STATUS OF THESIS

Title of thesis	BIOLOGICALLY INSPIRED OBJECT RECOGNITION SYSTEM

I HAMADA RASHEED HASSAN AL-ABSI

hereby allow my thesis to be placed at the Information Resource Centre (IRC) of Universiti Teknologi PETRONAS (UTP) with the following conditions:

1. The thesis becomes the property of UTP
2. The IRC of UTP may make copies of the thesis for academic purposes only.
3. This thesis is classified as

☐ Confidential

☑ Non-confidential

If this thesis is confidential, please state the reason:

____

____

____

The contents of the thesis will remain confidential for ___________ years.

Remarks on disclosure:

____

____

____

Endorsed by

______________________	______________________
Signature of Author	Signature of Supervisor
Hamada Rasheed Hassan Al-Absi	Assoc. Prof. Dr. Azween Abdullah
CIS Department	CIS Department
Universiti Teknologi PETRONAS	Universiti Teknologi PETRONAS
Bandar Iskandar, 31750 Trohoh	Bandar Iskandar, 31750 Trohoh
Perak, Malaysia	Perak, Malaysia
Date: ______________	Date: ______________

UNIVERSITI TEKNOLOGI PETRONAS

BIOLOGICALLY INSPIRED OBJECT RECOGNITION SYSTEM

by

HAMADA RASHEED HASSAN AL-ABSI

The undersigned certify that they have read, and recommend to the Postgraduate Studies Programme for acceptance this thesis for the fulfillment of the requirements for the degree stated.

Signature:	____
Main Supervisor:	Assoc. Prof. Dr. Azween Abdullah
Signature:	____
Head of Department:	Dr. Mohd Fadzil Bin Hassan
Date:	____

BIOLOGICALLY INSPIRED OBJECT RECOGNITION SYSTEM

by

HAMADA RASHEED HASSAN AL-ABSI

A Thesis

Submitted to the Postgraduate Studies Programme

As a Requirement for the Degree of

MASTER OF SCIENCE

DEPARTMENT OF COMPUTER & INFORMATION SCIENCES

UNIVERSITI TEKNOLOGI PETRONAS

BANDAR SERI ISKANDAR,

PERAK

SEPTEMBER 2010

DECLARATION OF THESIS

Title of thesis: BIOLOGICALLY INSPIRED OBJECT RECOGNITION SYSTEM

I HAMADA RASHEED HASSAN AL-ABSI

hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTP or other institutions.

Witnessed by

__________________________	__________________________
Signature of Author	Signature of Supervisor
CIS Department Universiti Teknologi PETRONAS Bandar Iskandar, 31750 Trohoh Perak, Malaysia	Name of Supervisor Assoc. Prof. Dr. Azween Abdullah CIS Department Universiti Teknologi PETRONAS Bandar Iskandar, 31750 Trohoh Perak, Malaysia
Date: ____________________	Date: ____________________

To My Family and the Memory of My Grandmother

ACKNOWLEDGEMENT

First and foremost, I would like to thank God the Almighty, for without His consent, it would be impossible to achieve what has been done in this work , for giving me the strength and determination to keep going even during the most difficult moments. May Allah accept this work, counts it as a good deed and make it useful.

I would like to express my utmost gratitude to my supervisor Assoc. Prof. Dr. Azween B. Abdullah for his constant guidance and support; he has guided, motivated, and advised me all times.

I would like to thank Universiti Teknologi PETRONAS for supporting this work by providing the Graduation Assistantship Scheme, the staff of the Computer & Information Sciences department and the postgraduate office for their support.

I would like to be grateful to My Parents, Brothers, Sisters and Everyone in my family who have supported me all the times during my study in Malaysia.

A special gratitude goes to Dr. Yasir Abdelgadir and Dr. Mahamat Issa Hassan for their advice, support and beneficial discussions. My Regards goes to everyone who has supported me to complete this thesis especially to the HISH group members.

“Credit is hereby given to the Massachusetts Institute of Technology and to the Center for Biological and Computational Learning for providing the database of facial images”.

Credit is also given to University of Essex for providing the face94 dataset of facial images.

ABSTRACT

Object Recognition has been a field of interest to many researchers. In fact, it has been referred to as the most important problem in machine or computer vision. Researchers have developed many algorithms to solve the problem of object recognition that are machine vision motivated. On the other hand, biology has motivated researchers to study the visual system of humans and animals such as monkeys and map it into a computational model. Some of these models are based on the feed-forward mechanism of information communication in cortex where the information is communicated between the different visual areas from the lower areas to the top areas in a feed-forward manner; however, the performance of these models has been affected much by the increase of clutter in the scene as well as occlusion. Another mechanism of information processing in the cortex is called the feedback mechanism, where the information from the top areas in the visual system is communicated to the lower areas in a feedback manner; this mechanism has also been mapped into computational models. All these models which are based on the feed-forward or feedback mechanisms have shown promising results. However, during the testing of these models, there have been some issues that affect their performance such as occlusion that prevents objects from being visible. In addition, scenes that contain high amounts of clutter in them, where there are so many objects, have also affected the performance of these models. In fact, the performance has been reported to drop to 74% when systems that are based on these models are subjected to one or both of the issues mentioned above. The human visual system, naturally, utilizes both feed-forward and feedback mechanisms in the operation of perceiving the surrounding environment. Both feed-forward and feedback mechanisms are integrated in a way that makes the visual system of the human outperforms any state-of-the-art system. In this research, a proposed model of object recognition based on the integration concept of the feed-forward and feedback mechanisms in the human visual system is presented.

ABSTRAK

Pengecaman objek telah menjadi sebuah bidang yang menarik kepada ramai penyelidik. Bahkan, ia telah dirujuk sebagai masalah terpenting dalam penglihatan mesin atau komputer. Para penyelidik telah membangunkan banyak algoritma untuk menyelesaikan masalah pengenalan objek yang dimotivasikan oleh penglihatan mesin. Di sudut yang lain, biologi telah memotivasikan para penyelidik untuk mengkaji system visual manusia dan haiwan seperti monyet dan memetakannya ke dalam model pengkomputeran. Sebahagian dari model-model ini adalah berasaskan mekanisma suap-depan komunikasi maklumat dalam korteks di mana maklumat disalurkan antara kawasan visual yang berlainan dari kawasan bawah ke kawasan atas menurut kaedah suap-depan; walau bagaimanapun, prestasi model-model ini telah banyak terjejas oleh peningkatan selerak di dalam pemandangan dan juga oklusi. Satu lagi mekanisma pemprosesan maklumat dalam korteks disebut sebagai mekanisma maklumbalas, di mana maklumat dari kawasan atas di dalam sistem visual tersebut disalurkan ke kawasan bawah menurut kaedah maklumbalas; mekanisma ini juga telah dipetakan ke dalam model pengkomputeran. Kesemua model ini yang berasaskan mekanisma suap-depan dan maklumbalas telah menunjukkan keputusan yang memberangsangkan. Bagaimana pun, semasa ujian terhadap model-model ini, terdapat beberapa isu yang menjejaskan prestasi mereka umpamanya oklusi yang menghalang objek dari dapat dilihat. Tambahan pula, pemandangan yang mempunyai kandungan selerak yang tinggi di dalamnya, di mana terdapat terlalu banyak objek, juga telah menjejaskan prestasi model-model ini. Bahkan, prestasi sistem telah dilapurkan menurun sehingga 74% apabila sistem-sistem yang berasaskan model-model ini didedahkan kepada satu atau kedua-dua isu yang disebutkan di atas. Sistem visual manusia, secara semulajadi, menggunakan kedua-dua mekanisma suap-depan dan maklumbalas dalam operasi memerhati keadaan sekeliling. Kedua-dua mekanisma suap-depan dan maklumbalas digabungkan dalam satu cara yang menjadikan sistem visual manusia mengatasi sebarang sistem terkini. Di dalam kajian ini, dikemukakan sebuah model yang telah dicadangkan mengenai pengenalan objek

berasaskan gabungan konsep mekanisma suap-depan dan maklumbalas di dalam sistem visual manusia. Model tersebut telah menunjukkan kebolehan mengenali objek contohnya wajah-wajah di dalam pemandangan kompleks seperti pemandangan yang berselerak dan pemandangan yang engandungi wajah-wajah yang sebahagiannya terselindung.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Identifying and recognizing objects in scenes have been one of the most famous research topics in machine/computer vision. Many research centers have been established around the globe with the goal of building and developing algorithms and techniques that can produce excellent results of object recognition. This interest in building applications with high recognition capabilities comes from the importance of object recognition in our lives. Object recognition has been employed in many applications that have high impact on the quality of life. Figure 1.1 shows an example of an object recognition system.

Although many algorithms have been developed to achieve high performance in recognizing objects, there are some issues and obstacles that affect the accuracy and robustness of these algorithms such as partially occluded objects, scenes with high clutter, objects with different shapes, variations in objects scales, orientation etc. (LeCun et al. 2004).

In order to overcome the aforementioned issues, computer scientists had to look for new methodologies that would facilitate to develop more robust systems. Therefore, and in line with the advance in neuroscience that led neuroscientist to understand the visual systems of cats (Hubel and Wiesel 1962), primates and finally humans, computer scientist introduced biological vision. This discipline refers to vision algorithms that have been inspired by the visual system of primates or humans (Louie 2003).

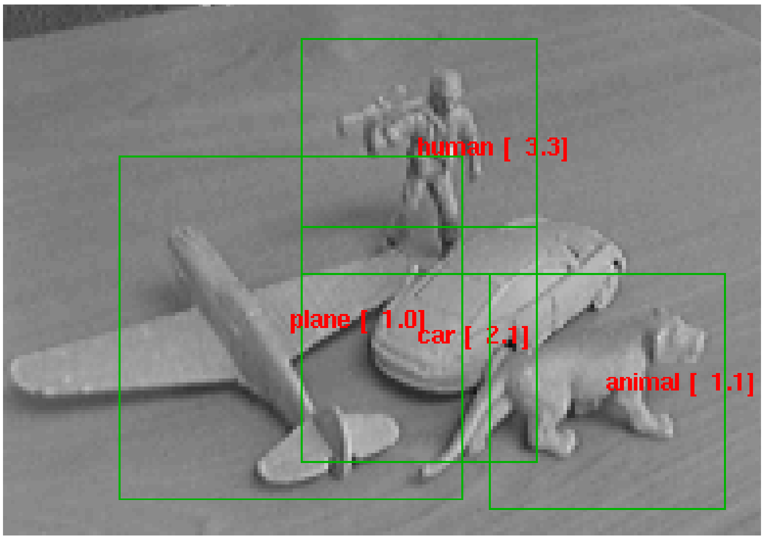


Figure 1.1: Example of an object recognition system[1]

Humans recognize different types of objects with ease and high accuracy. A person is able to recognize different types of objects around him/her such as the faces of relatives, different types of animals, differentiate car from others and so forth. The human visual system outperforms any state-of-the-art computer vision system in object recognition. The amazing ability of the human visual system has attracted neuroscientists to study this organ, try to understand how it works and identify its components and functionalities that contribute towards the evident performance.

(Hubel and Wiesel 1962) were the first to discover how information processing was done in the cat visual system. The discovery led to understand how the signals are communicated from the eye to the brain, and how the brain processes these signals in order to achieve recognition of objects. These discoveries were the first step towards understanding the visual system. After that, neuroscientist continued to investigate the visual systems of other animals such as monkeys before researchers started to involve the human visual system.

Biologically inspired system refers to systems that have been built with the inspiration of a natural living system (Bongard 2009) i.e. animals. In regard to object recognition, computer scientists have developed systems that are inspired by the visual

---

[1] http://www.lecun.com

systems of monkeys and humans. In fact, after the advances in neuroscience and the discoveries that led to the understanding of how the visual systems of monkey and human work, scientist utilized the information gained and developed object recognition systems based on the functions of the visual systems of monkeys at the beginning and then moved to mapping the functions of the humans visual system as well. In this research, a biologically inspired model based on the human visual systems is proposed in order to build a system that is capable of recognizing partially occluded object and objects that are in cluttered scenes.

1.2 Object Recognition Applications

The importance of object recognition is realized by looking at the applications that can be built with object recognition capabilities. The following are some applications of object recognition in different areas in our life.

1.2.1 Facial Recognition

In face recognition (Paliy et al. 2005), research has proved that facial recognition can solve or prevent many troubles. One of the applications of facial recognition is access control system (Bryliuk and Starovoitov 2002) that enables authorized personnel to access certain areas by identifying their faces, or even to access ones' computer (Figure 1.2). In addition to its ability of identifying faces in a robustness way, access control does not require fancy equipment in comparison to other access control methodologies which made it a cheap system. Face recognition can also be utilized in surveillance systems to identify criminals and make it easy to capture them.

Figure 1.2: Face recognition for access control[2]

1.2.2 Car License Plate Recognition

Object recognition technology has also been applied to car license plate recognition (Zheng and He 2006) and (Khalifa et al. 2007) (figure 1.3). This application is useful for the police to identify stolen cars, it is also used as an access authentication method to parking lots, security monitoring of road, and drive-through methodology to help in allowing customers to drive-through based on recognizing their car license plate.

Figure 1.3: Car license plate recognition[3]

[2] www.sharewareconnection.com

1.2.3 Object Recognition in Medical Applications

Object recognition technology has been applied in medical applications such as cancer recognition. In (Liu and Ma 2007) a breast cancer recognition system was proposed to detect early stages of the cancer so that it can be cured easily. The system provided a high detection rate; however, since breast cancer has many types, the system was not able to diagnose all of them. Another medical application is the lung cancer recognition system (Xia et al. 2006). Object recognition applications in medicine are increasing rapidly. Currently, it is being utilized as decision support systems to help doctors in diagnosing diseases that are difficult to be identified by the human eye. Figure 1.4 shows an image of a normal lung, and another image of a lung infected by a cancer that the system was able to recognize.

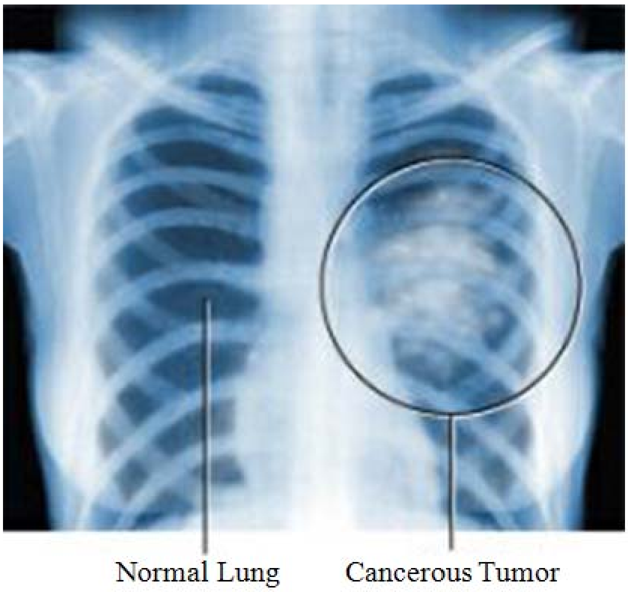


Figure 1.4: Object recognition in identifying lung cancer[4]

[3] www.plate-recognition.info
[4] www.hyscience.com

1.3 Problem Statement

Object recognition is a wide area in which researchers have developed many algorithms to achieve. Most of these algorithms are machine vision motivated. Biology has also motivated other researchers to come up with models that are inspired by the primates' visual system. However, by looking at the results of the aforementioned models, researchers are yet to come up with a model that can solve major problems in object recognition such as recognizing objects in cluttered scenes (Kreiman et al. 2007) and partially occluded objects.

1.4 Objectives

The specific objectives of the work can be summarized as follows:

1. Developing an object recognition model based on the human visual system.
2. Integrating the functions of feed-forward and feedback mechanisms in the human visual system in regard to recognizing objects.
3. Developing a prototype to test the features of the model and determine its robustness and efficiency.

1.5 Motivation

The discoveries in neuroscience that made the functionality of some parts of the brain, especially in the visual system, quiet understandable, motivated computer scientist to map these functionalities into computational models that mimic the way the human recognizes and categorizes objects. This research is an extension to those researches, and will introduce a new theory on the object recognition based on human and primate's visual system as well as develop a computational model.

In addition, object recognition has many applications in life. It can be used in face recognition (Bryliuk and Starovoitov 2002) and car number plate recognition (Khalifa et al. 2007). Developing a system that is robust and accurate that would be used in medicine to identify diseases such as cancer (Cahoon et al. 2000) could save lives.

1.6 Scope

This research will study the human visual system and develop a theory of object recognition based on the functions of the visual system in humans. After that, a model of object recognition will be developed based on the findings.

1.7 Research Approach

In order to develop the biologically inspired object recognition system, the following steps will be done:

1. Study the human visual system, its architecture, the visual areas and the function of each area that process the incoming signals from the retina. This step will give the understanding on how the human visual system operates and what are the processes that take place at each visual area in order to achieve the recognition of the captured objects.
2. Understand the feed-forward and feedback mechanisms that link the visual areas with each other. The output of this step is to identify the role of feed-forward and feedback mechanisms in passing the information from one visual area to another, the importance of each mechanism in achieving a more accurate result in object recognition, and the importance of integrating both processes in image understanding.
3. Develop the bio-inspired object recognition model based on the findings of steps 1 and 2.
4. Match the different processes of each component in the model with a corresponding algorithm.
5. Implement the model using the chosen algorithms using MATLAB programming. Then test it by applying it to an application domain.

1.8 Research Activities

The following research activities will be executed in order to achieve the objectives of this research:

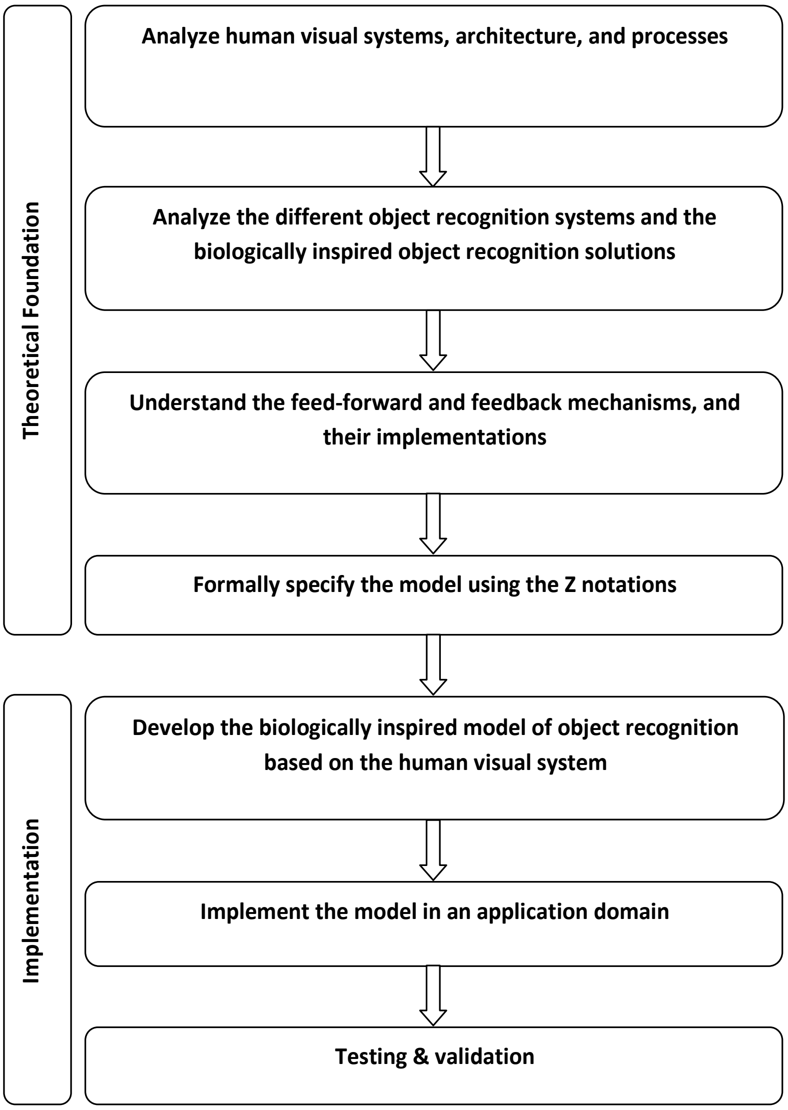


Figure 1.5: Research activities

1.9 Work Contributions

In this research, a new computational model of object recognition based on the human visual system is introduced. The model is based on the integration of the functions of the feed-forward and feedback mechanisms that connect the visual areas among each other. Previous work focused on the feed-forward mechanism and mapped its function into computational models. However, the performance of such models was affected by clutter and occlusion. To produce an object recognition system with human-like capabilities, both connection mechanisms should be integrated and that is the contribution of this research work.

1.10 Thesis Outline

The rest of this thesis is organized as follows: Chapter 2 provides a literature review on the biologically inspired recognition models especially on the feed-forward and feedback models that were mapped from the primates or humans visual systems. Chapter 3 introduce the methodology that has been followed in this research as well as the proposed model which is inspired by the human visual system and based on the integration of bottom-up (feed-forward) and top-down(feedback) functions in the visual cortex. Chapter 4 provides an application domain to test the proposed model which is a face recognition system. Chapter 5 presents an analysis on the results obtained in this research work. And finally, chapter 6 presents a conclusion of this research as well as some recommendations for future work.

.

CHAPTER 2

LITERATURE REVIEW

2.1 Object Recognition

Object Recognition has been a problem for both computer vision and biological vision. For many years, researchers have been developing different models and algorithms in order to achieve object recognition. Although there are so many techniques that have been developed, both computer vision and biological vision are still looking into building systems that can produce better results of object recognition. In this chapter, computer vision, biological vision and different algorithms / techniques that have been developed to achieve object recognition are discussed.

2.2 Computer Vision

The main aim of computer vision is developing intelligent applications that can understand the content of an image by extracting the information contained in it. Many algorithms are available which can achieve object recognition such as principal component analysis (PCA) that has been proven to perform well in recognizing objects such as faces (Aravind et al. 2002).

2.2.1 Feature Extraction

In computer vision, any system will start off by extracting features from the input image. This will help the classifier to decide on whether or not the intended object is in the scene. Many feature extraction algorithms are available such as Haar-like feature extraction algorithm (Wilson and Fernandez 2006) and Gabor filters (Ji et al. 2004) and (Zhang et al. 2007)

Haar-like features are one of the algorithms used to extract features from the image or input video (frames). These features use the change in contrast values between neighboring rectangular groups of pixels rather than using the intensity values of the pixel. Figure 2.1 shows the common Haar features.

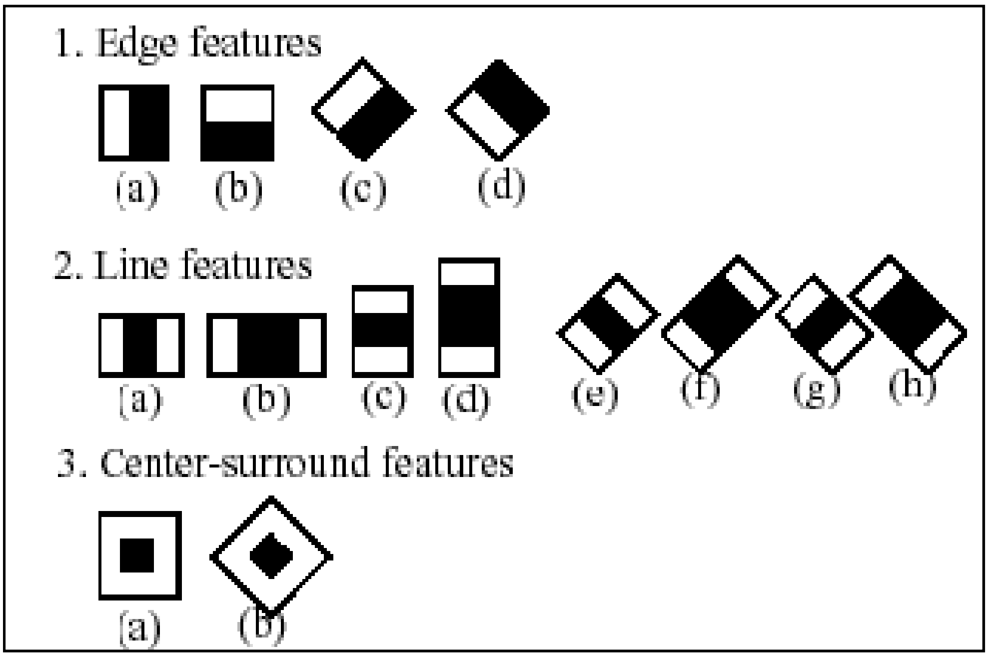


Figure 2.1 Common Haar-like features (Wilson and Fernandez 2006)

The simple rectangular features of the image can be calculated by using an intermediate representation called "*Integral Image*" (see equation 2.1) (Wilson and Fernandez 2006). This integral image is an array that contains the sum of the pixels'intensity values located to the left of a pixel and above the pixel at location (x,y).

If we assume that A[x,y] is the original image and AI[x,y] is the integral image then:

$$[x, y] = \sum_{x' \leq x, y' \leq y} (x', y') \qquad (2.1)$$

The computed feature value is then used as an input to a simple decision tree classifier that usually has tow nodes that can be represented as 1 or 0 (1 representing the existence of the object and 0 for the absence of the object). In fact, all features can be calculated in a fast constant time for any size for two auxiliary images (Lienhart et al. 2003).

Another set of features that are used to extract features are called Gabor filter (Ji et al. 2004; Zhang et al. 2007). This filter extracts features of different orientations and scales (figure 2.2). This filter has been used for edge detection and it has been proven to be sufficient (Ji et al. 2004).

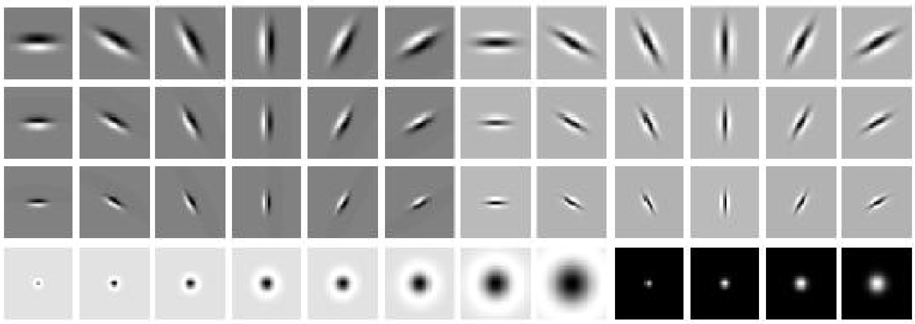


Figure 2.2: Gabor filter (Ji et al. 2004)

2.2.2 Principal Component Analysis

Principal Component Analysis (PCA) (Aravind et al. 2002; Smith 2002) is a statistical approach of identifying patterns in data and reforming the data in such a way as to express the similarities and differences. The approach guarantees dimensional reduction of original space without losing significant data characteristics as it is a powerful tool for data analysis. The approach is applied in number recognition/detection and it is reported to be one of the most robust, reliable and easily computed approaches.

Basically, PCA approach is to transform data into a new plane whereby patterns are more vividly emerged. This example illustrates the way PCA is performed on a set of data to show how clearly patterns emerge when data is transformed in the new plane. Ten sets of data in two dimensional planes are plotted and compared to the same set of data in the new plane (see figure 2.3). As seen in the new plane, data are divided by the line extending along the horizontal axis of the plane.

Table 2.1: Sample data to apply PCA

x	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
y	2.4	0.7	2.9	2.2	3	2.7	1.6	1.1	1.6	0.9

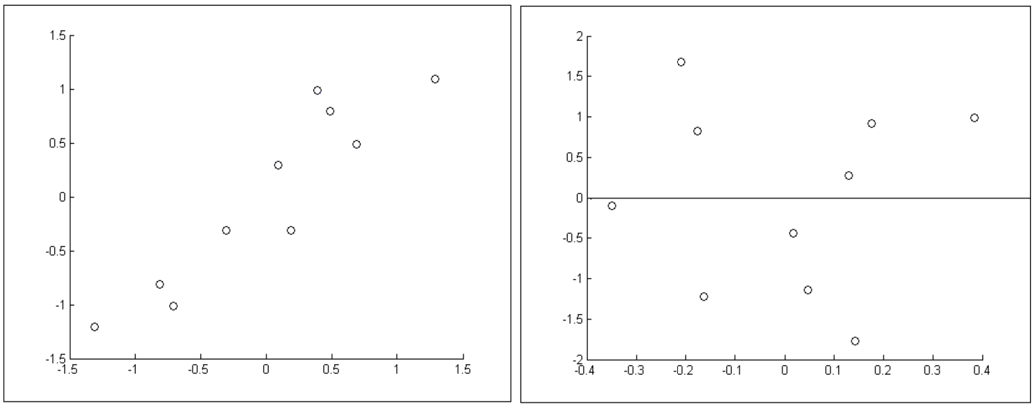


Figure 2.3: (Left) Data in a plane, (Right) Data in the new plane

2.2.3 Biological Vision

Biological vision is another technique that has recently been a topic of interest for many researchers. This discipline looks into the way the human or primate visual system works and maps it into a computational system. Human visual system outperforms any state-of-the-art systems in computer vision, therefore, researchers have been studying the way the information is processed in the visual system and tried to develop computational models. Most researchers have studied monkey’s brain and mapped its visual system functionalities into computational models; as the anatomy of monkey’s brain is similar to human’s brain (Tanaka 1997).

2.2.3.1 Feed-forward Models

Feed-forward models of object recognition are considered the most successful models that have been proven to be robust. It follows the feed-forward manner of information processing in the visual cortex. (Hubel and Wiesel 1962) were the first to discover how the visual system works in cats. They won the Nobel Prize in 1981 for their discovery. In 1999, (Riesenhuber and Poggio 1999) developed what is called the standard model of object recognition based on Hubel and Wiesel theory. (Riesenhuber and Poggio 1999; Riesenhuber and Poggio 2000) proposed a model based on (Hubel and Wiesel 1962) of simple cells to complex cells of the visual system. The model belongs to the feed-forward family that consists of hierarchical layers. Each layer has S units and C units. S units perform template matching of size and orientation. The outcome of the S unit is grouped and used as an input to C units that perform the MAX operation. The model in figure 2.4 is referred to as the standard model of object recognition in cortex.

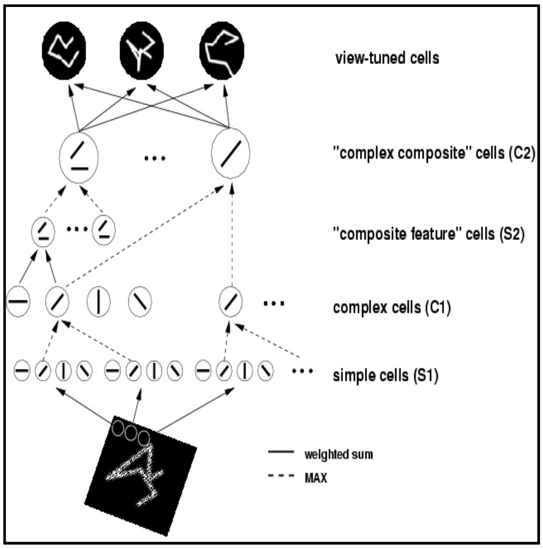


Figure2.4: Model of object recognition based on the feed-forward mechanism (Riesenhuber and Poggio 2000)

The simulation of the model has shown that the essential properties are robust. The results of the experiments on the model proved that it can be an extended model of the natural model proposed by (Hubel and Wiesel 1962).

(Serre et al. 2004; Serre et al. 2005) proposed a framework that introduces a set of features to ensure the robustness of object recognition; the proposed system is inspired by the standard model of object recognition (figure 2.4). According to their research, *"the computing of the features is done as follows:*

- *S1: Apply a battery of Gabor filters (figure 2.2) to the input image. The filter has 4 orientations and 16 scales which produces 64 maps. The 64 maps are arranged in 8 bands.*
- *C1: for each band, a max operation will be applied over each scale and position*
- *S2: compute Y for all image patches X at all positions to get S2 maps.*
- *C2: apply max operation over all patches to get shift and scale invariant features. Figure 2.5 illustrates the process of obtaining C2 features."*

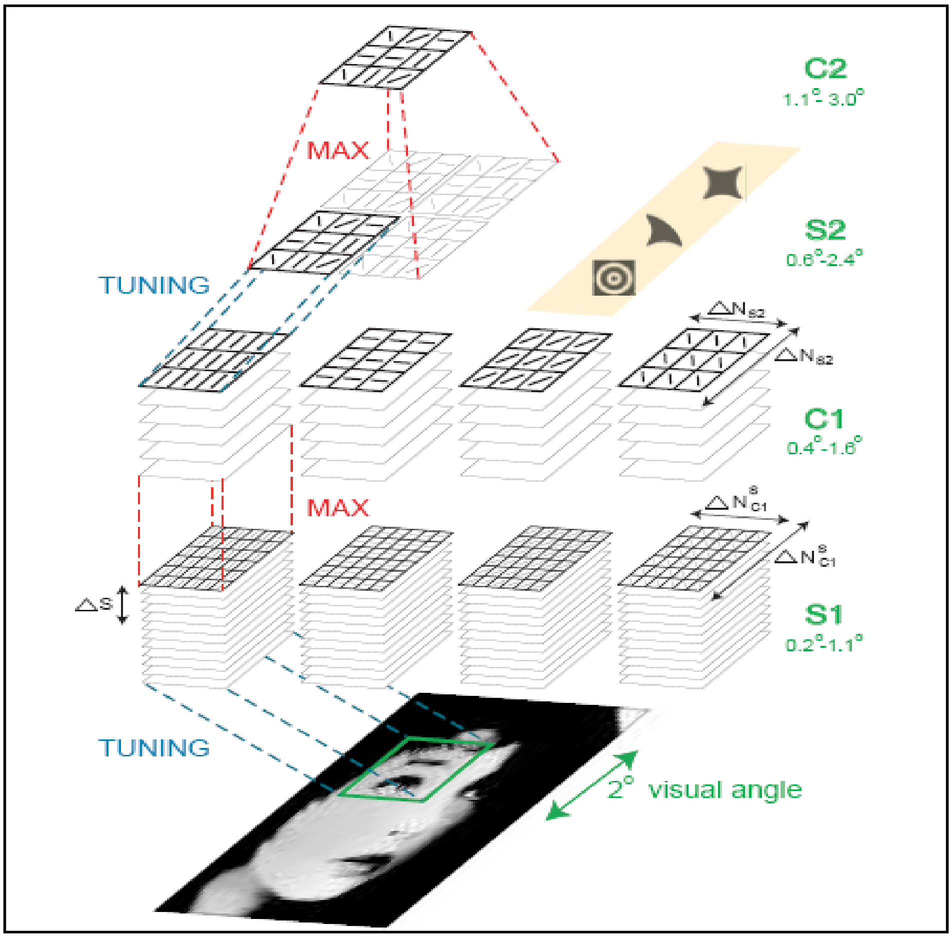


Figure 2.5: Obtaining C2 features (Serre 2006)

After the features have been obtained, the system runs a classifier on them. The results obtained from the system have been compared with other systems, and it shows that this feature provides consistent and better results than the other systems (Serre et al. 2005).

Another model, yet not different from the models in (Riesenhuber and Poggio 2000; Serre et al. 2007b; Serre et al. 2004; Serre et al. 2005) has been proposed by (Serre et al. 2007a) . It is also based on the standard model of object recognition and is part of the feed-forward family of models inspired by the visual cortex. Unlike the models proposed in (Serre et al. 2004; Serre et al. 2005) this model has more than 2 layers of the S and C cells which perform the same tasks as explained in (Serre et al. 2004; Serre et al. 2005) Figure 2.6 shows the new model. As shown in the figure, the model maps the information processing in the visual cortex (left) into a computational model (right).

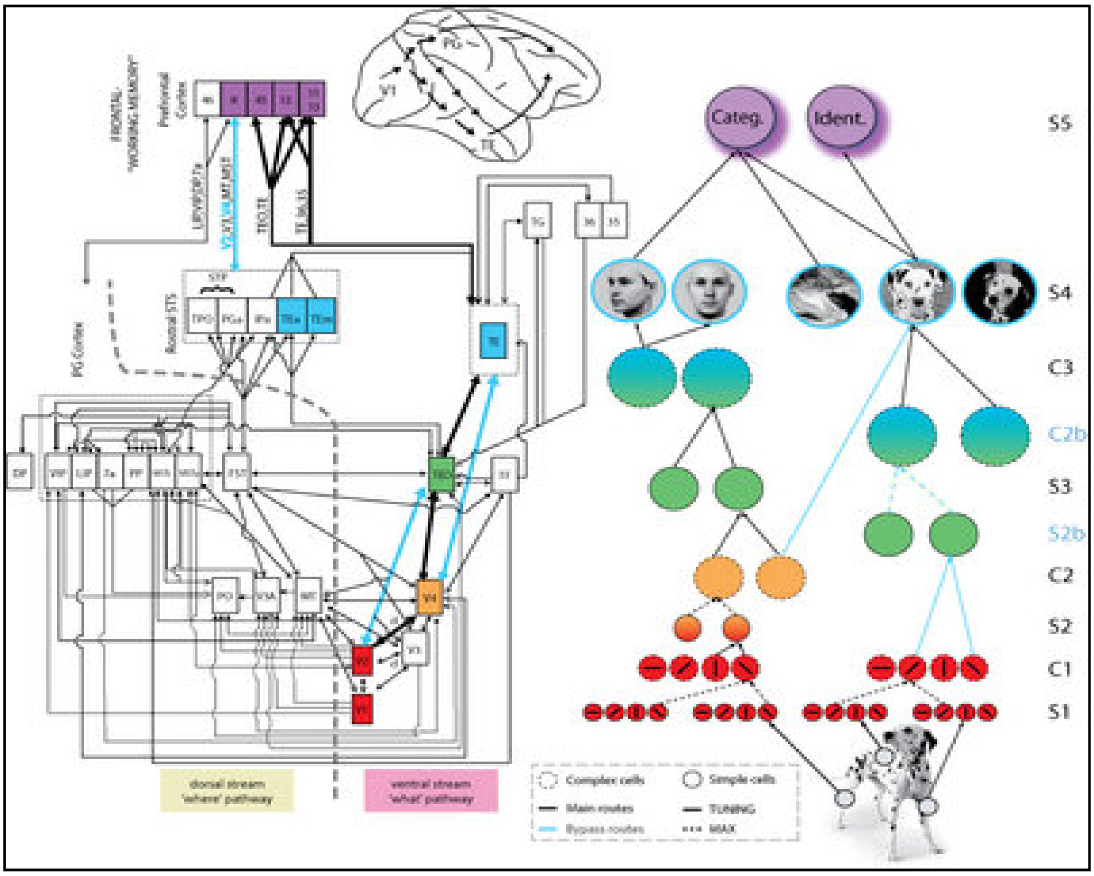


Figure 2.6: Model of object recognition (right) based on the feed-forward process in the ventral stream of the visual cortex (left) (Serre et al. 2007a)

Models that are based on the feed-forward model of visual system (mentioned above) have provided some good results. However, these results are only obtained when recognizing objects in scenes that have little amount of clutter and zero occlusion during the first glimpse. We humans sometimes cannot recognize objects at the first glimpse in clear scenes, and therefore, if there was clutter in the scene, it will be hard for us to recognize all objects. Hence, feed-forward mechanism of object recognition is not the best solution since it cannot handle all situations. Even in (Kreiman et al. 2007) it was mentioned that the performance of the feed-forward model dropped from 90% to 74% when the amount of clutter increased in the scenes that were used for testing.

(Lian and Li 2008) introduced an improvement of the model developed by (Serre et al. 2004; Serre et al. 2005). Their model consists of 4 layers as Serrer's model (Serre et al. 2005) ; however, it incorporates additional biological features that were not included in the model; these features or characteristic are: "the manner of neuron firing, feature localization, and merging unit features in the higher layers". As shown in figure 2.7, S1 units are calculated by applying a battery of Gabor filters (see equation 2.2) (Lian and Li 2008) on the input image; in their case only high frequency band will be extracted, and low frequency will be ignored to reduce the computational complexity.

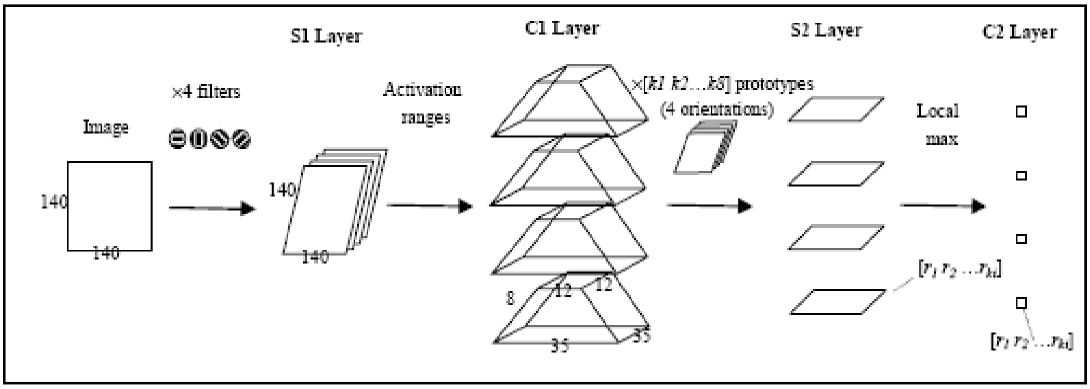


Figure 2.7: Lian & Li's improved model (Lian and Li 2008)

$$G(x, y) = \exp\left(-\frac{(X^2 + \gamma Y^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}\right) \qquad (2.2)$$

Their experiment showed that the performance of considering only high frequency bands is similar to considering the whole bands whether in high or low frequencies. In C1, they calculated statistical number of S1 firing units in different sizes, and then normalize them to sizes.

In S2, a prototype matching using a regularized RBF function (2.3) is performed between C1 patches and random sampled prototypes from C1 features of training images. Finally, in C2, they calculated the max over particular position for all C2 map.

$$R(X,P) = exp\left(-\frac{\parallel X - P \parallel^2}{2\sigma^2\alpha}\right) \qquad (2.3)$$

The speed of this model in recognizing objects was reported to be better than standard model of (Serre et al. 2007a), and the performance was quite similar to Serre's model (Serre et al. 2005).

In this paper, the model that was proposed also belongs to the family of feed-forward models of object recognition, and therefore, it has the same weakness of its inability of recognizing objects in clutter scenes. Although, they have added new features that helped in speeding up the process of extracting feature by only considering the high frequency bands, it produced a very much similar results compared to previous models of object recognition based on the feed-forward process of the brain.

2.2.3.2 Feedback Models

Another process that has been discovered in neuroscience is the feedback process in the visual system. In fact, both feed-forward and feedback are two processes that complete each other to help human and primates to recognize objects (Kim et al. 2004). Visual attention (Bermudez-Contreras et al. 2008) is associated with feedback. In fact, our visual system activates some neurons that correspond to relevant locations and features to attend to potentially significant objects (Saalmann et al. 2007). Attention acts as a filter that ignores any irrelevant information in scenes that have an increase amount of clutter.

Müller and Knoll introduced a biologically inspired system that uses visual attention to filter the scene and reduce any unwanted data; the remaining data will be used in further analysis such as object recognition. The system is developed to enable robots to recognize objects. It uses a mechanism to detect salient local feature based on the comparison of intensity and hue. This will help in creating a saliency map that highlights the relevant area in the image for further analysis, so this mechanism was applied in static and dynamic saliency. The attended region is mainly created using the attention detectors that are inspired by the bottom-up process. After that it uses the mechanism of top-down process of feedback in order to focus and makes the system able to ignore whichever area that has been analyzed before. Basically, it uses previous knowledge in order to focus the attention on other regions that have not been analyzed. The mechanism is called “inhibition of return”. The results of this application were good; the robot vision system could recognize almost all the objects that appeared in the scene with some faults (Müller and Knoll 2008). The introduced system that is “in a way” integration between the bottom-up and top-down processes. In fact, the paper shows evidence that top-down and bottom-up can be integrated to produce an object recognition system.

Similarly, the biologically inspired models in (Kim et al. 2004; Siagian and Itti 2007) are recognizing or categorizing objects using the same mechanism and the results were good as well.

Although the current models that depend on the feedback process are quiet good, neuroscience evidence shows the significance of recognizing objects by integrating both feed-forward and backward processes in order to get a better result and get the ability of the human and primates visual systems.

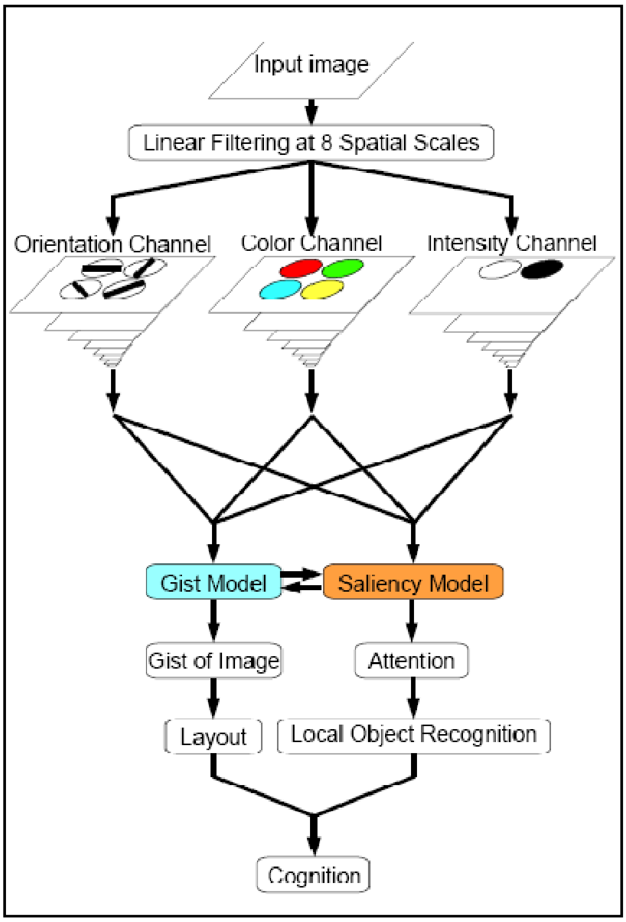


Figure 2.8: Attention as shown in the model proposed by (Siagian and Itti 2007)

2.2.3.3 Object Recognition by Bottom-Up and Top-Down

Neuroscientists have done researches to be acquainted with the visual system and how does it build an image and understand it. After it has been known how the brain formulates the image (Roorda 2002) (which helped in manufacturing the camera device) research moved to indentify how the visual system recognizes objects. Hubel and Wiesel's discovery (Hubel and Wiesel 1962) helped researchers to understand the initial steps in object recognition in the cortex. As mentioned in section 2.2.3.1,(Riesenhuber and Poggio 1999; Riesenhuber and Poggio 2000),(Serre et al. 2007a; Serre et al. 2004; Serre et al. 2005) and (Lian and Li 2008) models were inspired by the feed-forward mechanism of the human visual system that was discovered in (Hubel and Wiesel 1962).

In neuroscience, (Graboi and Lisman 2003; Kveraga et al. 2007; Rosenholtz et al. 2007) reported the importance of visual attention that is associated with visual feedback mechanism in recognizing objects. For example, by looking at figure 2.9, if a human were to be asked to recognize if there is a glimpse of a bottle of water in the scene, feed-forward process in the brain will not help since immediate recognition would be hard in this complex scene. However, increasing the amount of time by which a person looks at the scene, visual attention (which is part of the feedback process in the brain) will help in filtering the scene and ignoring unwanted objects such as humans. In the end, the person will recognize the object.

Figure 2.9: Objects in a natural scene with high amount of clutter (Source: (Rosenholtz et al. 2007))

As shown in the example, attention could improve the feed-forward to help in recognizing object in highly cluttered scenes. (Graboi and Lisman 2003) support the opinion that integrated model of top-down and bottom-up will produce better results of object recognition, since our brain employs this technique to recognize objects.

2.3 Human Visual System

The human visual system has been under research for a long time. It has amazing capabilities in perceiving the surrounding world and a complex anatomy that took neuroscientist years to understand how it works and what are the areas related to vision. The outstanding capabilities of the system in recognizing objects in an unusual or difficult situation have motivated computer scientist to try to understand the mechanism by which it operates and map those abilities into computational systems.

2.3.1 Anatomy of the Visual System

The visual system of humans consists of the following parts:

- **The eye:** this is the capturing device that captures the objects, the environment and everything that is around us. In fact, the human eye captures the light that is detected by the retina and transformed into electrical signals. These signals leave the eye and travel to the *lateral geniculate nucleus.*
- ***Lateral Geniculate Nucleus* (LGN):** acts as the middle man between the eye and the primary visual cortex. LGN is located at the thalamus on each side of the brain. The LGN transfers the electrical signals that have been received from the eye to the primary visual cortex. (Figure 2.10 shows the human visual system anatomy) (Serre 2006).
- **The visual cortex refers to the primary visual cortex:** (striate cortex) or area V1, and the extra striate cortical areas (V2, V3, V4 and V5). The primary visual cortex or area V1 receives the input information from LGN and then passes the information to two primary pathways, the dorsal pathway which is known as the "where" path way, and the ventral pathway which is known as the "what" pathway.

The dorsal pathway is associated with motion and location, while the ventral pathway is associated with object recognition and categorization.

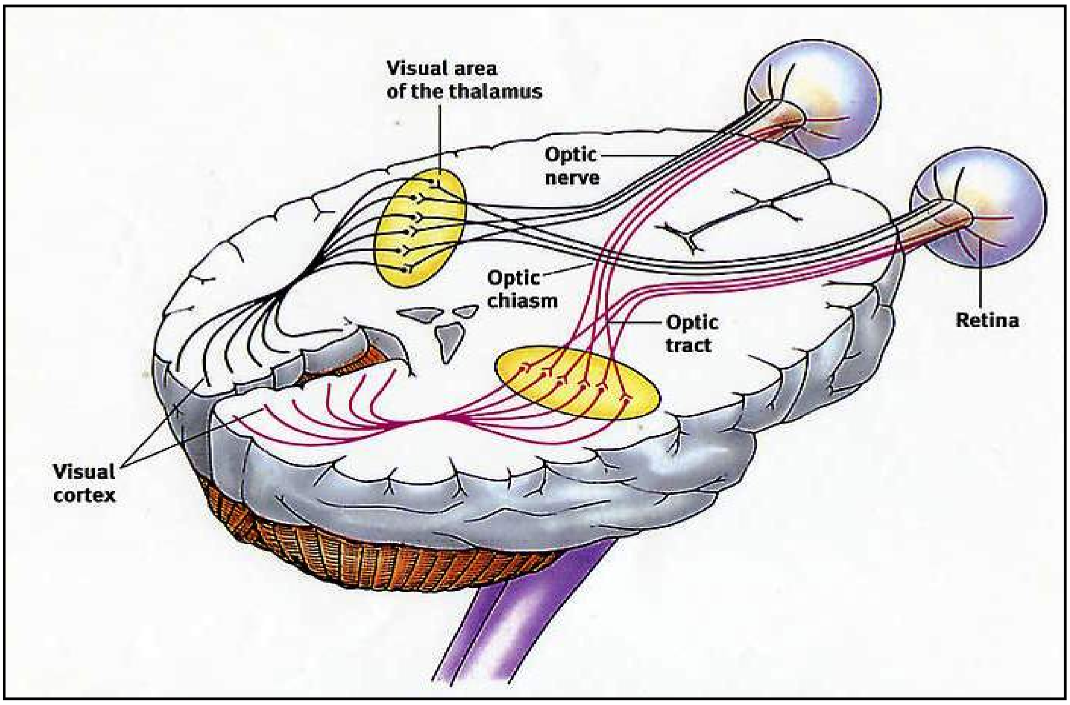


Figure 2.10: Visual path from the eye to the visual cortex

The visual cortex areas in the ventral stream (what pathway) that is associated with object recognition are[5]

1. Area V1: Receives input information from LGN and passes the output to other areas. It consists of selective spatiotemporal filters, which process the spatial frequency, orientation, motion, direction, speed, and other features.
2. Area V2: Receives information from area V1 and sends to other areas. The functionality of area V2 is similar to V1; however, V2 neurons' responses are adjusted by more complex properties such as the orientation of false contours.
3. Area V4: Part of the ventral stream, it receives input from V2 and primary visual cortex. V4 is adjusted for orientation, spatial frequency, color and object features of intermediate complexity.
4. Inferior Temporal Cortex: an area in the brain that is responsible on object representation in both human and monkey (Kreiman 2008).

[5] http://www.experiencefestival.com/visual_cortex

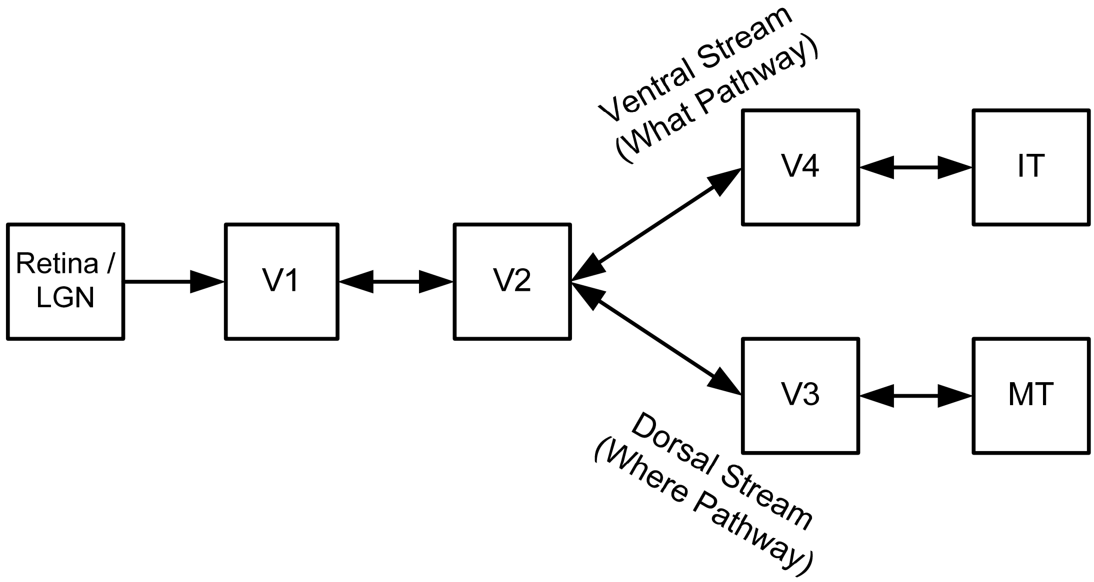


Figure 2.11: The organization of the ventral pathway of visual cortex

Figure 2.11 shows the above mentioned areas, their organization in the brain and the connection between them which is in feed-forward and backward. In addition to the visual areas, there are two main mechanisms of connection among the different visual areas namely feed-forward (figure 2.12) and feedback (figure 2.13). In the feed-forward mechanism, the information is being communicated among the visual areas in one way from the lower visual areas to the top visual areas whereas the feedback mechanism communicates the higher visual areas to the lower areas in a feedback manner. Most computer scientists have been focusing on the feed-forward mechanism and mapping the functions of the visual areas when they are communicating among each other in the feed-forward manner.

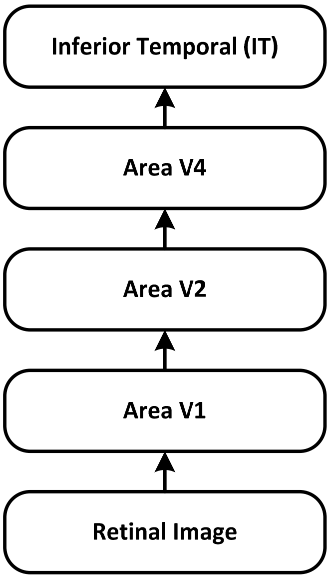


Figure 2.12: Feed-forward Connection among visual areas

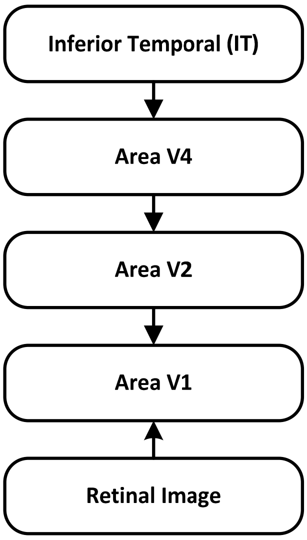


Figure 2.13: Feedback Connection between visual areas

Models that employ this mechanism has got a weakness since they can only obtain information from one side to another which does not show exactly how the human visual system works. These models' weakness appears when they are given the task of recognizing objects in high cluttered scenes and partially occluded objects. The reason behind this weakness is that the information are being communicated only one time in one way; however, the human visual system works by interacting with all the visual areas and allowing them to communicate in both ways in order to share information or request for new information at the end to the amazing capability that human experience all the times. Figure 2.14 shows the integration of both feed-forward and feedback in the ventral pathway.

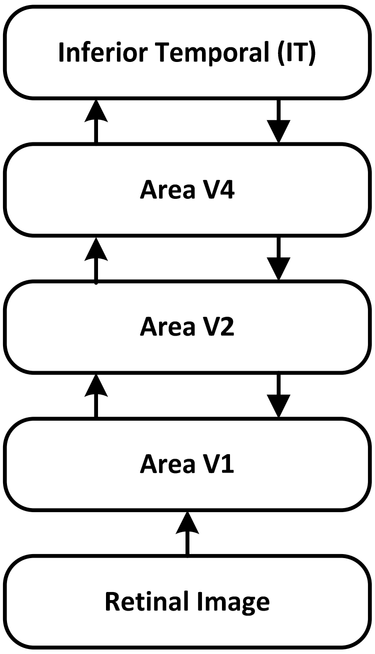


Figure 2.14: Connection among the visual areas in human (an integration of feed-forward and feedback mechanisms)

2.3.2 Object Recognition by Component

One of the capabilities of the human visual system is recognizing object by their components. This means if the brain cannot recognize objects that are not fully visible; it will match the features of the visible parts of the object with object's parts that have been stored in the memory. A theory of recognition by component was introduced by (Biederman 1987) where it is stated in the theory that humans can recognize object by dividing them into *geons* which mean a group of various shapes that can be brought together to form many objects.

In order to understand the theory of recognizing objects by their component, let's consider the pictures in figure 2.15. By looking at figure 2.15, the pictures a, b and c represents three different parts / components of a car. The human visual system will be able to identify the name of this object which is a car by recognizing these parts. In fact, the human visual system has an amazing capability that will name each part of the object's component.

(a) (b) (c)

Figure 2.15: a) middle part of a car, b) back part of a car, c) front part of a car. Recognition by Component (if the object is not fully visible, the human brain will be able to recognize it from its parts)

Recognition by component function which is part of the bottom-up process of information travelling in the human visual system is important in allowing the human to recognize objects in highly cluttered scenes as well as partially occluded objects. In fact, this function can be tested by any person by recognizing the objects that are not fully clear.

2.4 Summary

As shown, models of object recognition have been developed using the feed-forward mechanism that works for immediate object recognition by looking at a scene in a glimpse. The models were able to mimic the human visual system's ability of object categorizing in the first 150 milliseconds. The models produced good results; however, they had a weakness in recognizing objects in high cluttered scenes and objects that are partially occluded. Other systems that utilized the feedback mechanism to recognize objects have been reported to have good results as well. However, it has been discovered in neuroscience that the human visual system recognizes objects by using bottom-up and top-down mechanisms. The integration of both feed-forward (bottom-up) and feedback (top-down) connection mechanism with their associated would produce a system with more capabilities in recognizing objects in difficult situations.

CHAPTER 3

BIOLOGICALLY INSPIRED MODEL FOR OBJECT RECOGNITION

3.1 Introduction

The human visual system has an astonishing ability in recognizing object with various conditions. Each of the visual areas comprises the human visual system that has a specific role in recognizing the intended object(s). In addition to that, there are two mechanisms of communication among the different visual areas, namely feed-forward and feedback.

The organization of the visual system (see figure 2.11) is divided into two main pathways called the dorsal pathway and the ventral pathway. Both pathways are connected to areas V1 and V2. The difference between the two pathways is that the ventral pathway function is to recognize objects while the function of the dorsal stream is object tracking and motion detection. In this study, the focus is on the ventral pathway only where the model will be developed to recognize object regardless to whether or not it is moving.

In the next sections, an explanation on mapping the functions of the human visual system into a computational model which is based on the integration of the feed-forward and feedback mechanisms of information processing in the cortex will be demonstrated.

3.2 The Proposed Bio-Inspired Model for Object Recognition

In this section, the proposed model of object recognition will be discussed. As explained earlier, the functions of the two main mechanisms of connections between the visual areas have been mapped, the feed-forward and feedback mechanisms.

3.2.1 The Concept of the Model

As shown in chapter 2, an object recognition that is inspired by the human visual system would perform better if it employs the integration of the top-down and bottom-up mechanisms. The human visual system works by utilizing these techniques in order to recognize objects. Figure 3.1 shows the concept of the new model.

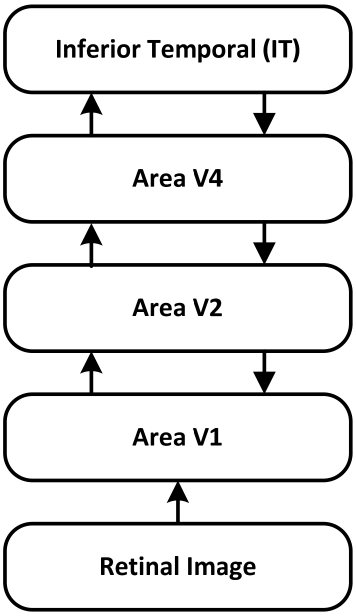


Figure 3.1: Integrated top-down and bottom-up model

As shown in the figure above, the retina will get the image from the outside world and pass it to the primary visual cortex (V1) via LGN. The primary visual cortex or V1 and area V2 will apply feature extraction on the incoming signals. Although both V1 and V2 apply feature extraction, V1 respond to simple features while V2 respond to complex features. In computer vision, the first step in processing an image is to extract all the features in the image. Likewise, areas V1 and V2 in the human visual system extract all the features from the input signals that represent the captured image. The features extracted include the edge features, the color features, the orientation etc. At the beginning, the features extracted by area V1 will be transferred to area V2 in a feed-forward manner to extract the complex features and form the final feature map.

The extracted features will be transmitted to area V4 through the feed-forward mechanism of information communication among the areas V2 and V4. In visual area V4, the features will be adjusted in terms of orientation and spatial frequency before it is passed to the inferior temporal cortex in a feed-forward manner.

When the final image features reach the inferior temporal or IT, it will be processed by the IT to obtain the feature's class / category in the case the object captured was a clear object. If the captured is not a clear object, the system will need more time and processing in order to obtain the exact decision with regards to those objects which were not clear. Figure 3.2 shows an example of a clear object.

The processing of the objects that were difficult to be recognized / categorized during the first round will involve the passing of information such as the features of those objects. The passing of this information to area V1 will be via the feedback connection mechanism. The feedback mechanism is associated with the visual attention.

Figure 3.2: Clear object (Source: (Serre et al. 2007a))

The visual attention is part of the function of the human visual system that works to help the visual system to focus the attention on the suspected objects within the scene and ignoring the rest of the information (Navalpakkam et al. 2005). Visual attention is a function of the human brain that is utilized by the human hearing system where, if there is more than one person talking at the same time, the hearing system will only focus the attention on the intended person and ignore the other sounds.

During the first round of information processing, IT will be able to identify the suspected objects in the scene through their features that have been extracted. The visual attention will pass this information to area V1 to extract the features for the second round (or maybe more rounds depending on the complexity of the scene) before sending them again to the IT. Area V1 will extract the features only at the regions that have been identified as suspect and will ignore the rest. After that, the new set of extracted features will be sent again to the IT in a feed-forward manner. Figure 3.3 shows an example of a complex scene.

Figure 3.3: Complex scene that requires more processing time (Source: (Serre et al. 2007a))

In the human visual system, the feed-forward processing will take up to 150 milliseconds (Serre 2006), and for clear objects, it will be able to recognize or categorize them within that period of time. However, as the complexity of the scene increases, the period of recognition will be increased as well.

As shown earlier, models that employ feed-forward mechanism only will not be able to recognize objects in complex scenes. Actually, those models have reported a drop in their performance when subjected to recognize objects in complex scenes. As a result, integrating both feed-forward and feedback mechanisms to form a computational model will result in performing better on complex scenes.

3.2.2 Bio-Inspired Model for Object Recognition

As mentioned in the previous chapter an integration of both feed-forward and feedback processes of the human visual system will produce object recognition with highly accurate results. The proposed model in this study employs the integration of the feed-forward and feedback mechanism of information passing in the human visual system (figure 3.4).

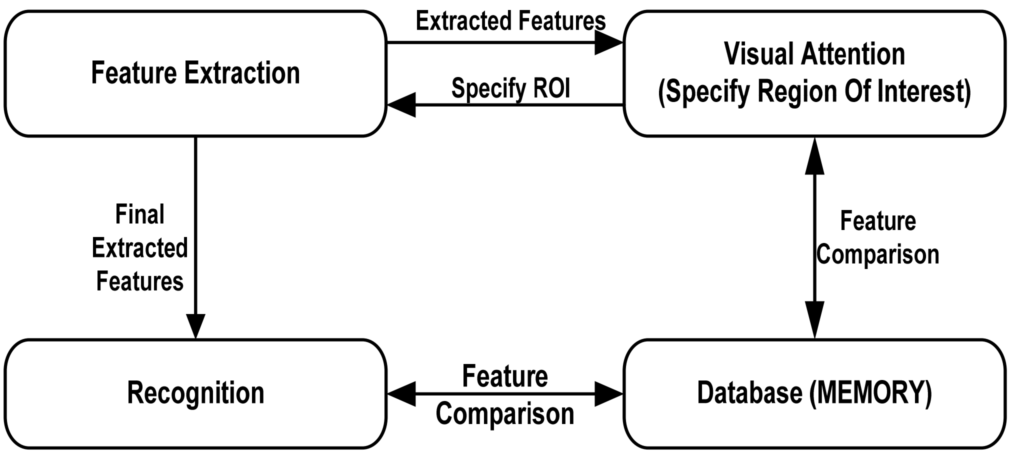


Figure 3.4: Bio-Inspired Model for Object Recognition (Abstract level)

The model consists of four components that correspond to the functions of the ventral pathway of the human visual system as well as the functions of the feed-forward and feedback such as visual attention (feedback). The components are: feature extraction, visual attention, recognition, and image database. In the next section, each component will be explained in terms of its function and role during the object recognition task. The concept of the model will help in obtaining a reliable, robust and efficient system with regards to complex scenes.

3.2.2.1 Feature Extraction (FE) Component

Feature extraction (FE) component will extract features of all objects in the input image and send them to the visual attention component which will specify the region of interest (ROI) of the intended object. After that, the ROI will be sent to the FE again for another round of feature extraction at the specified region. The final extracted features will be sent to the recognition components. The FE component acts as visual areas V1 and V2 whose job is to extract the features of all the objects that the human eye captures and sends them to the brain.

3.2.2.2 Visual Attention (VA) Component

The visual attention (VA) component's role is to identify the intended objects and specify the ROI for each object (Frintrop 2006), VA will get the features from the FE components and send them to the database. Based on the feedback of the database, the VA should be able to specify the ROI and send it to the FE component for further processing. The VA component acts as the feedback process in the human visual system where the IT area would categorize the objects and send a feedback to area V1 to further invistigate the attended region. The VA job is just as a classifier that gets the features as an input and produces the intended object(s) as an output.

3.2.2.3 Database (DB) Component

Storing the image of data will require a database. This database is in the form of a file that contains the data that represents all the intended objects. When the VA receives the features from the FE component, it will perform a comparison between the values of the features and the values stored in the database (file). The result of the comparison will determine whether or not the input image contains any objects that are similar to what is stored in the database. After that, it will send a feedback to the visual attention which will specify the ROI of each object. On the other hand, the database will be utilized by the recognition component to recognize objects. In fact, there will be two files (databases) one will be utilized by the classifier to determine whether or not the intended objects are available and the second database will be utilized by the recognition component to recognize the objects. In order to store images / objects in the database, a training phase must be performed, where all the intended objects data will be stored.

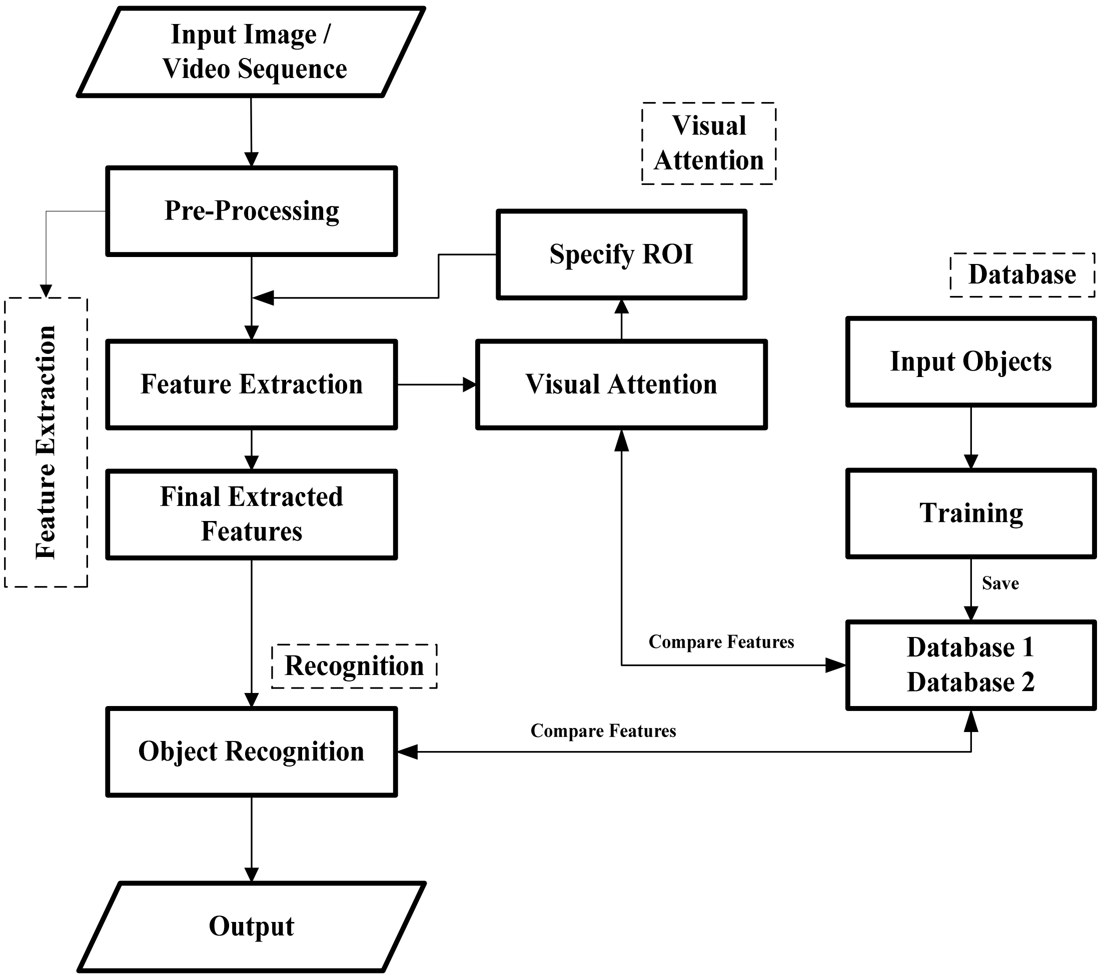


Figure 3.5: Interaction between the Components

3.2.2.4 Recognition Component

The final stage of the object recognition in this model lies at the recognition component. The final extracted features by the FE components (after focusing the attention on the ROI) will be sent to the recognition component. The recognition component will use the data stored in the database to compare them with the input features from the FE components, and recognize the objects according to their availability in the database. The model is further illustrated in figure 3.5 where the diagram shows the interaction between the different components of the model.

3.2.3 Model Formal Specification Using Z Notation

Formal specification language is a way of explaining any computer science system in a formal way. Many formal specification languages have been developed such as Z language (Spivey 1989). Z is a formal specification language that is based on the set theory. In this section, the formal specification of the model using Z specification language is shown:

Get Image from a Device

InputImage

aImage: IMAGE

---

aImage? $\in$ *IMAGE*

Feature Extraction

FeatureExtraction

aImage: IMAGE
aObject: OBJECT
aFeature: FEATURES
aFinalmap : FEATUREMAP
extract : IMAGE □ FEATURE
finalfeature: extract □aFinalmap

---

aFeature $\in$ *aObject* $\in$ *aImage*
aImage = aFeature $\cup$ *aObject*

Visual Attention

VisualAttention

aAttention: LOCATION
aFeature: FEATURE
aCoordinate: COORDINATE
roi: FEATURE □ LOCATION

LOCATION = aFeature → aCoordinate

Database

ImageFeatureDB

aImageID: IMAGE
aImageFeature: FEATURE
iDatabase : DATABASE

---

iDatabase = aImageID ∪ aImageFeature

AddNewImage

ΔImageFeatureDB
id?: aImageID; aImFeature? : FEATURE

---

newEntry! = id ∪ FEATURE

Object Recognition

ObjectRecognition

Ξ ImageFeatureDB
ΞFeatureExtraction
aResult: RecogObj

---

afinalmap ? ∈ iDatabase
aResult! = aImageID

3.2.4 Algorithms

In order to implement the proposed model, the model's components should be matched with an appropriate algorithm in order to demonstrate how the model works. In this section, the algorithms that have been identified for each component are discussed.

3.2.4.1 Feature Extraction

The FE component will apply a feature extraction procedure on the input image to get the features of all objects in the image. When the features are obtained, they will be sent to the VA component to allocate the desirable objects among others. For the clear objects, detected regions will be sent to the recognition algorithm to recognize them, but for the objects that are not clear which mean that they are classified as suspect; the detected regions of those objects will be subjected to a second round of feature extraction to confirm that they are among the desired objects.

The Haar-like features have been chosen in this study to do the task of feature extraction. In Haar-like features (figure 3.6), the main motivation of using features instead of pixels is the great speed that can be achieved by using integral image. For all rectangle features, values are computed as the difference between the black area and white area.

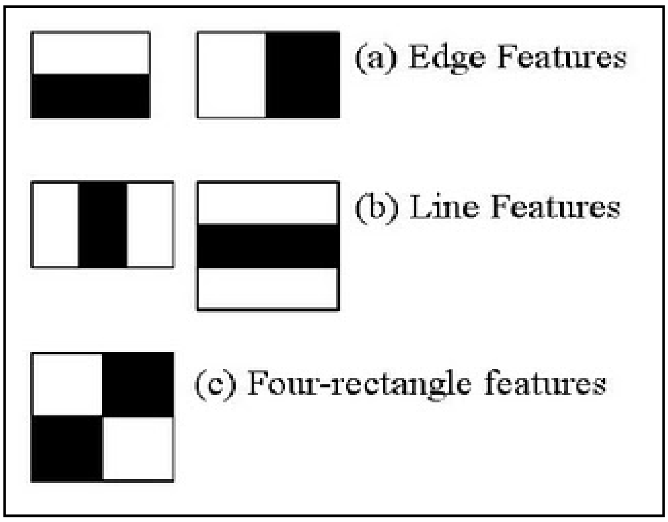


Figure 3.6: Haar-like Features

These features use the change in contrast values between adjacent rectangular groups of pixels. The simple rectangular features of the image can be calculated by using an intermediate representation called "*Integral Image*" where each feature can be computed at constant speed regardless of its scale or position. The *Integral Image* value at any location is the sum of all pixels above and to the left of (x,y) (see figure 3.7) .

If we assume that I[x,y] is the original image and II[x,y] is the integral image then:

$$II[x,y] = \sum_{x' \leq x, y' \leq y} I(x',y') \qquad (3.1)$$

As shown in figure 3.7, the *Integral Image* can be represented as a table that provides the area of the above and left of each pixel. As illustrated in the figure, only four points are needed to calculate rectangle sum and eight points are sufficient to calculate the difference of rectangle sum. Therefore, only points 1,2,3 and 4 are needed to calculate the rectangle sum of area D where 1,2,3 and 4 represent the areas A, A+B, A+C and A+B+C+D respectively. Thus, it is found that the area of D is 4+1-(2+3). The computed features are then sent to the VA component which is represented here as a classifier that will be discussed in the next section.

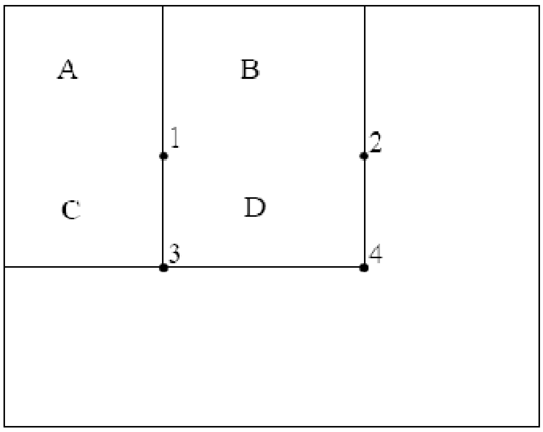


Figure 3.7: How *Integral Image* is used to calculate features

3.2.4.2 Object Classification

The visual attention in the proposed model acts as the classifier that will determine the availability of the intended object(s) and their category. In fact, the classifier acts as a filter that will select the features that represents the intended object(s) or the suspects object(s) (in complex scenes) and specify the regions that contain those objects. After it detects the objects, it will produce the ROI for each object and send it back to the FE component.

In this research, Haarcascde classifier is used. The classifier works very well with the Haar-like features. After obtaining the features using the *Integral Image*, the computed features will be passed to the classifier. The classifier has been trained using a set of positive and negative images. The positive image contains the intended object and the negative images represent scenes that do not enclose the intended object.

In order to build a strong classifier, the Adaboost algorithm of learning is used in order to build a strong classifier. The idea of this learning algorithm is to build a classifier that is a combination of multiple weak classifiers using a procedure called boosting. The boosted classifier is built as a weighted sum of weak classifiers. First, the weak classifiers are trained by selecting single feature. During the training the error rate is evaluated. Then, the classifier with the lowest error rate is chosen and the weight is updated until the final classifier is formed (Bardski et al. 2005). Figure 3.8 shows how the algorithm works.

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0,1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0,1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$:
  1. Normalize the weights,
  $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$
  so that w_t is a probability distribution.
  2. For each feature j ,train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
  3. Choose the classifier, h_t, with the lowest error ϵ_t.
  4. Update the weights:
  $$w_t + 1, i = w_{t,i} \beta_t^{1-e_i}$$
  where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is :

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

Figure 3.8: Adaboost Algorithm for classifier learning (Source: (Viola and Jones 2001))

In addition, figure 3.9 illustrates how the cascade of classifiers works (Viola and Jones 2001). Each stage represents a weak classifier that is trained with the positive and negative data. During the training of the cascade, at each stage, the classifier will reject the negative images (0) and pass the positive (1) images to the next stage (classifier) for further training / classification which gives a high detection rate at the end.

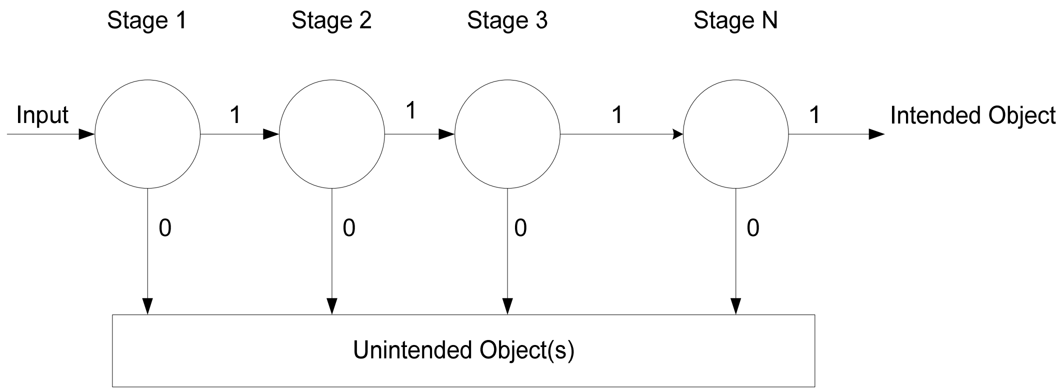


Figure 3.9: Cascade of classifier with N stages

After the training stage, the algorithm will produce an XML file that represents the intended object(s). This file is used in the comparison stage when a new image is presented to the system.

When the classifier receives the extracted features, it will compare those features with the XML file and identifies the intended object(s). Subsequently, the system will specify the region of interest based on the output of the classifier which will determine the area that contains the intended object(s) as well as the suspected objects (in complex scenes).

Once the system specifies the regions of interest for each detected object; it will send those regions to the feature extraction component which will send the final features for each region to the object recognition component. In the object recognition component, the algorithm will apply the feature comparison to decide whether or not the incoming region contains the object needed. In some cases there could be a region that contains unintended objects which were detected as false positive, if that happened, the recognition component will be able to detect this wrong detection and ignore the image by marking it as an unintended object.

3.2.4.3 Object Recognition Using Principal Component Analysis (PCA)

Principal Component Analysis (PCA) (Aravind et al. 2002; Smith 2002) is a statistical approach of identifying patterns in data and reforming the data in such a way as to express the similarities and differences. PCA is a method based on the information theory that extracts small set of features called "Eigen objects" which consists of the principal component for a given training set. Eigen object represents the differences among the individual data of the objects which are very important to perform the recognition.

The recognition task is performed by projecting the test object image into space spanned by the Eigen object called object space and then classified by comparing its position in object space with the positions of face images of the training set in the same object space.

PCA for object recognition:

Let individual objects in the training set be $\Gamma_1, \Gamma_2, \Gamma_3, \ldots . \Gamma_M$

The average object is defined as

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_{\mathrm{n}} n \qquad (3.2)$$

Each object class is different from the average object by

$$\Phi = \Gamma_{\mathrm{i}} - \Psi_{\mathrm{i}} \qquad (3.3)$$

This would compose a very high dimension set of vectors. The set will be subjected to PCA to create a set of M orthogonal u_n vectors which best describes the distribution of the data.

The k$^{\text{th}}$ vector u_k would be chosen so that

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \Phi_{\mathrm{n}})^2 \qquad (3.4)$$

is the minimum. The vector u_k and the scalar λ_k are eigenvectors and eigenvalue, respectively, of the covariance matrix C

$$C = \frac{1}{M}\sum_{n=1}^{M} \Phi_n \Phi_n^T \qquad (3.5)$$

The Eigen object span an M' dimensional subspace of the original N^2 image space. The M' significant eigenvectors are selected as those with the largest equivalent eigenvalues. A test object image is projected onto object space by the following operation:

$$w_k = u_k^T(\Gamma - \Psi) \qquad (3.6)$$

The weight will form a vector $\Omega^T = [\, w_1 w_2 w_3 \dots w_{M'}]$ that represents the contribution of each object of training set to the face under test. The image will be classified and recognized based on Euclidian distance 3.7 between image under test and others of training set (Tripathi et al. 2009; Turk and Pentland 1991).

$$\varepsilon_k^2 = \|\Omega - \Omega_k\|^2 \qquad (3.7)$$

How Eigen objects works

- Eigen object represents the significant differences among training set
- Each Eigen object represents only certain features which may or may not present in the original image
- Each object can be rebuilt by summing Eigen object with right portions (weights)
- Weight vector represents the degree each specific features "Eigen object" present in the original image
- To perform recognition weight space is built by calculating weight vector of each image
- For new image, weight vector is calculated and compared to those in weight space

3.2.5 Bio-Inspired Model vs. Other Models

The proposed model has an advantage over other models that have been developed. The integration of the feed-forward and feedback mechanisms in the human visual system and mapping it to the proposed model has an advantage in making this model perform better in complex scenes. In fact, the model mimics most of the features of the human visual system in the ventral stream, from feature extraction, visual attention and object recognition. Since the model has all the features mentioned above, it will be able to recognize objects in highly cluttered scenes and also it has the ability to recognize partially ocluded objects.

CHAPTER 4

FACE RECOGNITION: APPLYING THE BIOLOGICALLY INSPIRED MODEL OF OBJECT RECOGNITION

4.1 Introduction

In this chapter, the proposed model in chapter 3 will be implemented in an application to test the model's features and its ability to recognize objects and produce accurate and robust results in different situations particularly situations that involve partially occluded objects and high clutter scenes. The model will be implemented using the algorithms that have been identified and mentioned in the previous chapter. In addition, MATLAB has been chosen to be the tool of implementation. The application that has been chosen to test the model is a face recognition system.

Face recognition has many benefits in life. Developing an application that can guarantee more accurate results has been the target of many researchers in computer vision. Although many systems have been developed for face recognition, there have been some challenges that affect the performance of those systems. The issues are enclosed faces that are in cluttered scenes as well as partially occluded faces. Since the proposed model adopts the recognition by component which is one of the capabilities of the human visual system, it is undoubted that the system implemented in this chapter will be able to recognize faces in cluttered scenes as well as faces that are affected by occlusion. Thus, the following sections give detail explanation on how the proposed model can be implemented in a face recognition system and to what extend it can solve the problems of cluttered scenes and partially occluded faces.

4.2 Face Recognition

As explained earlier, the human visual system applies the recognition by component strategy where if a human brain cannot recognize the full shape of an object, it will look for features that represent any element of that object and subsequently it will try to construct the full shape in order to recognize the object. The same methodology can be applied to recognizing faces that are partially occluded or those in high cluttered scenes. For example, if the human eye can capture only half of the face, it is enough for the visual system in the brain to construct the other has and identify the person. Similarly, if the system captures and detects half a face, it will be able to identify the person. The idea is to train the classifier to be able to identify the face components (such as nose, eye, mouth etc) (Wilson and Fernandez 2006) from the features extracted in the scene and if the VA (classifier) confirms that one element is available, the VA component will identify that area as suspect and therefore, it will expand the ROI to cover the surrounding areas and send it back to the feature extraction to apply a second round of feature extraction on the specified region and pass it to the recognition component. The recognition component will be able to decide whether or not the region contains face or not by recognizing the face element such as profile face, nose, eye, mouth etc. If the area “looks like” a face, the recognition component will decide finally whose face is that if it is available in the database. Figure 4.1 shows the stages of the system.

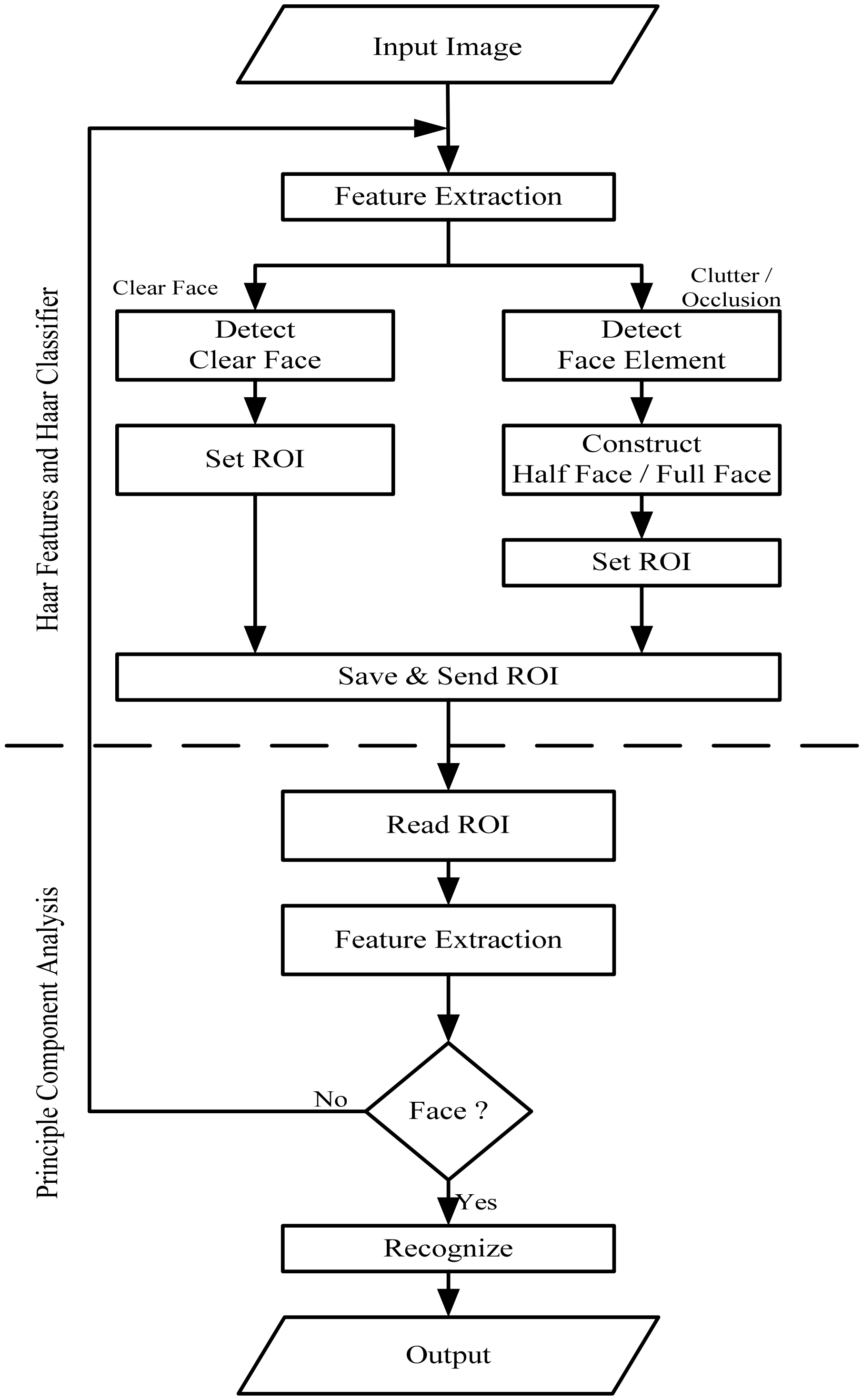


Figure 4.1: Face recognition system based on the proposed model

4.2.1 Feature Extraction

As shown in the proposed model in figure 3.4, the first stage is to apply feature extraction on the input images. For every image, the FE component will apply Haar-like features in order to compute the features of all objects in the scene. For this application the Haar-like features that are being used to compute the face features are the edge features and the line features. Figure 4.2 illustrates how to apply Haar-like edge and line features in computing the face features.

After the features have been computed using the *Integral Image* for all adjacent rectangles, it will be evaluated by the visual attention component (classifier) which is represented in the system by the cascade boosted classifier.

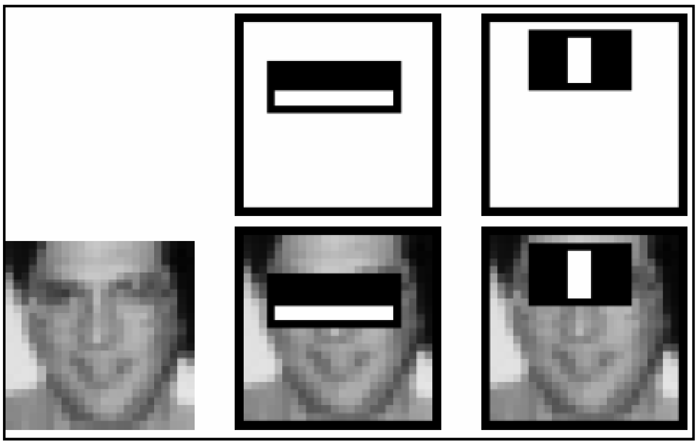


Figure 4.2: Face feature extraction using Haar-like features (Viola and Jones 2001)

4.2.2 Face Detection

The second part of the system is the face detection which is the function of the VA component in the proposed model. Face detection can be achieved by passing the features extracted to the cascade boosted classifier which will compare the computed features with those stored in the XML file that contains all the values of the training data. The XML which contains the data of the classifier is obtained after training the classifier algorithm. In this stage, the open source computer vision library (OpenCV) is utilized. OpenCV is an image processing library written in C language and was developed by Intel Corporation, however, since the implementation tool is MATLAB, the C code of the object detector had to be converted to MATLAB readable code. As a result, the MEX function of MATLAB has been used in order to compile the C code of the object detector to be called in MATLAB after it has been trained in C language. The OpenCV installation package contains ready-to-use classifiers for objects such as face, eye, nose, mouth, profile face etc. In this study, the face and eye classifiers have been adopted in order to detect the face and eye as an example of face element that can be utilized to construct a half face (in cluttered scenes) after it is detected by the classifier, and then recognize the face. Figure 4.3 shows how the cascade boosted classifier evaluates a new image to determine whether or not it contains a face.

Basically, the idea is to put the whole features into the object detector, which will check in the first round whether the features contain some intended objects. If faces or faces' elements are detected in the first round, it will be passed to the second stage and the same process is repeated. As for the features that were identified to not having the intended object(s), they will be classified as non-face and will be ignored in the next stage. Using this technique of having multiple stages in the same classifier results in more accurate results and therefore contribute towards the overall performance of the system.

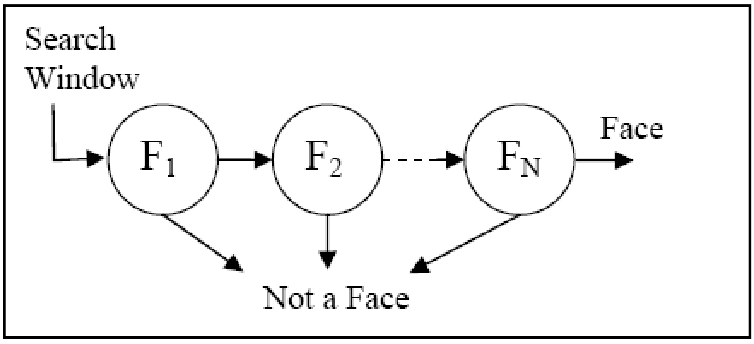


Figure 4.3: Face detection in Adaboost cascade classifier (Bardski et al. 2005)

4.2.2.1 Training a Classifier

According to (Bardski et al. 2005), in order to train the Adaboost classifier in OpenCV, the following steps must be followed:

- Collect a database that contains positive samples (faces / eyes). Figure 4.4 shows some positive samples of images that contain faces and eyes.
- Collect a database that contains negative samples, which are images that do not contain any instances of the intended object (in this case, faces or any face element).
- The data need to be converted into a format that is acceptable by the classifier (Images need to be converted into numerical data that represent the features values of all pixels). If the classifier could not read the data, then the output classifier will not perform well.
- After the data has been converted to the accepted format by the classifier, the training will start by extracting the Haar-like features then pass the computed features to the cascade of classifiers and finally produce the XML file.

The training procedure must be strictly followed in order to obtain the intended result at the end.

Figure 4.4: Positive samples used in training the face classifier and the eye classifier

Figure 4.5 illustrates the process of training a classifier based on Haar-like features and Adaboost algorithm. Once the training has finished, the classifier will produce an XML file that is the database that will be loaded to the object detector to determine whether a new image contains faces/eyes or not.

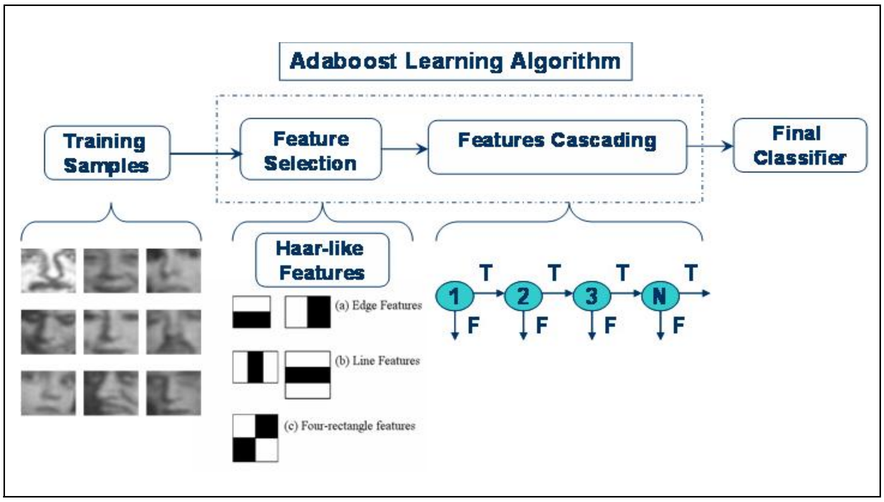
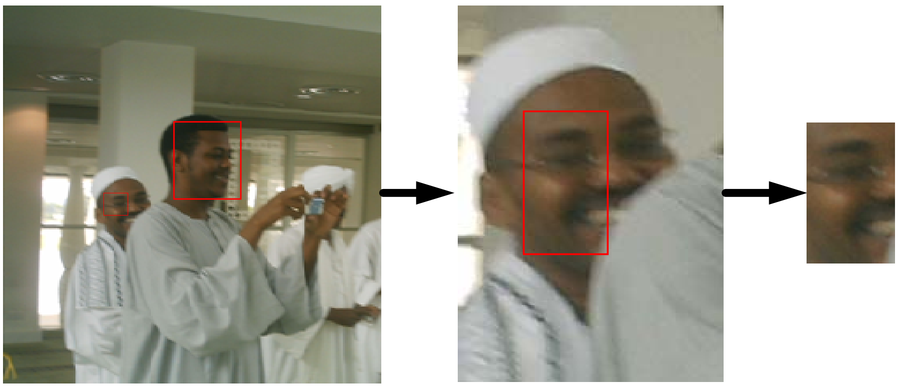


Figure 4.5: Training a classifier based on Haar-like features using Adaboost learning algorithm[6]

When applying the object detection to a new image, if the image contains faces or face components, the system will specify the region that surrounds the object. If it is a clear face, the region will be passed to the next component for further processing. On the other hand, if a face was not detected but instead the system found one of the face components i.e. the eye, it will specify the region of the eye and it will expand that region to cover areas above and below the eye to get a half face (if the half face is not occluded). The constructed region will be sent to the next component for further processing. Figure 4.6 illustrates constructing a half face from the detected eye.



(a) (b) (c)

Figure 4.6: a) Detect the eye, b) Construct half face, c) Extracted ROI of half face

[6] http://utarcvis.blogspot.com

As shown in figure 4.6, the eye in 4.6a is detected then the region above and below the eye was included in 4.6b as the region of interest and the half face construction was achieved in 4.6c. As for the other face in 4.6a which is a profile face, it will be directly sent to the classifier as it is clear.

4.2.3 Face Recognition

The face recognition task is to identify one or more persons or more in any given image. In this system, after the object detection has specified the ROI that contains the full face, half face or profile face, it will send this region to the face recognition algorithm which is PCA. To use PCA, the algorithm must be trained at the beginning with a set of images where at least one of the images in the set represents one person (face class). As mentioned in chapter 3, PCA will calculate the Eigen values on the dataset and constitute the face space, then compute weight space by projecting individuals onto face space. On the new image, the algorithm will calculate the weight vector by projecting onto face space and then classify the image by comparing the Euclidian distances on the trained faces from the new images. Figure 4.7 shows the output of the PCA on one image.

Figure 4.7: Face recognition with Principal Component Analysis

The image on the right is part of the training dataset while the image on the left is a test image. As shown, the PCA algorithm was able to detect the exact face which in fact is evidence on the robustness and accuracy of this algorithm in the face recognition task. Similarly, if the extracted region at the visual attention is half a face such as the image in figure 4.8, the system will be able to match the half face with its equivalent in the database and to recognize and display the full face. Figures 4.8 and 4.9 illustrate recognizing half a face.

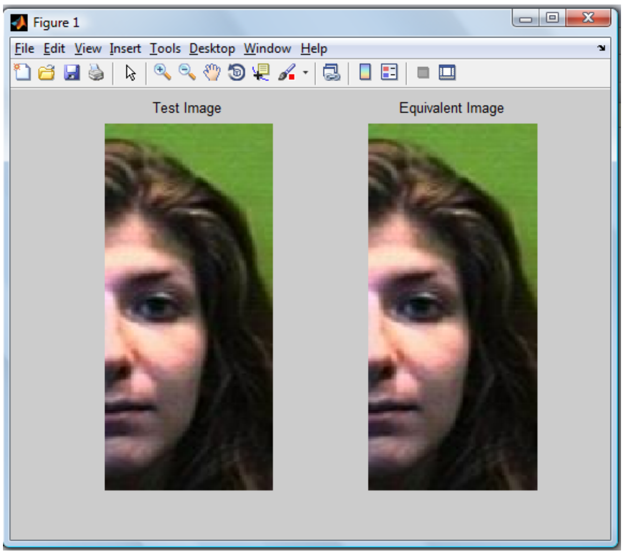


Figure 4.8: Matching half a face with its equivalent in the database

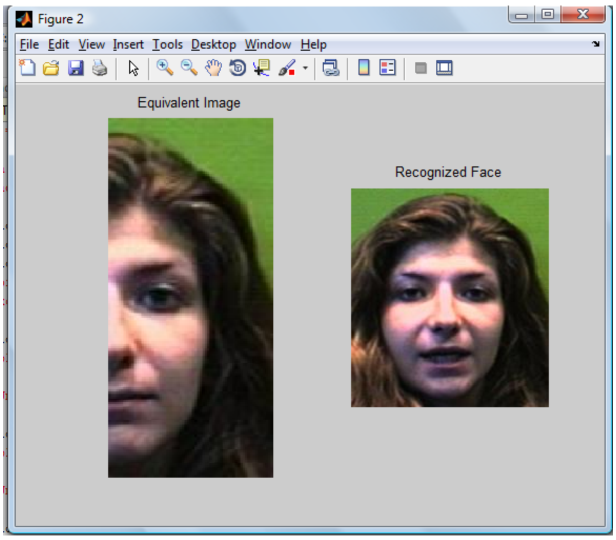


Figure 4.9: Recognizing full face from half a face

CHAPTER 5

RESULTS & DISCUSSION

5.1 Introduction

In this chapter, the results and analysis for the model will be discussed. The sections will discuss the results obtained for the object detection and the object recognition. Moreover, the ability of the model to detect partially occluded objects is demonstrated as well.

5.2 Object Detection

The first task by the system is to detect the intended objects. If the objects are not clear or affected by occlusion, the system will look for any element of the object and try to search for more features based on the element found. The system has been implemented in a face recognition system as shown in chapter 4. The first task is to look for a face or profile face or an eye. Then pass it to the recognition component. It has been found that the system performs well in detecting the faces and their elements i.e. profile face / eye. The system has been tested by applying it to 30 test images and the results obtained are shown in table 5.1

Table 5.1: Result of face and face element detection

Scenario	Number of testing images	Detected	Undetected	Accuracy
Full face	30	29	1	96.66%
Profile face	30	27	3	99.00%
Eye	30	26	4	86.66%

From the table, the accuracy of the detection for full face object is 96.66% and that is due to the complete features of the face that have been extracted and made it easy for the object detector to recognize most of the faces. For the profile face, the detection rate is 90% which is acceptable, the number of missed classified profile face is 3 which is normal in this kind of application where there should be false negative as the object is available but was not detected. For the eye detection, the accuracy is 86.66%. The reason behind this is the amount of clutter that affects the performance of the object detector. However, to some extend the system was able to detect the eye as it is part of the face. In feed-forward based model objects that are not clear in the image would not be detected.

The overall performance of the object detection is considered satisfying compared to the feed-forward based model where the performance of those models dropped to 74% (Kreiman et al. 2007) on unclear or partially occluded objects.

5.3 Object Recognition

PCA is used to get the region of the detected objects and perform the eigen values computing in order to obtain the exact person based on the face. In PCA, the algorithm should be trained first in order to calculate the eigen values for all the faces and add them to the database where these values will be used to compare them with new incoming images. Two datasets have been used in order to examine the PCA algorithms which are *face94 Face Recognition Dataset* (Spacek 2008) that has been developed by University of Essex and *MIT-CBCL Face Recognition Database*

(Weyrauch et al. 2004) that was developed at the Center for Biological and Computational Learning at MIT. Both datasets were tested to recognize faces under two main scenarios that are full face and half face. The full face or half face regions were obtained from the detection process (VAcomponent).

5.3.1 Face94 Dataset

Face94 dataset is part of a face recognition datasets that were developed for the purpose of training and testing face recognition algorithm for the computer science research projects at the University of Essex (Spacek 2008). The datasets contains images of 153 individuals divided into 20 female, 113 males and 20 male staff. For the purpose of testing PCA in this study, a total of 160 images were chosen out of the dataset that represents 8 male and 8 female individuals with 10 images each. The dataset was used to train the half face as well as full face. For the testing purpose, another 80 images were also chosen out of the dataset with 5 images per individusl. Figure 5.4 shows an example of the images used in the training phase in the face94 dataset.

Figure 5.1: Training images for PCA

The recognition component was tested for three scenarios: first when the face is recognized which means that it was among the training set; second, when the image is not available in the database, and lastly, when the image is not a face.

5.3.2 Face Available in the Database

Test images were used to test whether the algorithm could recognize the faces. The test images were part of the dataset that was used in testing. 80 images were used in the testing phase and 160 for training, and the algorithm was able to recognize 72 of them successfully with 90% accuracy. Figure 5.5 shows an example of recognizing face in the system.

Figure 5.2: Face recognition using PCA

Although both faces are not identical where the test part had some part occluded by adding some white area in the left side of the face image, yet the algorithm was able to recognize the equivalent image in the database. Moreover, the system was tested to recognize half faces and match them with the full face, and it was able to achieve. Figures 5.3 and 5.4 illustrate half face recognition.

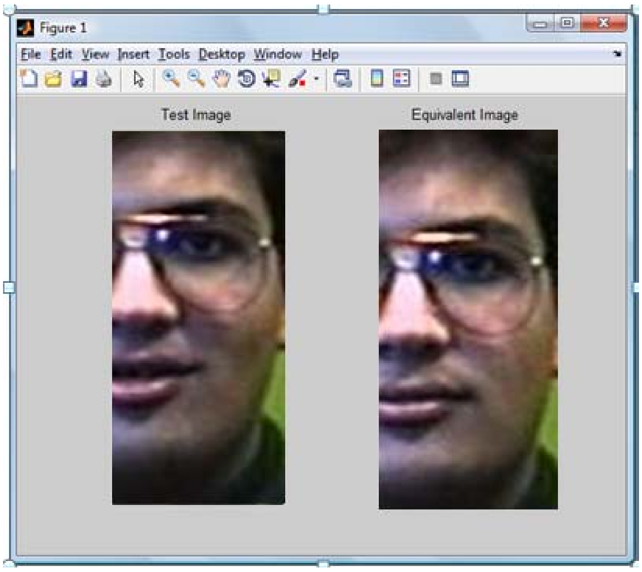


Figure 5.3: Half face equivalent in the database

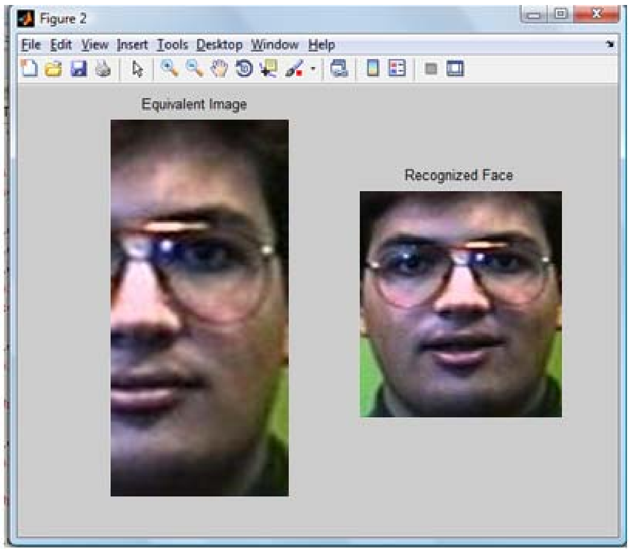


Figure 5.4: Half face matched with the full face

5.3.3 Face is not Available in the Database

The second scenario was to test whether or not the system is able to reject any face that is not available in the database. Out of 20 face images (that were not in the database) that were used in the testing, 11 images were identified to be in the database when they were not and the system displayed another image as the equivalent image. Figure 5.5 illustrates false recognition by the system.

Figure 5.5: False recognition

The system recognized that all the remaining 69 images were not in the database and displayed “unknown face”. The overall accuracy of this test is 86.25%. Figure 5.6 shows an example of correct recognition of an unknown image.

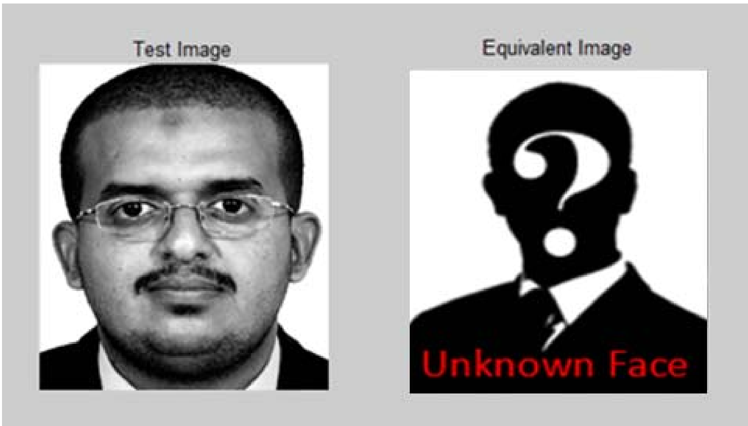


Figure 5.6: Recognition of unknown images

5.3.4 No Face in the Image

In order to test the capability of the system in recognizing face images only, a test to determine whether or not the system is able to recognize the non-existence of a face in an image was done. A test of 100 images that do not contain a face was done and the system was able to define 65 as non-faces. Figure 5.7 shows the capability of the system of identifying non-face images that could be passed to the algorithm from the detection algorithm.

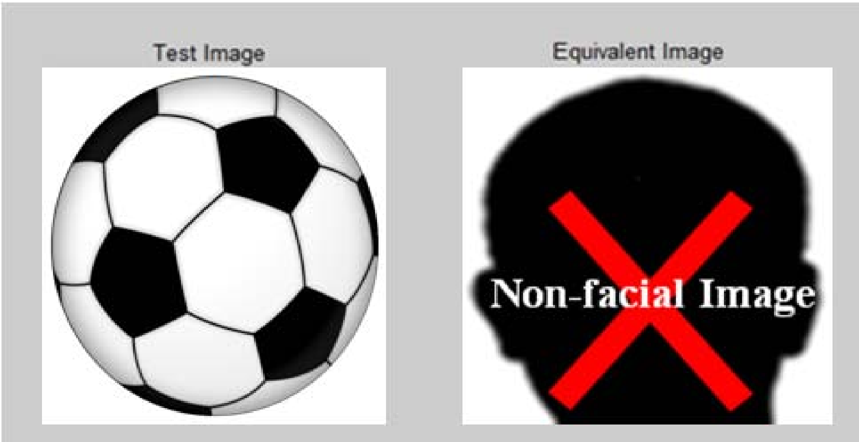


Figure 5.7: Identifying non-facial images

Table 5.2 shows a summary of the results obtained when the face94 dataset was used to test the system.

Table 5.2: Result of face recognition in the face94 dataset

Scenario	Number of training Set	Number of testing set	Number of corrected recognized	Accuracy
Face is available in database	160	80	72	90%
Face is not available in the database	160	80	69	86.25%
No face in the test image	160	80	65	81.25%

5.3.5 MIT-CBCL Face Recognition

MIT-CBCL face recognition dataset (Weyrauch et al. 2004) is another dataset that was used in this study to test the system. The data was developed at the Center for Biological and Computational Learning laboratory. It has 10 subjects and 2000 images per subject.

For the purpose of testing this system, a total of 500 images were used in the training set with 50 images per individual, and 100 images in the testing set with 10 images per individual. In addition, the same images were used to produce half images for both training and testing. Figure 5.8 shows an example of images used for the training of full face and figure 5.9 shows an example of the produced half face images that were used in the training phase.

Figure 5.8: Example of MIT-CBCL dataset for training full face

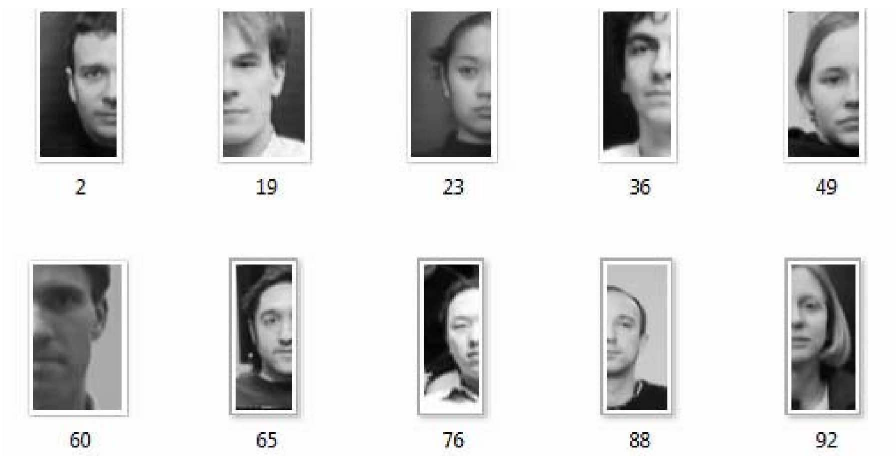


Figure 5.9: Example of the produce half face for training

The result of the system when it was applied to the MIT-CBCL face recognition dataset is summarized in table 5.3

Table 5.3: Testing of the system in MIT-CBCL face recognition dataset

Scenario	Number of training set	Number of testing set	Number of correctly recognized	Accuracy
Face is available in database	500	100	93	93.00%
Face is not available in the database	500	100	88	88.00%
No face in the test image	500	100	84	84.00%

As shown in table 5.3, the system was able to identify 93 images correctly out of the 100 images that were used in the testing phase for the MIT-CBCL dataset for full face and half face images which give 93% accuracy in this dataset. Figures 5.10 and 5.11 show the result of recognizing full face and half face respectively.

Figure 5.10: Result of full face recognition in MIT-CBCL dataset

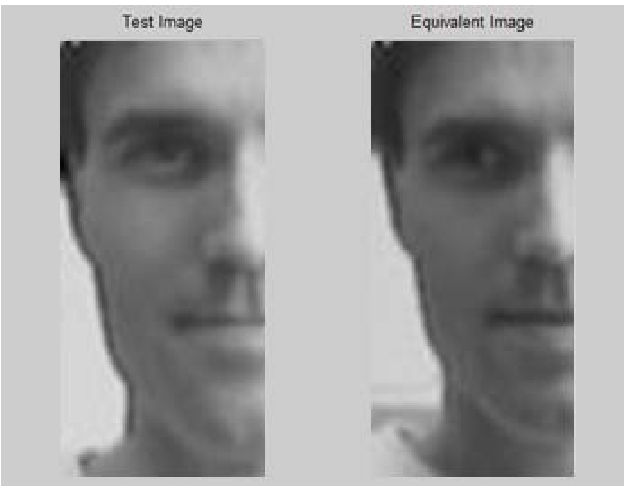


Figure 5.11: Result of half face recognition in MIT-CBCL dataset

Furthermore, the system was tested to identify faces that were not among the faces in the training dataset. Out of 100 images used in this test, the system recognized 88 images as not available in the dataset. As for the reset, the system wrongly matched them with images available in the training dataset. Figure 5.12 shows an example of wrongly recognized image.

Figure 5.12: False recognition of a face

Finally, the system was tested by subjecting it to non face images and it was able to recognize 84 images correctly as non face out of the 100 non face images that were used in this test.

5.3.6 Partially Occluded Images

The system was also tested on partially occluded faces, where the number of faces that were occluded was used as an input to the system. A small dataset of images that contains partially occluded faces under uncontrolled environment was collected. The purpose of these images was to illustrate the ability of the system to recognize objects in real situation. Figures 5.13 and 5.14 show an example of the images that were used in the training stage.

Figure 5. 13: Full face training set

Figure 5.14: Half face training set

Another set of images was used in the testing. The set contains faces of images used in the training stage which were partially occluded. The detection algorithm detected the eye and specified the ROI which was evaluated by PCA to determine whether a face existed or not and its availability in the database. In this test, 10 testing images were used to illustrate the capability of the system to perform the task. 7 faces were correctly recognized by the system. Figure 5.15 shows in example of the images that were used in the testing stage. In addition, Figures 5.16 and 5.17 illustrates an example of the system's performance in one of the images that were tested.

Figure 5.15: Example of testing images

Figure 5.16: Testing Image

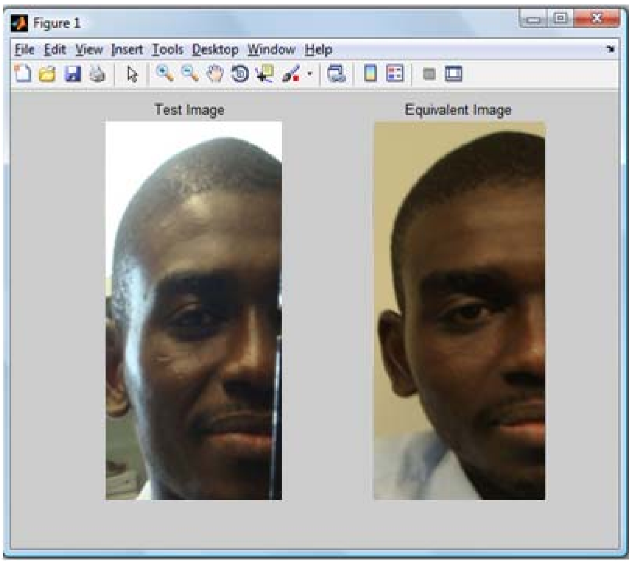


Figure 5.17: Detected half face and its equivalent

5.4 Summary

The results obtained in this chapter represent the proposed model in chapter 3 when it has been applied in a face recognition system. Two face recognition datasets were used, face94 from university of Essex and MIT-CBCL face recognition dataset from MIT. In addition, a small dataset was collected in order to test the capability of the system in recognizing partially occluded faces. The overall performance in the system demonstrates the capability of the integrated model of feed-forward and feedback processes in recognizing objects in complex scenes.

CHAPTER 6

CONCLUSION & FUTURE WORK

6.1 Introduction

This chapter concludes the work that has been presented in this thesis and summarizes some of the future works that could be done in order to enhance the model that has been developed.

6.2 Conclusion

As mentioned earlier, object recognition has been an interesting area of research that has attracted the attention of many researchers around the globe. Many methodologies have been employed in order to develop models and algorithms that are able to recognize objects. Researchers started in this area three decades ago and many algorithms have been presented. Most of these solutions developed to achieve object recognition were motivated by computer vision. Recently, a neuroscience research on the anatomy of the visual systems of primates and humans has led to the understanding of how the information is processed in the brain. Computer scientist mapped the functions of the visual system and designed biologically inspired object recognition systems. This research continued in exploring the findings of neuroscience and designed a model of object recognition based on the integration of two communication mechanisms that are being utilized by the human visual system. Feed-forward and feedback are two mechanisms of information passing between the visual areas in the brain. Previous works in biological vision presented models were based on the feed-forward mechanism. However, the models' performances were affected by the complexity of the images which they were subjected to.

With more evidence that support the opinion that the visual system integrates both feed-forward and feedback and with the potential of developing systems that could mimic the functions of the human visual system, a model of object recognition was presented in this work. The model integrates the functions of the feedback process with the feed-forward mechanism. Visual attention which helps humans to attend to important objects while ignoring others was mapped in this system as a function of the feedback process. Another function that was mapped is the recognition by components; where if the object is not fully visible, one or two components of that object could lead to recognizing it.

The model was implemented in a face recognition system. The results obtained have proven that the integration of the functions of the feed-forward and feedback helped in obtaining better results in complex scenes that contain partially occluded objects.

6.3 Contribution

This research work presented a model of object recognition based on the functions of areas of the ventral pathway in the human visual system. Previous models were based on the feed-forward or feedback mechanism. The model presented here is based on the integration of both feed-forward and feedback mechanisms of information communication among the different visual areas. The model employed the visual attention function as well as the recognition by component that the human visual system employs during the task of recognition.

6.4 Limitations

The work proposed in this thesis focused on the ventral pathway in the human visual system. The ventral pathway (or what pathway) is associated with object recognition and categorization. Another pathway in the human visual system is called the dorsal pathway (or where pathway) that is associated with object’s motion and location.

Both pathways complement each other during the task of perceiving the surrounding environment. The proposed model is able to recognize objects; however, it is not able to track objects during movement.

6.5 Future Work

Future improvement in this work might include the following:

- Apply the model in other application domains to further test its ability to recognize different sets of objects.
- Integrate some areas from the dorsal pathway (where pathway) in the human visual system to the existing model that could enhance its capabilities in tracking moving objects after they have been detected and recognized.

REFERENCES

Aravind I, Chandra C, Guruprasad M, Dev PS, and Samuel RDS. Numerical approaches in principal component analysis for face recognition using eigenimages; 2002 . pp. 246 - 251.

Bardski G, Kaehler A, and Pisarevsky V. 2005. Learning-Based Computer Vision with Intel's Open Source Computer Vision Library. Intel Technology Journal 9(2), pp. 119-130.

Bermudez-Contreras E, Buxton H, and Spier E. 2008. Attention can improve a simple model for object recognition. Image Vision Computing 26(6), pp. 776-787.

Biederman I. 1987. Recognition by components: A theory of human image understanding. Psychological Review 94(2), pp. 115-147.

Bongard J. 2009. Biologically Inspired Computing. Computer 42(4), pp. 95-98.

Bryliuk D, and Starovoitov V. 2002. Access Control by Face Recognition Using Neural Networks and Negative Examples. The 2nd International Conference on Artificial Intelligence. Crimea, Ukraine. pp. 428-436.

Cahoon TC, Sutton MA, and Bezdek JC. 2000. Breast cancer detection using image processing techniques. The Ninth IEEE International Conference on Fuzzy Systems. pp. 973 - 976

Frintrop S. 2006. VOCUS: A Visual Attention System for Object Detection and Goaldirected Search. Lecture Notes in Artificial Intelligence (LNAI). Bonn Germany: Springer Verlag Berlin/Heidelberg.

Graboi D, and Lisman J. 2003. Recognition by top-down and bottom-up processing in cortex: the control of selective attention. Journal of Neurophysiology 90, pp. 798-810.

Hubel D, and Wiesel T. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology. pp. 106–154

Ji Y, Chang KH, and Hung C-C. Efficient edge detection and object segmentation using Gabor filters; 2004. pp. 454 - 459

Khalifa O, Khan S, Islam R, and Suleiman A. 2007. Malaysian Vehicle License Plate Recognition. International Arab Journal of Information Technology 4(4), pp. 359-364.

Kim S, Jang G-J, Lee W-H, and ; ISK. 2004. How Human Visual Systems Recognize Objects - A Novel Computational Model. 17th International Conference on Pattern Recognition (ICPR). Cambridge UK. pp. 61-64.

Kreiman G. 2008. Biological object recognition. Scholarpedia 3(6).

Kreiman G, Serre T, and Poggio T. 2007. On the limits of feed-forward processing in visual object recognition. Journal of Vision 7(9).

Kveraga K, Ghuman AS, and Bar M. 2007. Top-down predictions in the cognitive brain. Brain and Cognition 65(2), pp. 145–168

LeCun Y, Huang FJ, and Bottou Le. Learning methods for generic object recognition with invariance to pose and lighting; 2004; Los Alamitos.

Lian Q-S, and Li Q. 2008. Object Recognition Based on Biologic Visual Mechanisms. Proceedings of the 2008 Congress on Image and Signal Processing. Sanya, China: IEEE Computer Society. pp. 386 - 390

Lienhart R, Kuranov A, and Pisarevsky V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection; 2003; Madgeburg, Germany. pp. 297-304.

Liu J, and Ma W. 2007. An Effective Recognition Method of Breast Cancer Based on PCA and SVM Algorithm. 1st international conference on Medical biometrics. pp. 57–64.

Louie J. 2003. A Biological Model of Object Recognition with Feature Learning: Massachusetts Institute of Technology.

Müller T, and Knoll A. Bioinspired early visual processing: The attention condensation mechanism; 2008 December 2008; Canberra, Australia.

Navalpakkam V, Arbib M, and Itti L. 2005. Attention and scene understanding. Neurobiology of Attention. pp. 197–203.

Paliy I, Sachenko A, Koval V, and Kurylyak Y. 2005. Approach to face recognition using Neural Networks. IEEE Workshop on Intelligence Data Acquisition and Advanced Computing Systems : Technology and Applications. Bulgaria. pp. 112 - 115

Riesenhuber M, and Poggio T. 1999. Hierarchical models of object recognition in cortex. Nature Neuroscience 2(11), pp. 1019-1025.

Riesenhuber M, and Poggio T. 2000. Models of object recognition. Nature neuroscience supplement 3.

Roorda A. 2002. Human Visual System - Image Formation. In: Hornak EJP, editor. The Encyclopedia of Imaging Science and Technology: John Wiley & Sons, New York. pp. 539-557.

Rosenholtz R, Li Y, and Nakano L. 2007. Measuring visual clutter. Journal of Vision 7(2), pp. 1-22.

Saalmann YB, Pigarev IN, and Vidyasagar TR. 2007. Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. Science 316(5831), pp. 1612–1615.

Serre T. 2006. Learning a dictionary of shape-components in visual cortex: comparison with neurons, humans and machines: Massachusetts Institute of Technology.

Serre T, Oliva A, and Poggio T. 2007a. A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Science. pp. 6424-6429.

Serre T, Wolf L, Bileschi S, Riesenhuber M, and Poggio T. 2007b. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), pp. 411-426.

Serre T, Wolf L, and Poggio T. A new biologically motivated framework for robust object recognition; 2004.

Serre T, Wolf L, and Poggio T. Object Recognition with Features Inspired by Visual Cortex; 2005 20-25 June; San Diego. IEEE Computer Society Press. pp. 994 - 1000.

Siagian C, and Itti L. 2007. Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2), pp. 300-312.

Smith L. 2002. A Tutorial on Principal Components Analysis.

Spacek L. 2008. Face Recognition Dataset.

Spivey JM. 1989. The Z notation: a reference manual: Prentice-Hall, Inc.

Tanaka K. 1997. Mechanisms of visual object recognition: Monkey and human studies. Current Opinion in Neurobiology 7, pp. 523-529.

Tripathi S, Singh L, and Arora H. 2009. Face Recognition Machine Vision System Using Eigenfaces. International Journal of Recent Trends in Engineering 2(2).

Turk M, and Pentland A. 1991. Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3(1).

Viola P, and Jones M. 2001. Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 511-I-518.

Weyrauch B, Heisele B, Huang J, and Blanz V. 2004. Component-based Face Recognition with 3D Morphable Models. Conference on Computer Vision and Pattern Recognition Workshop

Wilson PI, and Fernandez J. 2006. Facial feature detection using Haar classifiers. Journal of Computing Sciences in Colleges 21(4), pp. 127-133.

Xia K, Xu G, and Xu N. Lung Cancer Diagnosis System Based on Support Vector Machines and Image Processing Technique; 2006 December 18-20. pp. 143 - 146.

Zhang S, Wang J-h, Zhao S-g, and Luan X-j. 2007. Urinary Sediment Images Segmentation Based on Efficient Gabor flters. International Conference on Complex Medical Engineering, CME Beijing, pp. 812 - 815.

Zheng L, and He X. Number Plate Recognition Based on Support Vector Machines; 2006; Sydney, Australia, pp. 13.