Human Visual System applied to Facial Recognition

Fatima Zohra Allam, Latifa Hamami-Mitiche, and Hicham Bousbia-Salah

Abstract – Research in the field of biometrics is constantly expanding. Numerous studies have been carried out to develop different techniques with the aim of ensuring reliable and efficient recognition systems. Of our five senses, vision takes up the most of the neurons in our brain. This makes the visual approach challenging. This paper proposes to develop a new facial recognition technique based on the concept of Human Visual System in order to simulate or imitate human perception. Our approach exploits the behavior of the Human Visual System in biometric systems to improve individual recognition. It focuses on Perceptual Channel Decomposition in order to generate images, confined around a certain frequency range, perfectly uncorrelated. For the extraction of characteristic vectors, our technique uses the Principal Component Analysis. The principle of the visual approach consists of exploiting one or more characteristics of the peripheral parts of the Human Visual System. These characteristics can integrate the sensitivity of the Human Visual System to spatial frequencies, its sensitivity to local contrast. The Principal Component Analysis uses a set of seventeen channels, each adjusted to a band of given radial frequencies and orientation. The seventeen output images contain the same spatial information but are perfectly uncorrelated from a spectral point of view. In the implementation phase, we limited ourselves to explore this technique by using only four frequency rings. The results obtained are conclusive and satisfactory.

Keywords – Human Visual System, Perceptual Channel Decomposition, Principal Component Analysis.

I. INTRODUCTION

With the multiplication of applications in the biometric field, person recognition has attracted great attention in the scientific community, to the point that it is difficult to draw up an exhaustive assessment of all the work in this field. This is explained, on the one hand, by the enormous application interest, among others, in the field of security, video surveillance and general public multimedia applications, on the other hand, by the challenge that this represents for artificial vision algorithms. Biometrics brings together all computer techniques aimed at irrefutably determining the identity of a person based on their physical and/or behavioral characteristics, without having to use a card or password. Over the past decades, researchers have focused on developing automatic authentication systems capable of combating fraud to ensure security in different areas ranging from the most emblematic, such as crossing international borders, to the less burdensome, such as access to personal information. To meet these needs, biometrics, which is experiencing rapid development, appears to be a practical and effective solution. It makes it possible to measure the characteristics of people in order to characterize and authenticate them. It has established itself as a preferred technology in the field of recognition. This enthusiasm leads to the development of very varied biometric methods: from the most classic, such as facial recognition, the study of fingerprints, to the most exotic, such as gait recognition or recognition of the shape of the ears. Today, biometrics is establishing itself in the eyes of States as a security solution par excellence. Facial recognition is one of the most used biometric techniques. We notice that in daily life each of us identifies, throughout the day, different faces. So, when we meet a person, our brain will search our memory and check whether that person is listed or not. The facial modality is non-intrusive because it does not affect the privacy of the individual. It constitutes one of the most natural ways to recognize a person, unlike the iris which is certainly more intrusive. In this article, we propose an innovative alternative to create a biometric recognition system, different from those already developed in the literature. Traditional methods are based on the use of algorithms dedicated to the extraction of relevant characteristics such as the Gabor Filter, the Local Binary Pattern, and Zernike Moments. Our approach exploits the behavior of the Human Visual System. We are interested in processing spectral information rather than using the gray levels of the acquired image. This is a more elaborate model on image perception exploiting the behavior of the Human Visual System (HVS). The SVH coherently models key functions of perception such as Perceptual Channel Decomposition (PCD), Contrast Sensitivity Functions (CSF), etc. In our work, we favor approaches that seek to “simulate” human perception. Instead of using the gray levels of the facial image, these approaches exploit, through a decomposition into visual frequency sub-bands, characteristics similar to those used by the Human Visual System, such as Local Band-Limited Contrast (LBLC). Our article is structured as follows: In the second section, we introduce some common concepts about biometrics and biometric systems. We also present a state of the art on facial recognition. In the third section, we present another alternative to realize the biometric system for person recognition based on the processing of spectral information. This is a more elaborate model on image perception exploiting the behavior of the Human Visual System (HVS). The SVH coherently models key functions of perception such as Perceptual Channel Decomposition (PCD), Contrast Sensitivity Functions (CSF), etc. In this chapter, we favor approaches that seek to “simulate” human perception.

Manuscript received October 25, 2023; revised June 27, 2024.

F. Z. Allam, L. Hamami-Mitiche, and H. Bousbia-Salah are with Electronics Department, Ecole Nationale Polytechnique, Algiers, ALGERIA. (e-mail: mourad.haddadi@g.enp.edu.dz).

Digital Object Identifier (DOI): 10.53907/enpesj.v4i1.247

2716-912X © 2024 Ecole Nationale Polytechnique
Instead of using the gray levels of the facial image, these approaches exploit, through a decomposition into visual frequency sub-bands, characteristics similar to those used by the Human Visual System, such as Local Band-Limited Contrast (LBLC). In the fourth section, we explain the steps we followed for the implementation of the facial recognition system. The fifth section indicates the results obtained.

II. BIOMETRIC CONCEPTS

In this section, we present some common notions about biometrics:

1. Biometric Processes

Any biometric system operates in enrollment, verification and identification mode [1][2].

A. Enrollment Mode

The enrollment mode is a learning phase which aims to collect biometric information on the people to be identified (Figure 1). During this phase, the biometric characteristics of individuals are captured by a biometric sensor, then represented in digital form (signatures), and finally stored in the database [2].

![Fig.1: Block diagram of a Biometric System in Enrollment Mode](image)

B. Verification Mode

The verification or authentication mode is a one-to-one (1:1) comparison, in which the system validates the identity of a person by comparing the biometric data entered with the biometric template of that person stored in the system database [2]. Figure 2 illustrates the block diagram of a biometric system in Verification Mode.

![Fig.2: Block diagram of a Biometric System in Verification Mode](image)

C. Identification Mode

In this mode, the biometric characteristics of the person to be identified are acquired and compared to those of all customers known to the system and stored in a database (Figure 3). This is a one-to-many (1:N) comparison. Identity identification is typically used to prevent a person from using multiple identities [2].

![Fig.3: Block diagram of a Biometric System in Verification Mode](image)

2. Architecture of Biometric Systems

The architecture of a biometric system varies depending on the biometric characteristics used, application constraints and security requirements. However, these basic components are, generally, present in most biometric systems [3]. Figure 4 illustrates the Architecture of biometric systems.

![Fig.4: Architecture of biometric systems](image)

A biometric system is essentially a pattern recognition system that acquires biometric data from an individual, then compares a set of characteristics extracted from this data to a set of previously stored data to finally be able to execute an action or take a decision based on the result of this comparison [3][4].

A. Acquisition Phase

This step is done using various devices like cameras, fingerprint scanners, retina readers, etc. These devices capture the specific biometric characteristics of the individual. The quality of captured data is essential to ensure the accuracy and reliability of the biometric identification process.

B. Feature extraction Phase

This phase takes as input the data acquired by the capture module and extracts only the information in order to form a new representation of the data. Ideally, this new representation is supposed to be unique for each person and invariant to intra-class variations.

C. Matching Phase

The extracted biometric features are compared with the templates stored in the database as vectors, assessing their degree of similarity to that vector, measured in terms of difference or distance.

D. Decision Phase

The decision strategy plays a crucial role in the authentication of individuals. To do this, the user's biometric characteristics are extracted and compared with the vectors stored in the database. This comparison is made by measuring the degree of similarity between the extracted biometric characteristics and the stored vector, generally, in the form of difference or distance. In this phase, it is possible to verify the identity declared by a user or to determine the identity of an unknown person based on the degree of similarity between their biometric characteristics and those stored in the database.

3. State of the Art on Facial Recognition

Human beings generally find it easy to identify a person from their face. However, for machines, automatic face recognition poses a significant challenge. This problem has generated considerable research in recent years. Facial recognition is an image processing technology used to identify or verify the identity of an individual based on unique facial characteristics.
It uses algorithms to analyze facial characteristics, such as distance between eyes, nose shape, eye contours, and lip and mouth size, to create a unique digital signature for each individual.

Facial recognition is used in a variety of applications, including security, marketing, surveillance, human resource management, criminal research, etc. It can enable rapid and efficient identification of individuals, as well as real-time monitoring of activities.

4. Steps of Facial Recognition

In this section, we will explain the main steps of facial recognition [5]:

A. Acquisition
The image of the face is acquired using a camera or other imaging device.

B. Processing
The image is processed and normalized to remove unwanted distortions and noise.

C. Face Detection
The system uses face detection algorithms to detect the position and contours of the face in the image.

D. Feature Extraction
Facial features such as distance between eyes, size and shape of eyes, nose and mouth are extracted from the image.

E. Comparison
Extracted facial features are compared to those stored in a database of known faces to find a match.

F. Decision Making
Based on the match found, the system makes a decision on the identity of the person. In some applications, there may be additional steps, such as machine learning to improve recognition performance and verification of the person's identity using biometric methods.

5. Principal Component Analysis (PCA)

PCA is a very popular method of finding the principal components of faces from a set of example images. Each face can then be described by a linear combination of these eigenvectors. Although PCA is fast, simple and popular in pattern identification, it is not optimized for class separability (discrimination). This multivariate analysis technique can be used for facial recognition. The most used algorithms in the PCA method are summarized as follows [6]:

- Center the data: The facial images are first centered by subtracting the average of all the images.
- Calculate the covariance matrix using the centered data. It represents the relationship between different facial features.
- Calculate the principal components by diagonalizing the covariance matrix. These are vectors that represent the directions in which the data has the most variance.
- Reduce dimensionality: The principal components that have the highest eigenvalues are selected because they contain the most information about the data. These components are used to reduce the dimensionality of the data space.
- Project the data onto the new dimensions defined by the selected principal components. This step helps reduce data complexity and improve facial recognition performance.
- Classify the images by comparing their projection on the new dimensions with the projections of all other images in the database. Images that have similar projections are considered the same and can be used to identify a person.

III. STATE OF THE ART ON THE HUMAN VISUAL SYSTEM

Research on the human body has always aroused great interest, in particular, to understand its architecture and its interactions. The HVS has been studied in depth to discover the cells and neurons that compose it, in order to better understand the system as a whole and to better treat certain pathologies [7].

The HVS is a complex and sophisticated system that allows us to perceive and interpret visual information. Advances in imaging technology have made it possible to better understand how HVS works and to apply it in image processing. The latter uses computer techniques to enhance images and extract information from them [8][9]. These techniques often take inspiration from human visual perception to improve their accuracy and efficiency. Image segmentation involves dividing an image into different regions or objects and segmentation algorithms often take inspiration from human visual perception to detect object boundaries and contours. The ability of our visual system to process visual information quickly and efficiently is incredibly powerful. Technological advances have made it possible to develop artificial vision systems that attempt to imitate this ability to solve complex problems in various fields, such as image recognition, video surveillance, autonomous navigation, robotics. HVS is able to detect subtle features in images, such as shapes, textures, colors and movements. This capability can be used for practical applications such as facial recognition in security systems. In summary, the motivation to apply HVS is mainly due to its power, ease of use, and ability to process complex visual information.

A. Modeling of the Human Visual System

HVS modeling is an interdisciplinary area of research that aims to understand how humans visually perceive and interpret their environment. Advances in computing and Artificial Intelligence technologies have made it possible to develop increasingly sophisticated models to describe visual processes and associated behaviors [7][8]. HVS models are generally based on insights from psychology, neuroscience and computer science. They use data obtained from psychophysical experiments, behavioral studies, brain images and other sources to create mathematical and computational models that replicate human visual perception [9][10].

An example of a well-known model is the Marr model, which describes the SVH in three levels of representation

- At the primary level, visual information is processed in the primary visual cortex from the retina, using algorithms such as spatial and temporal filtering, edge detection, orientation selectivity and luminance perception.
- At the 2.5D representation level, visual information is processed to create a two-dimensional representation of the visual scene, taking depth information into account. Algorithms used at this level include stereovision, motion disparity, textures and contours.

- At the 3D representation level, visual information is used to create a three-dimensional representation of the visual scene. The algorithms used at this level include depth perception, 3D scene reconstruction, scene segmentation into objects and object recognition.

Although this model is a simplification of the actual complexity of the HVS, these three levels are interconnected and work together to produce visual perception.

B. Luminance perception

The HVS is capable of perceiving a wide range of luminance, from a few thousandths of candelas per square meter (cd/m²) under night vision conditions (Scotopic) to several thousand candelas per square meter under bright light conditions (Photopic) [11][12].

Luminance perception involves the complex interaction between the physical properties of light, the properties of photoreceptor cells in the retina, and the visual information processing mechanisms in the brain that create a subjective perception of brightness and contrast in the visual environment.

C. Definition of Contrast

Contrast is one of the most studied parameters in psychovisual experiments and also the most exploited in image quality measurements. Visual contrast is defined as the relative difference in luminance between two regions of the image, divided by the average luminance of these two regions [9][13].

Contrast is an important parameter in the study of visual perception and image quality. Contrast, generally used in HVS models, depends on several parameters such as shape, spatial frequency, orientation of the stimulus used and also the complexity of the neighborhood. Several definitions have been proposed since Michelson's but no universal definition is known to date (Figure 5).

1. Michelson Contrast

It is defined as the ratio of the difference in luminance between a light region and a dark region, divided by the average of the luminance of these two regions [9][14]. This definition was introduced by Michelson and is widely used in the fields of human vision, imaging and photography.

\[ C_M = \frac{L_{max} - L_{min}}{2L_{ave}} \]  

(1)

Where: \( L_{max} \) and \( L_{min} \) correspond to the maximum and minimum luminance values respectively.

2. Contrast by Weber Fechner

The response of the Human Visual System depends more on local variations of luminance \( \Delta L \) in relation to the surrounding luminance \( L \), than on absolute luminance values. This property is known under the Weber-Fechner law. It is one of the first quantitative models of luminance perception [9].

We perceive a minimum deviation \( \Delta L \) for a target of luminance \( L + \Delta L \) on a uniform background of luminance \( L \), called adaptation. The \( \Delta L/L \) ratio is practically constant over a wide range of brightness. Mathematically, the Weber contrast is modeled by the relation (2):

\[ C_w = \frac{\Delta L}{L} \]  

(2)

3. Peli Contrast

Decomposition into frequency channels makes it possible to take into account the properties of human vision, which is more sensitive to certain frequencies than others by decomposing the image into several frequency channels. Specific contrast treatments can be applied to each channel in order to optimize the contrast throughout the image [11][15].

Once the image has been decomposed into several frequency channels, the Peli contrast can be calculated for each channel individually using the formula given in (3). This contrast measurement can then be used to guide adaptive contrast treatments that take into account the properties of human vision.

\[ C_i(m, n) = \frac{L_i(m,n)}{\Sigma_{k=1}^{i-1} L_k(m,n)} \]  

(3)

Where : \( i \) represents the \( i^{th} \) channel radial

The denominator represents the local average luminance which corresponds to all channels of spectral support lower than that of the \( i^{th} \) channel.

D. Multi-Channel Organization

The sensitivity of different cells of the HVS to certain types of information, such as color, orientation or frequency, suggests that there is a prior organization of the information before its processing. The results of numerous psychophysical experiments support this idea and present the HVS as a multi-channel system [16].

This means that the visual system is organized into several channels, each specialized for processing a specific type of information. This organization allows faster and more precise perception of the visual environment, as well as more effective adaptation to complex and varied visual situations.

It is true that single-channel modeling is not sufficient to completely explain the behavior of the HVS in the face of these stimuli. Indeed, HVS uses a set of channels to analyze input signals, each sensitive to a specific orientation and spatial frequency.

These channels are separable in a polar representation, meaning they can be described by their amplitude and phase at different spatial orientations. The characteristics of these channels allow the HVS to process visual information selectively, responding only to stimuli that match the characteristics of the activated channels [9][11].

By combining information from different channels, the brain is able to reconstruct a complete image of the visual environment.
E. Perceptual Channels Decomposition

The decomposition into perceptual channels reflects the spatio-frequency selectivity of the HVS. It implements a division of the spatio-frequency plane as obtained by a Fourier transform.

This decomposition proposes a frequency tiling into seventeen channels. It is strongly inspired by Watson’s Cortex transform. However, the number and characteristics of the channels are different by dynamic division into radial frequencies. The angular selectivity varies as a function of the radial frequencies unlike the Watson decomposition. The DCP is based on a set of psycho-physical experiments.

Under standardized viewing conditions, there are four radial spatial frequency domains (LF, II, III, IV) [9]:

- The LF (Low Frequency) domain corresponds to spatial frequencies between 0 and 1.5 cpd (cycles per degree).
- Domain II corresponds to frequencies between 1.5 and 5.7 cpd.
- Domain III corresponds to frequencies between 5.7 and 14.2 cpd.
- Domain IV corresponds to frequencies between 14.2 and 28.2 cpd.

The angular selectivity depends on the spatial frequency domain considered as follows (Figure 6):

- For the LF domain, there is no angular selectivity.
- For domain II, the angular selectivity is 45°, which defines four oriented channels.
- For domains III and IV, the angular selectivity is 30°, which defines six oriented channels for each domain.

Under standardized viewing conditions, there are four radial spatial frequency domains (LF, II, III, IV) [9]:

A. Extended Yale Face Database B

The “Extended Yale Face Database B” is a collection of facial images used for facial recognition research. This database contains images of 133 people, each with 20 images of their face. The images were taken from different angles and with varying lighting conditions, allowing the effectiveness of facial recognition algorithms to be tested in real-world situations.

Using this database, we can test the algorithms’ ability to identify people from images of faces taken under different lighting conditions. The results can be used to improve the efficiency and accuracy of facial recognition systems, which is crucial in many security and surveillance applications.

B. Architecture of Face Recognition System based on SVH

Figure 7 illustrates the architecture of the facial recognition system based on SVH [17]:

1. Acquisition System

The first step to loading an image into Matlab is to use the "imread" function. This function allows us to read an image from a file stored on our computer and store it in a Matlab variable. We can then use other Matlab functions to process or display this image, such as "imshow" to display it on the screen [17].

2. Facial Pretreatment

The image is converted to grayscale using the “rgb2gray” function. Image noise is reduced by applying a Gaussian Filter. This helps improve the accuracy of face detection. A face detector is initialized using the Viola-Jones detection method, which is a popular technique for detecting faces in images. Bounding boxes of faces in the image are detected using the face detector.

Next, we initialize a “for” loop to extract each face from the image using the detected bounding boxes. In each iteration of the loop, the coordinates of the bounding box are stored in the variable “bbox”. The Region Of Interest (ROI) corresponding to the bounding box is extracted using the “imcrop” function.
Finally, the ROI is displayed using the “subplot” function for each face detected in the image [17].

3. Decomposition into Perceptual Channels (PCD)

To perform the decomposition into channels, we use the wavelet transform which allows us to decompose the image into different frequency bands, each containing specific information about the image. The “dwt2” function is used to perform this wavelet decomposition of the “img” image using the ‘db6’ wavelet family (Daubechies 6) with four domains.

The four domains correspond to the wavelet decomposition coefficients of the image, each having a different angular selectivity. The decomposition coefficients are divided into four matrices [17]:

- **LL1**: Matrix of approximation coefficients (Low-Low) which contains the low frequency information of the image. This matrix represents the image at a lower scale with lower resolution.
- **HL1**: Horizontal detail coefficient matrix (High-Low) which contains high frequency information in the horizontal direction of the image.
- **LH1**: Vertical detail coefficient matrix (Low-High) which contains high frequency information in the vertical direction of the image.
- **HH1**: Matrix of diagonal detail coefficients (High-High) which contains the high frequency information in the diagonal direction of the image.

Figure 8 represents Perceptual Channels Decomposition of a Face Image [17]:

4. Feature Extraction

After decomposing the image into four domains of different selectivity, we applied the HOG method to extract features using different sizes of feature extraction cells [17].

5. Feature Vector

After feature extraction using HOG method, we can apply Principal Component Analysis (PCA) method for feature selection and dimension reduction of feature vectors.

6. Classifier

The classifier is an essential element in learning. It is an algorithm or model that learns to classify or predict data based on input characteristics or variables. In our work, we use the SVM (Support Vector Machine) classifier to perform the classification of our data. The SVM is a linear or non-linear classifier which seeks to find the optimal hyperplane to separate the different classes in the feature space.

7. Database used

In our work, we use a database of 30 classes. The database is divided into two parts:

- a training database
- a test database.

The learning base is used to train the system; it uses the database. The test database uses the remaining 30 of the database to test the system [17].

Here are the predictions of our classifier
V. RESULTS AND EVALUATION

To evaluate and see the effect of our HVS based face recognition system, we first need to create a validation baseline for our system. We will then check if our system is able to identify the face images that have been provided to it and that have already been used to train the system through an SVM classifier [17].

Modifies the database structure as follows:
- 10% for validation
- 20% for testing
- 70% for training

To clearly visualize the Confusion Matrix, we use 15 classes with 20 images each (Figure 11).

![Fig.11: MSVM Classifier predictions (15 classes)](image)

Figure 12 illustrates the confusion matrix for 4 classes.

![Fig.12: MSVM Classifier predictions (4 classes)](image)

Figures 13 and 14 show the confusion matrices for 15 and 4 classes respectively [17].

VI. CONCLUSION

Biometrics is a field that is both exciting and complex and the number of research is growing universally. Currently, the use of biometric systems such as facial or fingerprint recognition are essential in high security environments because they make it possible to recognize or verify the identity of people, with a significant degree of reliability. In this paper, we are interested in developing a robust and efficient face-based biometric system. Beforehand, we introduced some general concepts about biometrics. We have seen that a typical biometric system can be represented by four main modules, namely, the capture module, the feature extraction module, the matching module and the decision module. Subsequently, we presented a state of the art on facial recognition. We followed a well-defined structure by dividing the database into three distinct parts: training, testing and validation. In our approach, we applied SVH to decompose images into different frequency scales. Then, we extracted the features using the Histogram of Oriented Gradients (HOG) method. This step allowed us to represent each image by a vector of significant characteristics. To reduce the dimensionality of the data and improve the efficiency of the model, we applied principal component analysis (PCA). This technique allowed us to select the most discriminating components and use them to represent faces in a reduced-dimensional space. Finally, we used a classifier to perform face prediction. Different classification methods can be used; such as support vector machines (SVM). Our face recognition system based on the PCA method gave good results in terms of accuracy and performance. We obtained an accuracy of 100%.
These results are encouraging and very satisfactory and show the effectiveness of the algorithms developed for the extraction of discriminating characteristics. This implementation brightens the horizon for carrying out other work exploiting other biometric modalities by combining them. It is worth noting that this project opens so many doors to explore. In this sense, we can consider as future perspectives to make possible improvements which are mainly summarized as follows:

- Test the capacity and performance of the approach developed on other databases.
- Apply other techniques to facial recognition, other than global approaches.
- Create a biometric system on FPGA type electronic components to respect space constraints and real-time processing.
- Make the biometric system portable thanks to improving the performance of FPGA programmable circuits.

REFERENCES


Fatima Zohra Allam-Chergui (M. Sc 2006, Doctorate 2023) is an Assistant Professor at the Ecole Nationale Supérieure des ICT et de la Poste. She is the author of several International Publications. His research focus is based on Image Processing, more precisely, Biometric Images. Its objectives are multiple, among others, the implementation of Biometric Systems on embedded circuits of the FPGA type, representation and projection of signals from the time domain to the frequency domain with the aim of exploiting the contribution of spectral information rather than spatial.

Latifa Hamami-Mitiche (M.Sc. 1988, Ph.D. 2002) is a full Professor in the Department of Electronics at Ecole Nationale Polytechnique of Algiers, Algeria. She is author more than 150 scientific papers. Her research interests include Medical Image Processing, ISAR Image Processing, Optical Character Recognition, Biometrics and Embedded System, FPGA and ASIC Design, System on Chip.

Hicham Bousbia-Salah is currently Associate Professor, Researcher in Signal Processing at Ecole Nationale Polytechnique, Algiers. He earned his Engineering degree in Electronics at Ecole Nationale Polytechnique, Algiers in 1981, his Master degree in Computer Engineering at Wayne State University in USA in 1984 and his PhD in Electronics in Signal Processing from the Ecole Nationale Polytechnique, Algiers in 2006. Currently, he is interested in the field of blind separation of multiple sources from convolutive mixtures and its implementation in signal processing.