implemented face recognition algorithms

3.1. Histogram of Oriented Gradients (HOG)

The HOG descriptor is one of the most popular descriptors used for feature extraction in local region by taking orientation histograms of edge intensity. The feature vector computation by the HOG descriptor proceeds as follows:

- *Calculation of the gradient:*

At first, to obtain the HOG descriptor, the gradient image is calculated by an appropriate filter mask (e.g., Laplacian, Sobel, and Prewitt filters) in order to extract the edge gradients and orientations. Next, depending to their gradients and orientations, a grid of histograms is created, where every histogram votes the respective gradients using its length into small spatial area called ‘‘cell’’. Divide the local image into regions called cells, and then calculate the amplitude of the first order gradients of each cell in both the horizontal and vertical direction r (e.g. equation (3.1) and (3.2)). The most common method is to apply a 1D mask [-1 0 1].

$$G_x\ (x, y) = \mathrm{I}(x+1,y) - \mathrm{I}(x-1,y) \tag{3.1}$$

$$G_y\ (x, y) = \mathrm{I}(x,y+1) - \mathrm{I}(x,y-1) \tag{3.2}$$

where, $\mathrm{I}(x, y)$ is the pixel value of the point (x, y), $G_x\ (x, y)$ and $G_y\ (x, y)$ denote the horizontal gradient amplitude and the vertical gradient amplitude, respectively.

The amplitude and gradient direction of the pixel (x, y):

$$G\ (x,y) = \sqrt{G_x\ (x,y)^2 + G_y\ (x,y)^2} \tag{3.3}$$

$$\theta\ (x,y) = \tan^{-1}\left(\frac{G_y\ (x,y)}{G_x\ (x,y)}\right) \tag{3.4}$$

- *Each block 16 * 16 is divided into 2 * 2 cells*

16 * 16 Gaussian weight blocks to adjust the amplitude of the gradient, each range of bins is $\left[\frac{k\pi}{9}, \frac{(k+1)\pi}{9}, k = 0,1,\ldots\ldots 8\right]$. The amplitude of the gradient and the orientation of each pixel in the cell are voted in 9 bins with the tri-linear interpolation. The tri-linear interpolation is important, which makes the image block of the edge zone better allocated to improve the stability of the features.

- *Concatenate the histograms of all blocks*

The histograms of the four cells of each block concatenate on a vector of 36-D characteristics which are normalized by L2-Hys to reduce the influence of the local variation of the illumination and contrast of the foreground.

The HOG descriptor has been one of the best process accepted to describe image local texture, and it is one of the best features of shape and edge information. The HOG descriptor can describe the shape of the face by the distribution of edge direction or light intensity gradient. It displays great success in face recognition.

3.2. Locally Binary Patterns

Locally binary patterns (LBP) are a one of the high texture descriptor approach, it's have been highly used in different applications. The LBP descriptor was initially revealed for face recognition and texture classification, due to their robustness to binary coding of threshold intensity values. Generally, the LBP descriptor works in a $3x3$ pixel matrix $(p1 \ldots p8)$ of an image. The pixels of this matrix are thresholded with the value of the center pixel p_0 (i.e., use the intensity value of the center pixel $i(p_0)$ as reference for thresholding) to produce the binary code. If a neighbor pixel's value is lower than the center pixel value, it is given a zero, if not one. Finally, this code binary multiplied by powers of two and then summed to obtain a locally binary pattern descriptor value for the center pixel. The LBP defined in a matrix of size 3x3 and it is as given by Equation (3.5).

$$LBP = \sum_{p=1}^{8} 2^p s(i_0 - i_p),\ with \qquad s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x \leq 0 \end{cases} \tag{3.5}$$

where i_c and i_p are the intensity value of the center pixel and neighborhood pixels, respectively. Also, the *Figure* 3.2 illustrate the LBP descriptor.

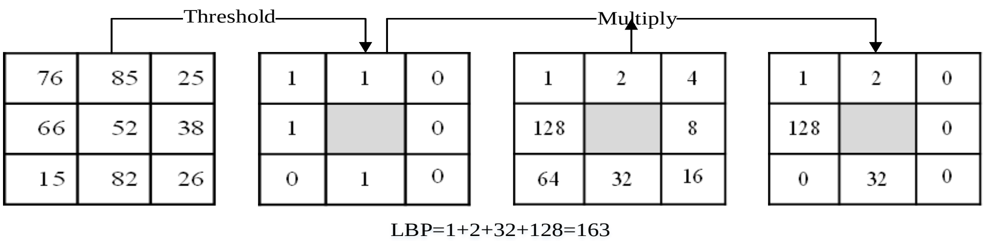

Figure 3.2: The LBP descriptor

The most influential benefits of LBP are its low computational complexity, convenient multi-scale extension, and invariance to monotonic gray-scale changes. LBP descriptors have three patterns: rotation invariant uniform pattern, rotation invariant pattern, and uniform pattern. LBP is defined as a set of binary commands comparison of pixel intensities between the central pixel and its eight surrounding pixels in the image. In fact, the LBP descriptor was used in this report for face authentication or verification. In this paper, for HOG and LBP descriptors, the matching task is performed with the Chi-squared distance. This distance is used to find similarity between two histograms [17]. The HOG and LBP descriptor inherits the advantages of Chi-squared distance, which well handles the unbalanced nature of HOG and LBP histograms [18]. The Chi-Squared ($\chi 2$) was successfully used for local descriptors matching, shape classification, object categories and texture classification[21], and boundary detection.

Generally, in the natural histograms the difference between small bins is greater significant than the difference between large bins. The Chi-Squared ($\chi 2$) is a histogram distance that takes this into account. To compare two histograms $S_1 = (u_1, \ldots\ldots.. u_m)$ and $S_2 = (\mathrm{w}_2, \ldots\ldots.. \mathrm{w}_m)$ Chi-Squared ($\chi 2$) distance can be defined as:

$$\chi 2 = D(\mathrm{S}_1, \mathrm{S}_2) = \frac{1}{2}\sum_{i=1}^{m}\frac{(u_i - w_i)^2}{u_i + w_i} \tag{3.6}$$

where m is the dimensionality of the spatially enhanced histograms.

The $\chi 2$ histogram distance is performed to test the difference between an observed frequencies and distribution.

3.3. Correlation filters

The correlation filters are highly used in face recognition task. There are two main approaches for correlation filters: the Joint Transform Correlator [22], [23] and the Vander-Lugt Correlator. In this article, we will use the second approach “Vander-Lugt Correlator” for face recognition. The Vander Lugt Correlator (VLC) have been many used for face recognition applications due to their ability of instantly detecting and their discrimination capability and estimating target objects [15]. In addition, the similarity matching based on the PCE criterion is used to obtain the recognition results.

The VLC technique, as shown in *Figure* 3.3, has three planes, the input plane P1, the Fourier plane P2, and the correlation plane P3. Two converging lenses (L1 and L2) are used to pass from one plane to the other. The first lens realizes the FT of the target image. The second realize the FT^{-1} of the multiplication result (spectrum of target image * the correlation filter).

The flowchart of the VLC technique is presented as follows: firstly, a Fourier transform is applied to the target image to get a target spectrum. After that, a multiplication between the target spectrum and the correlation filter obtain with the Fourier transform of a reference image is affected, and place this result in the Fourier plane. Next, it provides the correlation result is recorded on the correlation plane, this multiplication effected by inverse Fourier transform.

The correlation result, described by the peak intensity that uses to determine the similarity degree between the target and reference images, *Figure* 3.3 illustrate the VLC technique. To evaluate the correlation, the Peak to Correlation Energy (PCE) defined as the energy in the correlation peaks intensity normalized to the overall energy of the correlation plane.

$$PCE = \frac{\sum_{i,j}^{N} E_{peak}(i,j)}{\sum_{i,j}^{M} E_{corralation_{plane}}(i,j)} \tag{3.7}$$

where M and N are the size of correlation plane and the size of the peak correlation spot, respectively.

To enhance the matching process Horner and Gianino [24] proposed a phase-only filter (POF). The POF can produce correlation peaks marked with enhanced discrimination capability. The POF is an optimized filter defined as:

$$H_{POF}(u,v) = \frac{S^*(\mathrm{u},\mathrm{v})}{|S(\mathrm{u},\mathrm{v})|} \tag{3.8}$$

where $S^*(\mathrm{u},\mathrm{v})$ is the complex conjugate of the FT of the reference image.

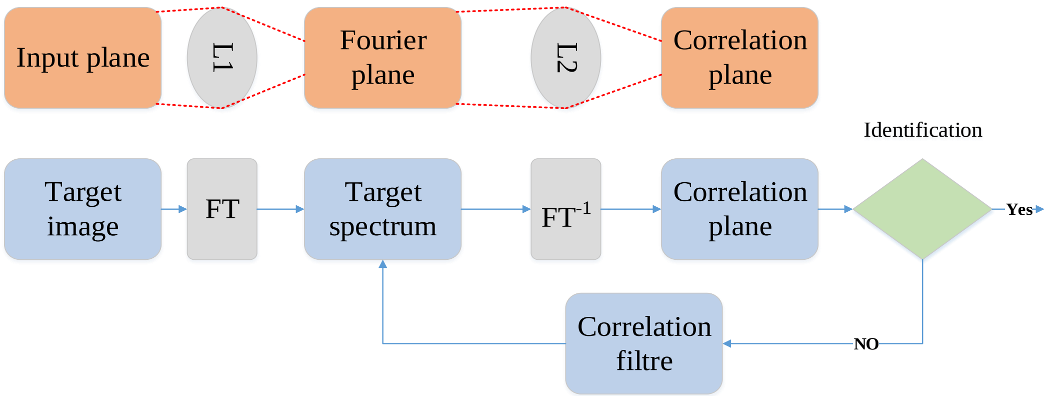

Figure 3.3: Vander Lugt Correlator

The correlation peak indicates clearly when the face image was in its corresponding face image (in *Figure* 3.4 (a)). Besides, the correlation peak has less outstanding, when the face image was correlated with other face image, peaks were less outstanding (in *Figure* 3.4 (b)).

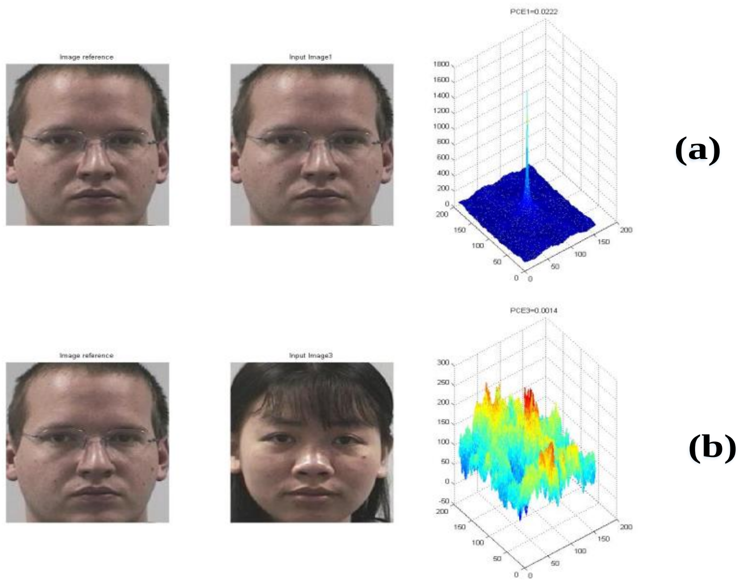

Figure 3.4: Results of the Correlation technique

Higher correlation peaks were associated with lower noise levels and smaller ones with higher noise levels.

3.4. Scale Invariant Feature Transform (SIFT)

The scale-invariant feature transform (SIFT) descriptor is presented by [25]. This descriptor establish the scale space of images and extract the local extrema with the Difference-of-Gaussian (DoG) function. This function is used to detect the interest points in the face image. The difference of Gaussian (DoG) in scale space is applied in order to detect the extrema and obtain the feature points (local extrema of the DoG) that are invariant to orientation and scale. The DOG function is the difference between two Gaussian kernels with different scales, it is defined as:

$$\begin{aligned} D(x\,y) &= (G(x,\ y,\ k\sigma) - (G(x\,y,\sigma)) * I(\ x\,y) \\ &= L(x,\ y,\ k\sigma) - L(\ x,\ y,\sigma\ ) \end{aligned} \tag{3.9}$$

with

$$G(x,\ y,\ k\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where I (x, y) is the gray level of pixel (x, y), σis the scale space factor, k is the ratio factor between two neighboring scales, G (x , y, σ) is Gaussian kernel function, and L (x, y, σ) is the image scale space.

Then, a feature vector is extracted, at each interest point. Then, to provide invariance against rotation, the local orientation of the image is estimated using the local image properties over a neighborhood and, over a number of scales around the point of interest. Next, based on local image information at the characteristic scale, a descriptor is computed for each detected point. In addition, this method performed a histogram of gradient orientations of sample points in a region around the interesting point, determines the interest points have the highest orientation value to constructs a feature vector and uses these directions as the dominant direction of the interesting point.

Typically, by put the key point in the center, an adjacent 16x16 region is performed. SIFT divides this region into 4×4 sub-regions with 8 orientation bins in each. The vector obtained by SIFT descriptor has 128 elements since there are 4 x 4 = 16 histograms each with 8 bins. Thus, the meaningful descriptors are extracted from the image that are yet robust, highly distinctive, and compact to change in camera viewpoint and illumination. *Figure* 3.5 shows the representation of the SIFT descriptor for a 16 × 16 pixel patch and a 4 × 4 descriptor array.

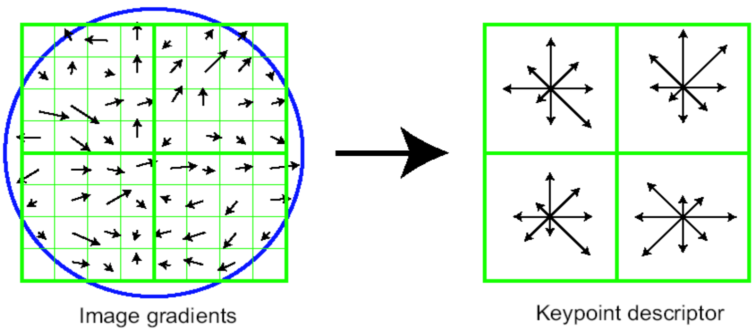

Figure 3.5: SIFT descriptor for a 16 × 16 pixel patch and a 4 × 4 descriptor array

3.5. Speeded Up Robust Features (SURF)

SURF is a technique used generically for the detection of features, but also used for the description. This descriptor has been constructed on the basis of on Scale Invariant Feature Transform (SIFT) descriptor. Also, the SURF it is very robust, and several times faster in fetching the results than SIFT. SURF descriptor has been used mainly to extract points of interest, to track objects, to locate and recognize faces or people, and to reconstruct 3D scenes. Also, this descriptor is invariant to scale and rotation. Its algorithm outputs a descriptor vector as well and it finds key point locations.

The SURF descriptor has three main steps include interest point detection, local neighborhood description, and matching. Initially, the SURF descriptor used an intermediate representation of the face image, to improve the speed of the SURF algorithm. This representation called "Integral Image" as shown in *Figure* 3.6, it is contains the sum values of the grayscale pixel of the face image. Its representation reduces a computational time advantage as it only depends on three integer operations.

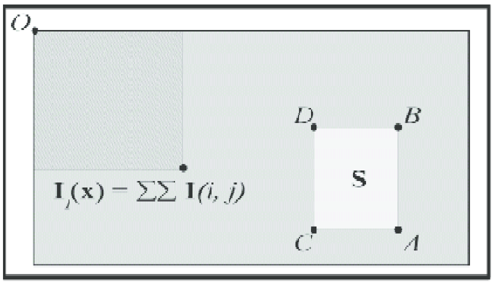

Figure 3.6: Evaluation of Integral Image

When the integral image is used, to compute the surface integral from the face image, it is necessary to read only four-pixel values so as (Jan et al., 2009). Next, to detect the interesting point, SURF employs the determinant of Hessian blob detector. In the image I, x = (x, y) is the given point, the Hessian matrix H(x, σ) in x at scale σ, it can be defined as (e.g. equation (3.10)):

$$H(x, y) = \det \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \qquad (3.10)$$

The approximated determinant of the Hessian describes the blob response in the image. To approximate the image partial derivatives, SURF utilizes a second order Gaussian filter, since this filter allows both space and scale analysis. The responses to Haar wavelets are used for orientation assignment, before the interest point descriptor is formed from the wavelet responses in a given surrounding interest point neighborhood. Finally, the similarity matching based on FLANN distance (Euclidean distance) is used to obtain the recognition results for the SIFT and the SURF techniques.

3.6. Binary Robust Independent Elementary Features (BRIEF)

Binary robust independent elementary features (BRIEF), is belongs to the family of binary descriptors include BRISK and LBP, which only performs simple binary comparison test and uses Mahalanobis or Hamming distance instead of Euclidean distance. Thus, it is only necessary to compare the intensity between two-pixel positions located around the detected interest points for building a binary descriptor. This allows obtaining a representative description at very low computational cost. Also, to represent an image patch as a binary string, this descriptor relies on a relatively small number of intensity difference tests. Generally, a binary descriptor for a patch of pixels of size S × S is performed by concatenating the results of the following test:

$$\tau = \begin{cases} 1 & if\ I(P_j) \succ I(P_i) \\ 0 & otherwise \end{cases} \tag{3.11}$$

where $I(P_i)$ I(pi) denotes the (smoothed) pixel intensity value at P_i, and the selection of the location of all the P_i uniquely defines a set of binary tests. In addition, the number of binary pixel pairs and the binary threshold are two setting parameters used with BRIEF descriptor. Also, this descriptor is very efficient both to compute and to store in memory. However, it does not provide rotation invariance. Finally, in this study we use the BRIEF descriptor with the Hamming distance for face recognition.

We present the software results of the six algorithms of feature extraction for the three approaches: the VLC for the correlation approaches and the LBP, HOG, SIFT, SURF, and BRIEF for the local approaches. The results of our implement methods for face recognition are shows in *Figure* 3.7.

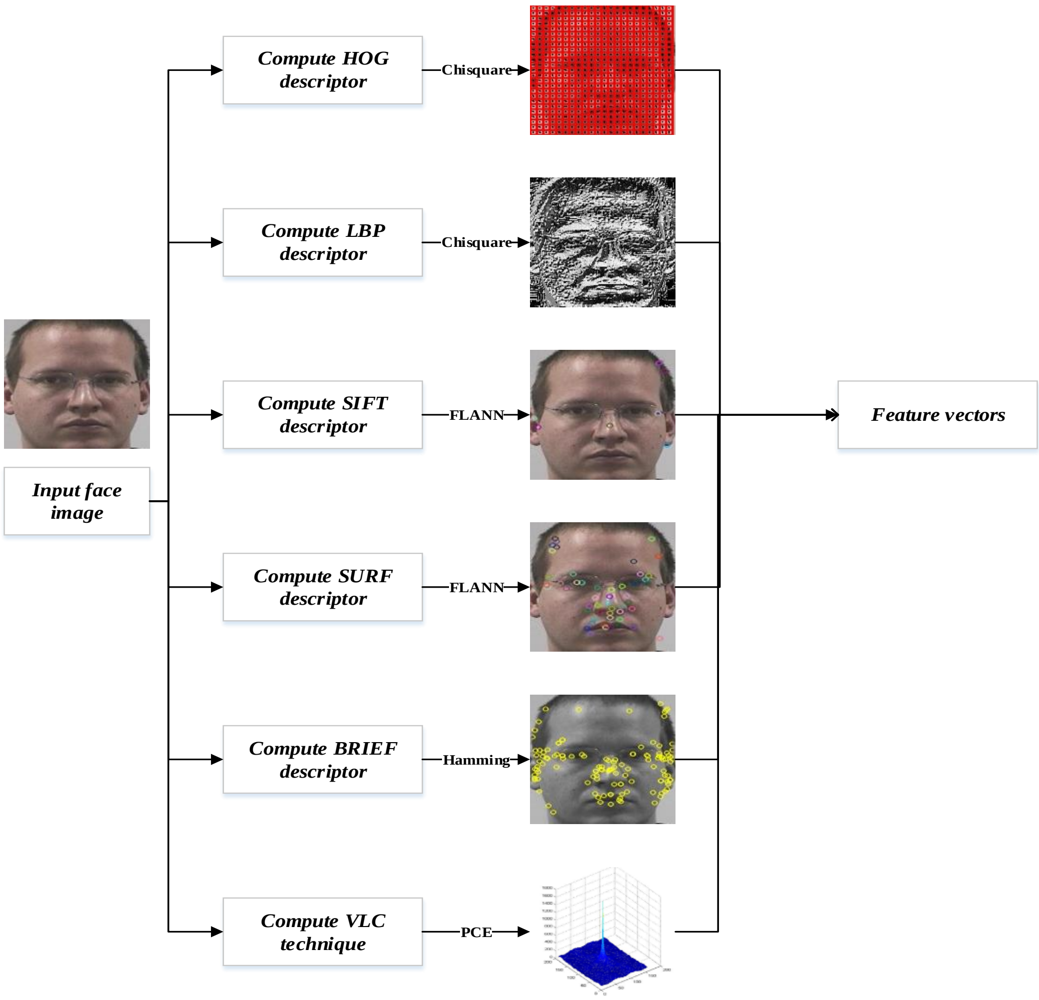

Figure 3.7: The software results of our implement methods for face recognition

4. Results Analysis

We implemented the proposed methods in C++ and OpenCV for testing a wide variety of images of faces. These experiments have been successfully evaluated on Intel PC platform equipped with a CPU Core i3 3.2-GHz.

4.1.Pointing Head Pose Image Database (PHPID)

In order to evaluate the performance of the six algorithms, we use the Pointing Head Pose Image Database (PHPID)[27]. In addition, in face recognition research community, the PHPID is one of the most used databases. It is contains 2790 monocular face images of 15 sets of persons with tilt angles from -90 to +90 degrees and variations of pan. Every set contains two series of 93 different poses (93 images) of the same person. In this study, we have two series per person in order to be able to train and test our algorithms on unknown and known faces. The face images in the database PHPID have various skin color and wear glasses or not. In addition, this database differs in the degree of illumination, expression, and variation of pose present in the face images. The *Figure* 4.1 shown some faces images used for evaluation.

Figure 4.1: Examples of PHPID faces images used in our tests

4.2.Experimental settings

Before the evaluation of the different methods, a preprocessing step is performed. In order to extract the region of interest of the position of the face, a geometrical normalization of the faces is used to resize the image to 154x188. The preprocessing step is performed in order to reduce the local shadowing, highlights, and effect of the illumination variation.

For the HOG and LBP methods, we use the same size of the region of interest locate with applying the grayscale conversion and normalization steps. The Chi-Squared distance is applied for recognition. Concerning VLC method, we use the same sized window of the region of interest to extract the correlation result and the peak-to-correlation energy (PCE) for recognition.

For the SIFT, SURF, and BRIEF methods, we used the same descriptors (SIFT, SURF, and BRIEF) to detect and describe the feature vectors and we performed the FLANN distance (SIFT and SURF) and Hamming distance (BRIEF) for recognition.

4.3.Evaluation techniques

In order to evaluate the performance of these six techniques for face recognition, two major evaluation plots are used the PR (Precision (P) and Recall) curve and the ROC (receiver operating characteristics) curve. The PR curve determines the relation between the detection precision and the detection rate (recall), while the ROC curve determines the relation between the True Positive Rate (TPR) and the False Positive Rate (FPR). In our study, we need two class recognition case (i.e., true or false face). The true positive means the true positive to be accepted by the technique, while the false positive means the false face to be accepted as true face. The detection rate and recall have the same meaning as the term true positive. In addition, we summarize a detailed description for evaluation in *Table 4.1*. However, we use a specific threshold for a given case to decide the positive or negative result for recognition. We could achieve high TPR and FPR, and vice versa with low threshold. Then, in this study, we use specific thresholds to represent each point on the ROC curve.

Table 4.1: The definition of parameters used in this study

Recognition	*Accepted (positive)*	*Rejected (negative)*
Desired face	True positive (TP)	False negative (FN)
Un desired face	False positive (FP)	True negative (TN)

The definition of the TPR, the FPR and precision are:

$$TPR = Recall = TP / (TP + FN) \tag{4.1}$$

$$FPR = FP / (FP + TN) \tag{4.2}$$

$$P = TP / (TP + FP) \tag{4.3}$$

4.4.Results and discussion

In this study, three steps are presented for each method: preprocessing, feature extraction using the LBP, HOG, and VLC, SIFT, SURF, and BRIEF, and classify with Chi-Squared, FLANN, and Hamming distances or with PCE. Thus, in order to demonstrate the effectiveness and feasibility best method used for face recognition, we discuss and compare the six face recognition algorithms in terms of recognition, implementation, and application under various conditions and environments.

The PHPID dataset provides trustees characteristics points such as mouth positions, nose, and eyes. We use the eye labels for crop the face images and resized these to the size of 154×188 pixels. In addition, we applied a gray scale conversion as a preprocessing in order to alleviate a large illumination variation contained in PHPID dataset. We compute the TPR and FPR by changing the threshold values to draw the ROC curves for 30 faces without rotation. We use two subset containing 30 face of 15 persons with zero-degree rotation under various illumination conditions. The mean values of FPR and TPR of the PHPID dataset with fixed thresholds over all the two subsets for performance evaluation in *Table 4.2*.

Table 4.2: The results under the different threshold for 30 faces (2 desired face and 28 undesired face) for various techniques.

Methods/ parameters	TP	FN	FP	TN	TPR	FPR	P
VLC	25	5	34	386	0.83	0.91	0.42
LBP	23	7	31	389	0.76	0.92	0.42
HOG	20	10	130	290	0.66	0.69	0.13
SIFT	28	2	9	411	0.93	0.97	0.75
SURF	23	7	10	410	0.76	0.97	0.69
BRIEF	26	4	21	399	0.86	0.95	0.55

Besides, we use the PHPID dataset with rotation (vertical and horizontal) 45° to 45° in this case study, where the dataset is divided into three subsets for validation, with two subsets containing 42 pairs of genuine matches and 42 pairs of impostor matches of the same person and the other subset containing 42 pairs of another person for verification. The mean values of FPR and TPR of the PHPID dataset with rotation with fixed thresholds over all the three subsets are shown in *Table 4.3-4.4-4.5*.

Table 4.3: The results under different threshold for 126 faces (84 desired face and 42 undesired face) for VLC and BRIEF techniques

VLC ***(T=PCE)***								**BRIEF** ***(T= Hamming distance)***							
T	**TP**	**FN**	**FP**	**TN**	**TPR**	**FPR**	**P**	**T**	**TP**	**FN**	**FP**	**TN**	**TPR**	**FPR**	**P**
T=0.0010	84	0	42	0	1	1	0.65	**T=160**	79	5	40	2	0.94	0.95	0.66
T=0.0012	83	1	42	0	0.98	1	0.65	**T=150**	66	18	37	5	0.78	0.88	0.64
T=0.0014	75	9	39	3	0.89	0.92	0.65	**T=140**	51	33	31	11	0.65	0.73	0.62
T=0.0016	68	16	25	17	0.80	0.59	0.73	**T=130**	45	39	28	14	0.53	0.66	0.61
T=0.0018	56	28	13	29	0.66	0.30	0.81	**T=120**	42	42	27	15	0.5	0.64	0.60
T=0.0020	53	31	4	38	0.63	0.09	0.92	**T=110**	40	44	27	15	0.47	0.64	0.59
T=0.0022	31	53	3	39	0.36	0.07	0.91	**T=100**	36	48	27	15	0.42	0.64	0.57
T=0.0024	23	61	3	39	0.27	0.07	0.88	**T=90**	36	48	27	15	0.42	0.64	0.57
T=0.0026	18	66	3	39	0.21	0.07	0.85	**T=80**	36	48	26	16	0.42	0.61	0.58
T=0.0028	13	71	2	40	0.15	0.04	0.86	**T=70**	34	50	26	16	0.40	0.61	0.56

Table 4.4: The results under different threshold for 126 faces (84 desired face and 42 undesired face) for SIFT and SURF techniques

SIFT ***(T= FLANN distance)***								**SURF** ***(T= FLANN distance)***							
T	**TP**	**FN**	**FP**	**TN**	**TPR**	**FPR**	**P**	**T**	**TP**	**FN**	**FP**	**TN**	**TPR**	**FPR**	**P**
T=450	74	10	41	1	0.88	0.97	0.64	**T=0.4**	78	6	38	4	0.92	0.90	0.67
T=440	64	20	38	4	0.76	0.90	0.62	**T=0.3**	66	18	31	11	0.78	0.73	0.68
T=430	45	39	33	9	0.53	0.72	0.57	**T=0.2**	53	31	23	19	0.63	0.52	0.69
T=420	32	52	25	17	0.38	0.59	0.56	**T=0.1**	39	45	18	24	0.46	0.42	0.68
T=410	21	63	19	23	0.25	0.45	0.52	**T=0.09**	38	46	15	27	0.45	0.35	0.71
T=400	18	66	14	28	0.21	0.33	0.56	**T=0.08**	36	48	13	29	0.42	0.30	0.73
T=390	14	70	10	32	0.16	0.23	0.58	**T=0.07**	35	49	13	29	0.41	0.30	0.72
T=380	11	73	7	35	0.13	0.16	0.61	**T=0.05**	27	57	12	30	0.32	0.28	0.69
T=370	10	74	3	39	0.11	0.07	0.76	**T=0.02**	13	71	5	37	0.15	0.11	0.72
83/T=360	9	75	2	40	0.10	0.04	0.81	**T=0.005**	2	82	1	41	0.023	0.02	0.66

To evaluate our techniques implement for face recognition, we plot the ROC curve using TPR and FPR for performance evaluation in *Figure 4.2*.

When reading a ROC curve, the points farther away from the diagonal line (y = x) and closer to closer to the point of coordinates (0, 1) are better, as that indicate lower FPR (fewer incorrect classifications) and higher TPR (more correct classifications).

Table 4.5: The results under different threshold for 126 faces (84 desired face and 42 undesired face) for LPB and HOG techniques

LBP ***(T= Chi-Square distance)***								**HOG** ***(T= Chi-Square distance)***							
T	**TP**	**FN**	**FP**	**TN**	**TPR**	**FPR**	**P**	**T**	**TP**	**FN**	**FP**	**TN**	**TPR**	**FPR**	**P**
T=1.3	83	1	41	1	0.98	0.97	0.66	**T=2500**	61	23	4	38	0.72	0.09	0.93
T=1.2	83	1	39	3	0.98	0.92	0.68	**T=2400**	61	23	4	38	0.72	0.09	0.93
T=1.1	80	4	30	12	0.95	0.71	0.72	**T=2300**	57	27	3	39	0.67	0.07	0.95
T=1	70	14	21	21	0.83	0.5	0.76	**T=2200**	48	36	2	40	0.57	0.04	0.96
T=0.9	59	25	14	28	0.70	0.33	0.80	**T=2100**	43	41	1	41	0.51	0.02	0.97
T=0.8	47	37	5	37	0.55	0.11	0.90	**T=2000**	39	45	0	42	0.46	0	1
T=0.7	32	52	3	39	0.38	0.07	0.91	**T=1900**	34	50	0	42	0.40	0	1
T=0.6	24	60	0	42	0.28	0	1	**T=1800**	30	54	0	42	0.35	0	1
T=0.5	8	76	0	42	0.09	0	1	**T=1700**	25	59	0	42	0.29	0	1
T=0.4	3	81	0	42	0.03	0	1	**T=1600**	23	61	0	42	0.27	0	1

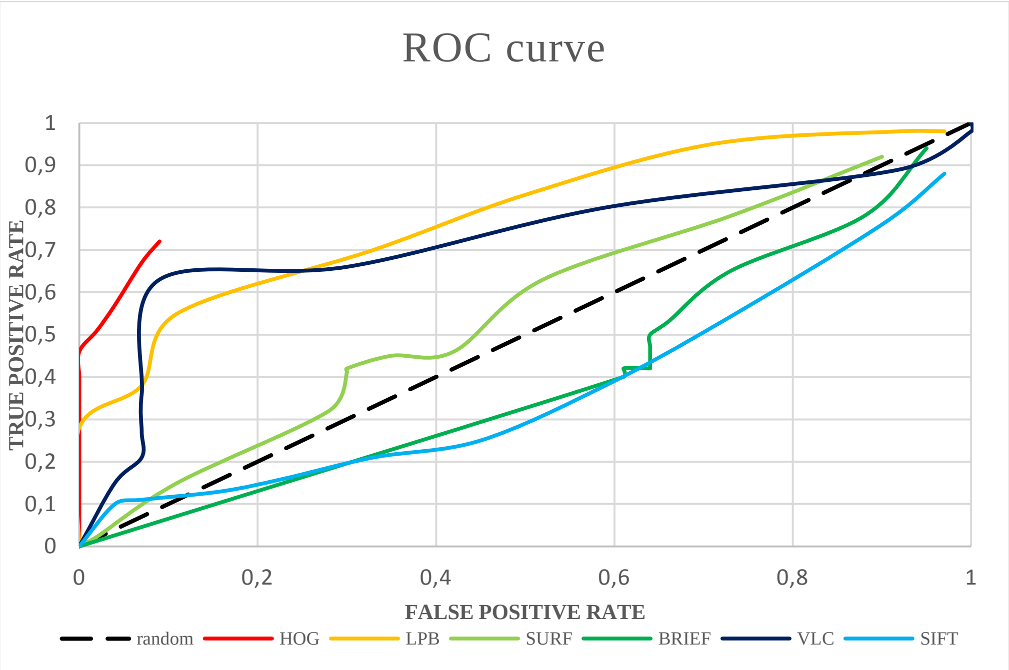

Figure 4.2: The ROC curves using TPR and FPR, the vertical axis shows the true-positive rate TPR and the horizontal axis show the false-positive rate FPR

We calculated the TPR and FPR by varying the threshold values to plot the ROC curves. The ROC curve for the six techniques shows that LBP and VLC outperforms HOG, SIFT, SURF, and BRIEF in face recognition task.

More specifically, VLC achieves 61 % TPR at 46 % FPR, LBP achieves 59 % TPR at 41% FPR whereas HOG, SIFT, SURF and BRIEF obtain 55 %, 40%, 48% and 58% TPR at 9 %, 52 %, 44 % and 95 % FPR respectively, Table 4.6. The *Figure 4.2* shows ROC curves for the six methods, the VLC and LBP achieve better performance than HOG, SIFT, SURF, and BRIEF.

Table 4.6: The average values of TPR, FPR, and P for different techniques.

Parameters / Methods	TPR	FPR	P
VLC	0.61	0.46	0.75
LBP	0.59	0.41	0.80
HOG	0.51	0.09	0.89
SIFT	0.40	0.52	0.63
SURF	0.48	0.44	0.69
BRIEF	0.58	0.95	0.72

4.5.Computational time

In this section, we compare the complexity of the most widely used methods for face recognition: HOG, LBP, VLC, SIFT, SURF and BRIEF descriptors. The face recognition times for different methods as shown in *Table 4.7*.

Table 4.7: Face recognition times for different methods.

Methods	HOG	LBP	VLC	SIFT	SURF	BRIEF
Times(seconds)	0.039	0.099	0.046	0.41	0.116	0.304

For the computational complexity of the six many methods implemented in this study inside the identification (or recognition) phase, the VLC which is based on the fast Fourier transform (FFT) and HOG which is calculate histograms by using the angle orientation of the edges are the less complex methods, as compared to the other methods include LBP which is a simple thresholding and encoding, and the SIFT, SURF, BRIEF which consists of two steps such as key points detection description.

Considering the VLC method, the LBP requires a computational complexity which is 2 times greater (the calculation of VLC features and the construction of correlation plane is very fast when compared to the compute the a simple thresholding and the binary code to get the feature vectors), the SIFT, SURF, BRIEF requires a computational complexity which is 4 times greater, 2 times greater, and 3 times greater, respectively (the detection and the description of the interest points is the most complex processes due to the techniques used in these descriptors (e.g., DOG, wavelet transform, Harr filter, Histogram of gradients). Also, the HOG and the VLC methods probably have the same time, but the VLC method outperforms in recognition rates on the PHPID database.

5. Conclusion and future work

This chapter establish a comparative study about face recognition of various vision algorithms used in face recognition with feature mining. The changes in angle and lighting make the recognition of the faces very complex, which often differ in subtle ways. Therefore, the face recognition system should take serval elements includes illumination change, pose change, facial expression change, mustache, beard, mask in front, occlusion due to the scarf and scaling factor. Experiments on the PHPID dataset, a challenging database of real-world human faces, demonstrated that the implemented methods yield the best results in the various conditions. Since the VLC and LPB conveys more efficiency and discriminative power than other descriptors like HOG, SURF, SIFT and BRIEF, and for feature extraction does not involve any quantization in the computation of descriptors. The results clearly show that VLC and LBP outperformed the other studied methods on all the sebsets of PHPID dataset. Finally, we conclude in this work that

VLC and LBP are the best methods in terms of implementation and application in real time in the future work.

6. Reference

[1] Alfalou, A., Brosseau, C., Katz, P., & Alam, M. S. (2012). Decision optimization for face recognition based on an alternate correlation plane quantification metric. *Optics letters*, *37*(9), 1562-1564.

[2] Elbouz, M., Bouzidi, F., Alfalou, A., Brosseau, C., Leonard, I., & Benkelfat, B. E. (2013, April). Adapted all-numerical correlator for face recognition applications. In Proc. of SPIE Vol (Vol. 8748, pp. 874807-1).

[3] Lu, X., Duan, X., Mao, X., Li, Y., & Zhang, X. (2017). Feature Extraction and Fusion Using Deep Convolutional Neural Networks for Face Detection. *Mathematical Problems in Engineering*, *2017*.

[4] Solanki, D. V., & Kothari, A. M. (2014). A Survey on Face Recognition Techniques and Its Applications. Journal of Image Processing & Pattern Recognition Progress, 1(3), 16-21.

[5] Abhishree, T. M., Latha, J., Manikantan, K., & Ramachandran, S. (2015). Face recognition using Gabor filter based feature extraction with Anisotropic Diffusion as a Pre-processing Technique. *procedia Computer Science*, *45*, 312-321.

[6] Rahim, M. A., Azam, M. S., Hossain, N., & Islam, M. R. (2013). Face recognition using local binary patterns (LBP). *Global Journal of Computer Science and Technology*.

[7] Huang, C., & Huang, J. (2017). A Fast HOG Descriptor Using Lookup Table and Integral Image. *arXiv preprint arXiv:1703.06256*.

[8] Feng, Y., An, X., & Liu, X. (2016, October). The application of scale invariant feature transform fused with shape model in the human face recognition. In *Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016 IEEE* (pp. 1716-1720). IEEE.

[9] Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. Computer Vision–ECCV 2010, 778-792.

[10] R Verma, R., & Kaur, R. (2014). Enhanced Character Recognition Using Surf Feature and Neural Network Technique. *IJCSIT) International Journal of Computer Science and Information Technologies*, *5*(4), 5565-5570.

[11] Singh, N. A., Kumar, M. B., & Bala, M. C. (2016). Face Recognition System based on SURF

and LDA Technique. *International Journal of Intelligent Systems and Applications*, *8*(2), 13.

[12] Vishwakarma, N., 2016. Face Recognition System Using Principal Component Analysis (PCA) 5, 1759–1766.

[13] Liu, Y., Pados, D. A., & Yeh, C. H. (2016, April). Two-stage Tensor Locality-Preserving Projection Face Recognition. In *Multimedia Big Data (BigMM), 2016 IEEE Second International Conference on* (pp. 214-218). IEEE.

[14] Alfalou, A., & Brosseau, C. (2010). Understanding correlation techniques for face recognition: from basics to applications. In *Face Recognition*. InTech.

[15] Wang, Q., Alfalou, A., & Brosseau, C. (2017). New perspectives in face correlation research: a tutorial. *Advances in Optics and Photonics*, *9*(1), 1-78.

[16] Král, P., & Vrba, A. (2017). Enhanced Local Binary Patterns for Automatic Face Recognition. *arXiv preprint arXiv:1702.03349*.

[17] Ren, J., Jiang, X., & Yuan, J. (2015). A chi-squared-transformed subspace of LBP histogram for visual recognition. *IEEE Transactions on Image Processing*, *24*(6), 1893-1904.

[18] RAMI, H., HAMRI, M., & MASMOUDI, L. (2014). Tracking Object Using Exponential Forgetting Factor In The LBP Code. *Journal of Theoretical and Applied Information Technology*, *63*(3).

[19] Du, G., Su, F., & Cai, A. (2009, October). Face recognition using SURF features. In *Proc. of SPIE Vol* (Vol. 7496, pp. 749628-1).

[20] Işık, Ş. (2014). A comparative evaluation of well-known feature detectors and descriptors. *International Journal of Applied Mathematics, Electronics and Computers*, *3*(1), 1-6.

[21] Zhang, J., Marszalek, M., Lazebnik, S., & Schmid, C. (2005). *Local features and kernels for classifcation of texture and object categories: An indepth study*. Technical Report RR-5737, INRIA Rhône-Alpes.

[22] Butt, J. A., & Wilkinson, T. D. (2006). Binary phase only filters for rotation and scale invariant pattern recognition with the joint transform correlator. Optics communications, 262(1), 17-26.

[23] Leonard, I., Alfalou, A., & Brosseau, C. (2012). Spectral optimized asymmetric segmented phase-only correlation filter. *Applied optics*, *51*(14), 2638-2650.

[24] Horner, J. L., & Gianino, P. D. (1984). Phase-only matched filtering. *Applied optics*, *23*(6), 812-816.

[25] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

[26] Svab, J., Krajnik, T., Faigl, J., & Preucil, L. (2009, November). FPGA based speeded up robust features. In *Technologies for Practical Robot Applications, 2009. TePRA 2009. IEEE International Conference on* (pp. 35-41). IEEE.

[27] Gourier, N., Hall, D., & Crowley, J. L. (2004, August). Estimating face orientation from robust detection of salient facial structures. In *FG Net Workshop on Visual Observation of Deictic Gestures* (Vol. 6).