APPEARANCE BASED GAZE ESTIMATION USING NEURAL NETWORK

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
ABDULMENEM ABUBAKR

In Partial Fulfilment of the Requirements for the Degree of Master of Science in Electrical and Electronic Engineering

NICOSIA, 2017

APPEARANCE BASED GAZE ESTIMATION USING NEURAL NETWORK

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
ABDULMENEM ABUBAKR

In Partial Fulfilment of the Requirements for the Degree of Master of Science in Electrical and Electronic Engineering

NICOSIA, 2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: ABDULMUNEM ABUBAKR

Signature:

Date:

ACKNOWLEDGMENTS

To my parents and family….

ABSTRACT

Gaze estimation and detection of the eye movements have obtained different accomplishments in the applications of human machine interface. Eye movements and gaze is one of the important interaction methods used by humans. They have ability to understand different things from the gaze and movements of the eye. Human eye movements can be implemented as an alternative for the hands in the conversation between human and computer. The interest in the gaze detection and its implementation to interact with digital devices is increasing rapidly. Artificial neural networks are drawing the future of artificial intelligence in the modern science. They are considered of the most promising, fastest developing and widely implemented artificial technologies. The implementation of artificial neural networks in the gaze detection researches has gained more support and interest in the last few years. It is been investigated and studied by many researchers all over the world. This work presents the use of artificial neural networks in the gaze estimation of different participants. MPIIGaze collected from Max Plank Institute of Informatik will be used to train and test the capability of the neural network for the gaze estimation. Different experiments will be carried out and their results will be printed and discussed.

Keywords: Artificial networks; back propagation; gaze; human machine interface; yaw angle

ÖZET

Bakmanın (uzun uzun bakma) nedenleri ve göz hareketlerinin yakalanması, insan makine bağlantısı uygulamalarında farklı başarılar elde etmiştir. Göz hareketleri ve bakma insan ilişkilerinde önemli bir yöntemdir. İnsanların bakarak ve göz hareketleriyle farklı şeyleri anlama yetenekleri vardır. İnsanın göz hareketleri bilgisayar ve insan arasındaki iletişimde ellere bir alternatif olabilir. Bakma nedeninin değerlendirilmesi ve digital araçlarla iletişimde uygulanması hız kazanmaktadır. Yapay sinir ağları yapay zekanın geleceğini modern bilime çekmektedir. Yapay sinir ağları en hızlı gelişen, yaygın olarak kullanılan ve ilerisi için umut veren yapay teknolojilerdir. Son birkaç yılda, bakmanın nedenleri araştırmaları daha çok destek ve ilgi görmektedir. Dünyada bu konuyla ilgili bir çok araştırmacı çalışılmalarını ve incelemelerini yürütmektedirler. Bu çalışma, farklı katılımcılar üzerinde, bakma nedenleriyle ilgili yapay sinir ağlarının kullanımını araştırmaktadır. Max Plank Institute of Informatik'ten alınan MPII Gaze, sinirsel ağların bakma nedenleriyle ilgili eğitim ve test yeterliliğinde kullanılacaktır. Farklı denemeler yapılacak ve sonuçları yayınlanıp tartışılacaktır.

Anahtar Kelimeler: Yapay ağlar, back propagation, bakma, insan-makine bağlantısı,yan açı

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN: Artificial Neural Network

CNN: Convolution Neural Networks

MPII: Max Plank Institute of Informatik

MSE: Mean Squared Error

x

CHAPTER 1

INTRODUCTION

1.1 Introduction

The eye of the human being is not just used for vision although it is its main purpose. Social sciences claim that the eye is one of the most powerful communication methods in human beings. Eye contact is well known for social scientists for its high performance during discussions, meetings, and public speech. When somebody poses the question, who is he? While looking at some person, we assume that the listener knows about whom the questioner is speaking even if he doesn't explain. The listener will most likely understand by studying the gaze of the listener without the need for any effort. The impact of eye contact on the listener is very well known between people.

Science is expanding in a very impressive manner and is continuously in the search for new ideas that could simplify the life of humanity. The interface between humans and machines has moved very large steps toward being more easy and efficient. Different types of computer keyboards, touch screens, microphone recognition systems, touch pads, and many other methods of interface have made our life easier. However, on the contrary to the machine human interface which became highly easy and developed; these methods are still in need for more and more development to increase the speed and efficiency of the human machine interface.

Nowadays, the development of new easier and more efficient human machine contact is attracting the interest of many researchers in different countries. They are continuously trying to develop new methods to make easier the control of computer device, typing, surfing, and using different machines. The use of eye movements to interact with computer and gaming devices has been implemented using different technologies. Although these types of technologies are very efficient, their main drawback is the need to fix special parts near the eye to detect the muscles movements. The new trend in the science is to try to use the gaze

movements detected by special cameras without fixing any parts on the body. Gaze detection algorithms are being widely studied and developed to detect and implement the gaze in the control of computer devices.

Artificial neural networks have been used for decades and developed throughout the years to become more efficient and flexible and to be suitable for different types of applications. Artificial neural networks are non linear systems that copy the function and structure of biological brain system and apply its logic on artificial basis. They have the ability to adjust themselves with different systems and learn common patterns in multiple targets. Neural networks are implemented nowadays in different fields of science and represent one of the main artificial intelligence methods that are widely used.

Back propagation learning algorithm of artificial neural network is a well known learning algorithm that offers high performance in the supervised learning of artificial neural network. It has been widely investigated to train neural networks and adjust them to perform different tasks. It has been considered one of the most efficient types of learning algorithm for decades. New algorithms are being implemented nowadays like Deep Learning algorithms. However, most of these algorithms are using the back propagation as their main learning tool. They are all considered as extensions of the original back propagation algorithm.

Artificial networks are being suggested to be used for the applications of gaze detection based on imaging systems. They are believed to have the capability to detect the gaze of the user under different conditions based on the image of the face and eyes. The detected gaze direction can then be used as the control of the intended device such as laptop or personal computer. The works on the subject are still in the development phase although different researchers claim their findings of good results in the same context.

CHAPTER 2

ARTIFICIAL NEURAL NETWORKS

2.1 Related Works

The interest to study the eye gaze activities in human to human discussions began around the 1960's. The researches have become more intensive and accurate in the past two decades due to the increasing needs of developing efficient and accurate eye gaze detection systems. Some researchers have shown that gaze can be used to produce synchronisation between people (Hugot, 2007). The eye gaze is now accepted to be very important means of communication between humans (Alnajar, Gevers, Valenti, & Ghebreab, 2013). Scientists have proved that looking at someone starts a communication channel with him; however, most of cultures consider gazing at somebody as an aggressive action (Fu, 2015).

In the 1930s, Tinker and his group started applying imaging techniques in the study and analysis of eye movements while reading. The group has noticed different responses of the eye and reading speed upon changing the size of font, page size, and other properties (Hartridge & Thompson, 1948). The major step in the development of eye tracking systems was the creation of head fixed eye tracker system (Moran & Card, 1983). More and more interest is given to the subject of eye and gaze tracking methods (Pomerleau & Shumeet, 1993). They are being studied to replace the actual human machine interface devices for disabled people and handicaps. Many researches are being generated and published yearly presenting new ideas on the use of different digital algorithms using the gaze as an interaction method with computers and machines.

3D gaze detection based on facial features tracking of one camera system was presented by (Chen & Ji, 2008). The face feature points are tracked in order to estimate the 3D gaze direction. The proposed algorithm needs one time calibration for the user. A high accuracy achievement was reported by this research article. In (Alnajar, Gevers, Valenti, & Ghebreab, 2013) a new method for automatic calibration of gaze detection system was presented. The

estimated gaze of the person vision is compared with similar ones to detect the gaze direction. The reported average accuracy of the system with 10 images was high and promising. In (Schneider, Schauerte, & Stiefelhagen, 2014); a person independent and calibration free gaze detection system was presented. The paper used the Direct Cosine Transform for face extraction. A manifold alignment algorithm was applied and Histogram of Gradient feature was also extracted in this work. The gaze estimation in the wild based on the appearance was presented in the work of (Zhang, Sugano, Fritz, & Bulling, 2015). This work has proposed a new approach in terms of database collection and algorithm implementation. Data base images were collected in real conditions and over a long period of time to ensure realistic estimation results. SURF face cascade method for face detection was first implemented in this work to identify the face in the image. Different algorithms were presented in this work like Random Forest, k-Nearest Neighbour, Adaptive Linear Regression, Support Vector Regression and others.

2.2 Artificial Neural Network

It can be simply explained as a structure built of interconnected processing cells or units that are called neurons. They were built and designed in an attempt to mimic the main function and construction of the biological brain. The neural network is designed through the capability to do processing in the aim of recreating or restructuring the human brain flexibility and power based on artificial means (Zurada, 1992). The artificial neurons are tied to each other through weighted links. The weights are numerical values that are thought of as memory of the artificial network. The weights of the neural network are adaptable with the task and can be adapted in order to produce the desired output. Neural networks are mathematical models of the human being neural networks. The compound and perfect organization of the brain makes it capable to carry out difficult tasks with help of bulky structure of neurons. An artificial neural network is composed mainly of the next components:

- ➢ Input layers
- ➢ Output layers
- ➢ Weights
- ➢ Biases
- ➢ Hidden layers

➢ Propagation functions

Each one of layers can contain one weight minimum that is totally connected with previous and next weights. The basic functional structure of the artificial neural network is presented in **Figure 2.1**. In the figure, input layers, hidden layers, output layers, weights, and interconnections are clearly presented. One can notice that each one of layers is constructed of a distinct number of weights. The weights of each layer are totally or partially connected with the other layers weights (Zurada, 1992).

The use and structure of artificial neural networks have developed through years since its first invention before the mid of the twentieth century. Different types and structures have been presented and discussed offering a vast variety of neural network structures and functionalities.

| Input layer | Hidden layers | Output layer |
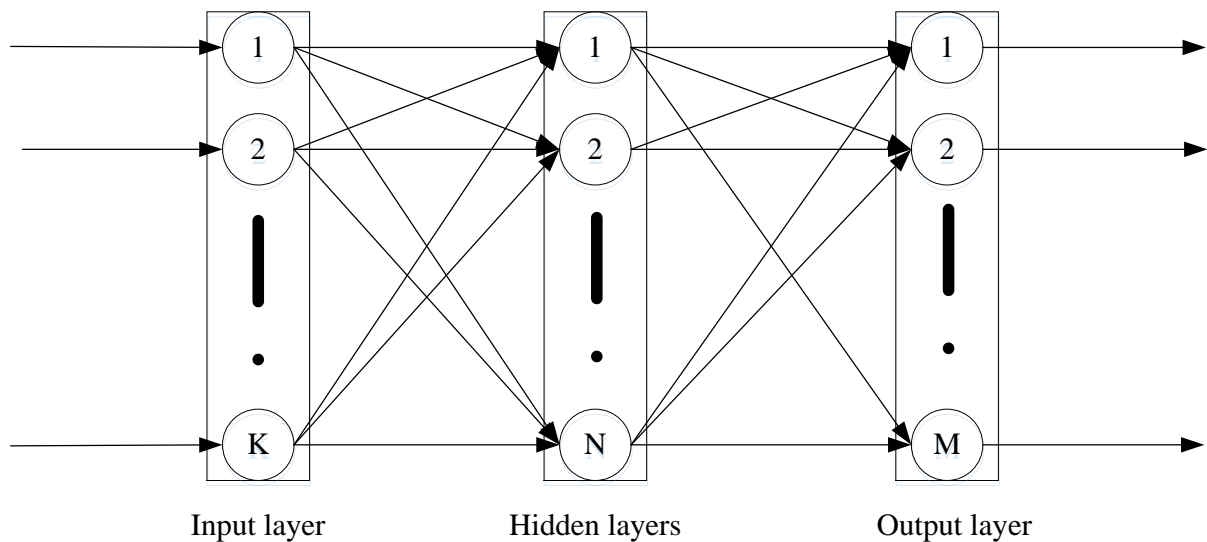
Figure 2.1: Basic architecture of artificial neural network

History of Neural Network

The history of the artificial neural networks goes back to the first decades of the 20th century. The oldest built mathematical representation of the brain as it was imagined in the form of artificial neural networks was implemented in the researches of Pitts and McCuloch in the year 1943. This model was also found in the papers of Hebb during the end of the fifth decade of the last century. Rosenblatt also have discussed this model in his works in the year 1958. A

research paper on the possible way that neurons may function has been prepared by Pitts and McCuloch in the middle of the last century. This paper was an attempt to give a brief description of the theory of human's brain. They imagined the brain as an electrical circuit and built a model of that circuit to explain their idea. Their model was actually basic but it was able to explain their idea in its simplest forms. In his papers, Hebb was sure that his theory about the strength of neural networks is true. He built his own theory dictating that the connections between different neurons become stronger after their implementation; it means that the neuron becomes stronger as much as it is being more used. His theory was an acceptable explanation of the way in which human brain can learn, understand, analyze and memorize information. The theory of Hebb is still accepted in the science and social fields as logical explanation of how brain works. The first neural network model was implemented and discussed in these studies and it was given the name of "computing mechanism or machine". This model was also known by the name of "learning with teacher" as it was similar to the process in which students in school learn new ideas and gain their knowledge.

 After these researches and theories were established and had their basic foundations; different researchers and scientists started to be more interested in the new established artificial intelligence theories. This can be seen in the researches of Widraw and Hoff in the year 1959. They offered two new model of artificial intelligence of their own design. The first one was called "Adaptive Linear Elements Network" and well known nowadays by the abbreviation "ADALINE" neural network. The second model that was invented by Widraw and Hoff is the "Multiple Adaptive Linear Elements Networks" or "MADALINE". The ADALINE network was sacrificed to be used in the case of binary pattern classification where the output is supposed to be binary. MADALINE model was an extended version of the binary ADALINE that can be applied in the real applications of artificial intelligence. There continuous researches led them to the invention of a new learning algorithm in which the results can be examined prior to the adjustment of the network's weights. This algorithm was presented in their works during the year 1962.

With development of researches in the same area, Anderson and Kohonen presented their works describing an intelligent system similar to ADALINE and MADALINE in 1972. In their systems, they employed matrix theories to simplify their systems. They built a vector of

ADALINE neurons responsible to control multiple outputs. They built their system in an analogue circuit form. After all these great inventions, the development of the artificial networks has gone slowly due to the lack of powerful personal computer technologies. After the year 1980, the researches on neural networks were launched again. They can be found in Hopfield's researches. Hopfield researches were pointed toward using bidirectional lines of connections to create a neural network mechanism (Clabaugh, Myszewski, & Pang, 2000). Similar ideas to that of Hopfield network were proposed in the year 1986. The ideas have invented the well known back propagation algorithm that is used in most of neural networks nowadays. It is considered an extended version of the basic Widrow Hoff algorithms of multiple layers. The back propagation is used to feedback an error function toward the layers of the network. The error function is used to apply slight modifications on the connections of the neural network. These modifications are done in a way that leads the network to converge toward the desired output. The back propagation suffers from slow rate of learning that needs thousands of loop iterations before it converges (Clabaugh et al., 2000).

2.3 Function of Artificial Neural Networks

The function of the artificial neural network is the realization of human like systems. Such systems are responsible to learn things, patterns, relationships, and classify them based on examples. A neural network is considered as transfer function that is able to understand the relationship between inputs and outputs. Being provided with suitable examples, the neural network will be responsible to update itself according to these examples and to learn the pattern in these examples. After learning the patterns from the examples, it is supposed to be able to generalize these patterns for unknown sets of inputs and generate their correct outputs. One of the critical components of the ANN is the learning function. There are many different learning algorithms that depend on the training method used of ANN.

2.4 Similarity with Human Brain

The artificial neural network is an attempt to imitate the structure and task of the biological brain. It is using the same analogy and organisation of the human brain. The biological brain is constructed from a huge number of neurons. The neurons of the brain are interconnected and exchange information between each other. Each neuron of the brain nerves is tied with more than

tens of thousands of neurons. **Figure 2.2** presents a part of the human brain that shows the interconnection between neurons. The dendrites connect to the lines and the axons which are connected to the neural cells. The information is transmitted between neural cells in the form of electrochemical reactions. These reactions are transmitted to the body of the neural cell causing it to fire an activation signal. The activation signal is transmitted to the next cell (Roberts, 2015).
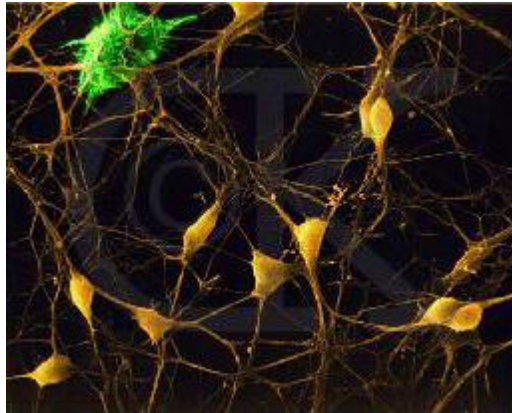
Figure 2.2: Small part of a biological brain structure (Clabaugh et al., 2000)

Figure 2.3 shows the different components of the biological neuron in the human brain. Different parts of the neuron are presented in the figure. The figure shows the cell body, the axon, the dendrites and the synapses of the neuron.
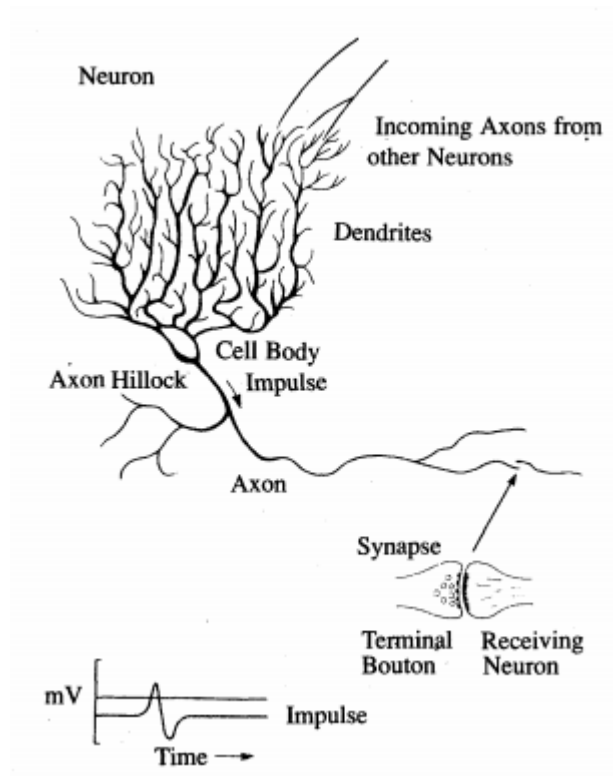
Figure 2.3: Components of the biological neuron (Clabaugh et al., 2000)

2.5 Artificial Networks

As said previously, artificial neural networks were inspired from the way human brain does the mental thinking. The main motivation for this inspiration is to create an intelligent model that can help solving complex scientific problems. Different functional structures of the artificial neural networks are analogous to the biological nervous structures because they all require suitable training before being able to give results (Kiran, 2009). The neural networks find the sum of their different input signals to generate the proper output like biological neurons does. Suitable threshold or transfer functions are used to determine the strength of the output. If the generated signal is strong enough, it will be passed to the transfer function. The basic structure of a simple neural network grid is illustrated in Figure 2.4. Summation function, activation function, weights, inputs, and outputs are all shown in this figure. The actual output of this neuron is the output generated from the output function. The summation function of the neural network is given by (Colin, 1996):

$$TP = \sum_{k=1}^{N} \omega_k x_k \qquad\qquad (2.1)$$

Where; TP is the total potential of a neuron, each input signal x_k in this function is associated with a neural weight ω_k that decides the importance of this signal in generating the final output.
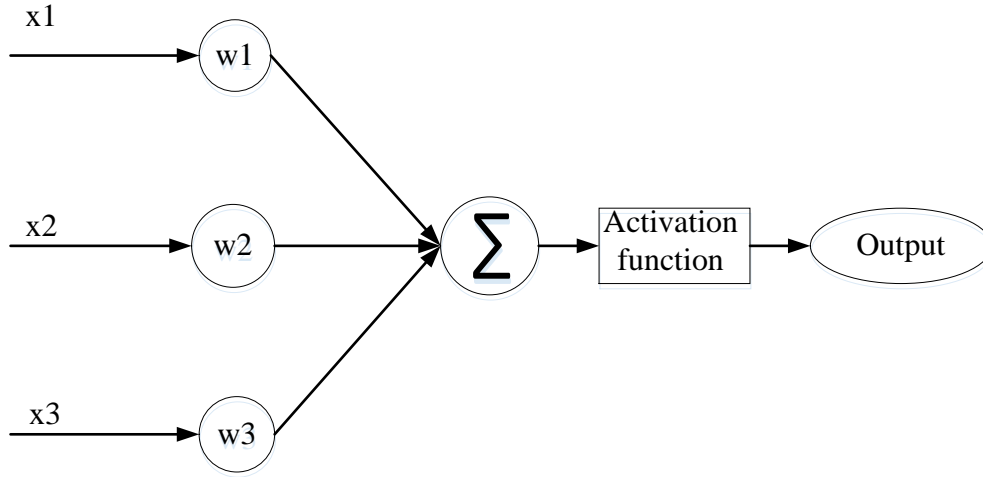
Figure 2.4: Basic structure of artificial neural network

2.5.1 Multi layer perceptron

Multilayer perceptron artificial networks are very famous for their simplicity and good efficiency. They are well known for their flexibility, and the possibility of their programming in different high level programming languages. The multilayer perceptrons are a type of supervised feed forward structures. The multilayer perceptron is constructed mainly of three important parts. These parts are mainly: weights, layers, and activation functions. These three parts are connected together with the help of the training rules that control the way that the neural network can update itself. For better understanding of these different components of a neural network, each one of them is going to be discussed separately (Anderson & McNeill, 2010).

2.5.2 Layers of ANN structures

The artificial networks are built up using connecting weights between functional parts that are called layers. Layers contain the important functions that generate the output of the neural network. The signals move between different layers through the weights. Mainly there are three different layer types based on the position in which they exist although they all do the same functions. These three types are:

➢ Input layers: they are the first layers in the artificial structure that are connected directly with the inputs. The rule of these layers is the submission of input signals to the next parts of the ANN. They don't process the input signal; thus the output O_i of these layers is equal to their input I_i and can be given by:

$$O_i = I_i \tag{2.2}$$

➢ Hidden layers: hidden layers can be considered the heart of the neural network structure. It is built of different processing layers. They ensure the connection between the input layers and the output layer of a network. The inputs to the hidden layers are found through the multiplication of input layer's output with the hidden weights. They can simply be given by:

$$I_h = \sum_n \omega_{hn} O_{in} \tag{2.3}$$

Where; ω_{hn} is the hidden weight matrix, O_{in} is the output of the first layer, and I_h is the input of the hidden layer. The output of each hidden layer O_h can be found through:

$$O_h = f(-I_h) \tag{2.4}$$

Where; f is a transfer function that has different types and properties. Different types of transfer functions will be discussed and presented later in this chapter.

➤ Output layer: These are the final layers in the artificial networks. There inputs are received from the hidden layers while their outputs are submitted to the output of the network. They process the information before submitting them to the final results. The inputs to the output layer I_o are given by:

$$I_o = \sum_i \omega_{oi} O_{hi}$$ (2.5)

Where; ω_{oi} is the weight matrix of the output layer, and O_{hi} is the output vector of the last hidden layer. Other transfer functions are applied on the input of the output layers whose output is the final neural network's output of the system. **Figure 2.5** illustrates the structure of layers in the artificial neural networks. It is important to mention here that the input and output layer sizes are related to the sizes of input vectors and output vectors of the network. The hidden layer sizes and number of hidden layers is an arbitrary choice that needs to be modified time after time while doing different trainings of the network until finding an optimum choice that gives better results.
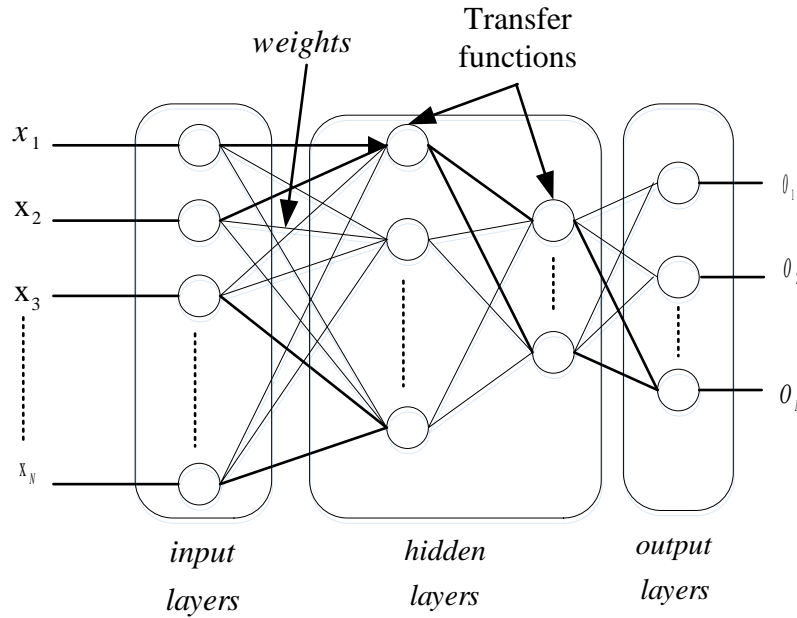
Figure 2.5: Layers structure in the artificial neural network

2.5.3 Synaptic weights

The weights of neural network are the memory that saves information about the different processes. The memory of the network is initialized arbitrary with random values before a planed training is applied. During the training of the network, weight values are updated to find the desired output. The final weights or memory are then considered to be true and saved for future uses. In the mathematical representation of the weights in neural network, they are represented in the form of matrix of weights.

2.5.4 Transfer functions

As it was mentioned in previous part of this chapter, transfer functions are one of the important parts of the different layers. They are used to generate the output of the neuron starting from its inputs. Transfer functions are responsible to generate an image of the input of the neuron by deciding whether to activate or deactivate the output. Some transfer functions have soft output curves while others are sharp in their output features. There are different types of transfer functions:

2.5.4.1   Hard limit transfer function

In this type of transfer functions, the output is binary and can take one of two different values. That means it can either be true or false. When the sum of the input signal to the layer is greater than a given value called threshold, the output changes its state. If the sum is less than that value, the output returns to its initial state. The output function is given by (Zurada, 1992):

$$f(O) = \begin{cases} 0, & O < \theta \\ 1, & O > \theta \end{cases} \tag{2.6}$$

The form of this type of transfer function is illustrated in **Figure 2.6**. The hard limits were used earlier in activation of artificial neural networks for their simple structures. Different other types of activation functions can be used nowadays and replace this type of functions.
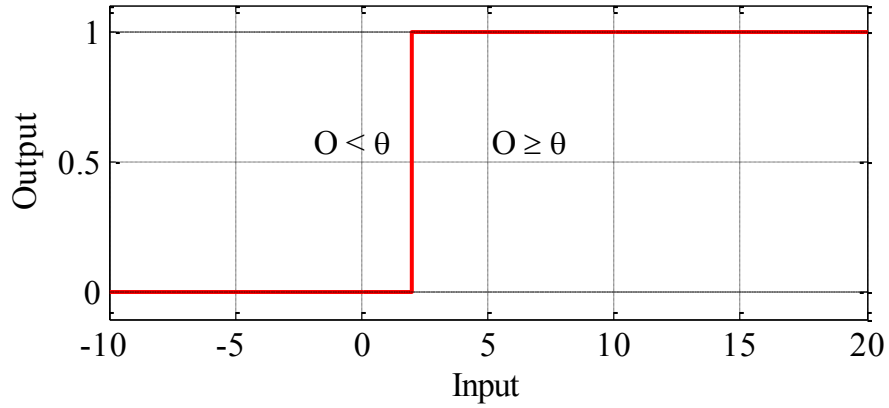
Figure 2.6: Hard limit transfer function general curve

2.5.4.2 Sigmoid transfer functions

This type of functions is an enveloped function that can range between two values. These two values can be 0 and 1 or -1 and 1. The main examples of these functions are logarithmic hyperbolic sigmoid and the tangential sigmoid. **Figure 2.7** and **Figure 2.8** illustrate the curves of these two functions. The output of these functions can be flattened or compressed by varying the constant or slope of the function. The mathematical description of the two functions is given by (Zurada, 1992):

$$f(O) = \frac{1}{1+e^{-aO}} \quad , \log arithmic \tag{2.7}$$

Where; "a" is a factor that control the shape of the curve, and "O" is the input value of the transfer function.
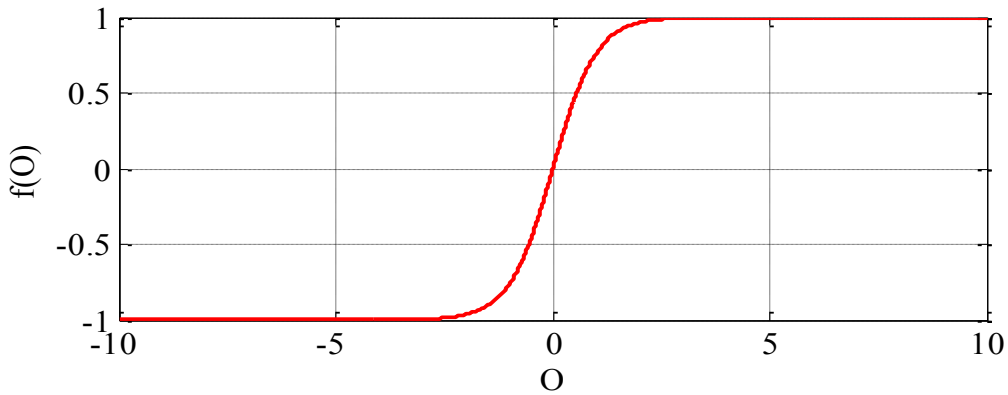
Figure 2.7: Tangent sigmoid activation function curve

14

$$f(O) = \frac{1 - e^{-aO}}{1 + e^{-aO}} \qquad , \text{Tan } gential \qquad\qquad (2.8)$$
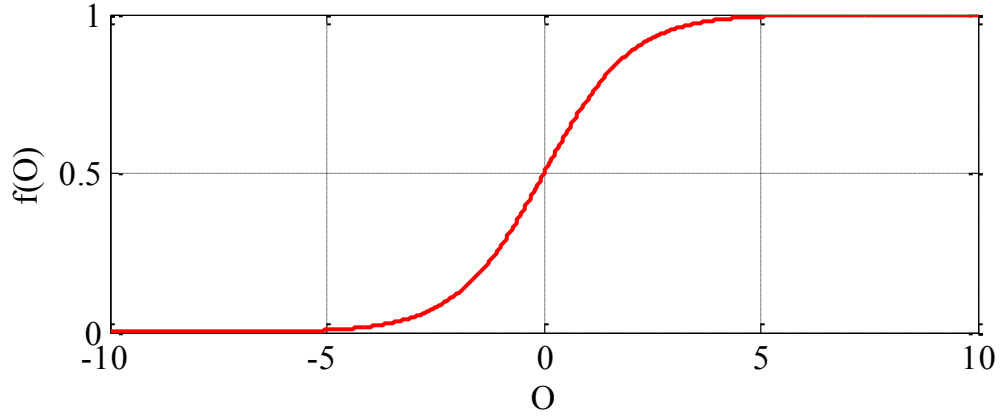
Figure 2.8: Logarithmic transfer function

2.5.4.3 Linear transfer function

In this type of transfer functions, the output is a linear function of the input as illustrated in Figure 2.9. It shows a linear saturated transfer function whose upper limit is 5 and lower limit is -5. Beyond these two limits, the output of the transfer function remains unchanging.
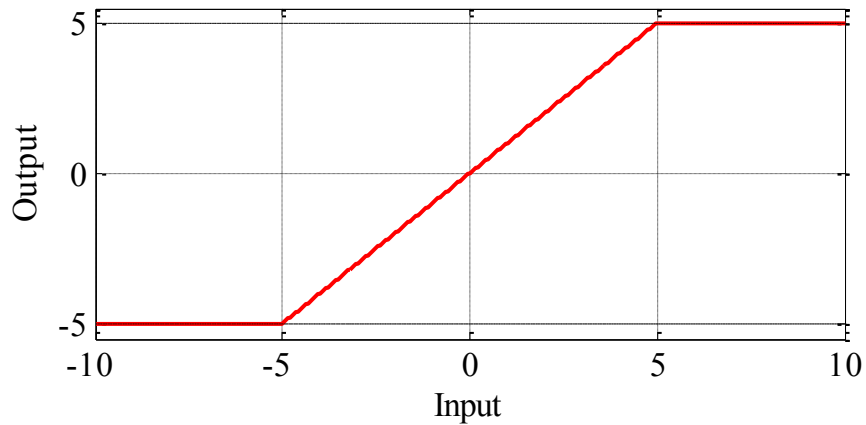
Figure 2.9: Linear saturation transfer function

The linear transfer function is described mathematically by:

$$f(O) = \begin{cases} -K & O \leq -K \\ aO & -a*K \leq O \leq a*K \\ K & K \leq O \end{cases} \qquad\qquad (2.9)$$

15

Where, a and K are different constants that control the behaviour of the function. It is important to mention that sigmoid functions are the most used due to their good performance and easier derivation compared to the other types.

2.6 Error Back Propagation Training Algorithm

The teaching process of a neural network is known as training. There are different training algorithms of the ANN among which the back propagation algorithm is very well known and famous for its high accuracy. It uses a feed forward iteration followed by an error back propagation in the inverse direction to update the neural weights. The development of ANN in the last two decades of the twentieth century is the result of the invention of back propagation learning algorithm. The training of back propagation network is complex and expensive in terms of processing; however, it is capable of simulating functions with different accuracy levels (Gupta, 2006).

The principle of the back propagation algorithm is simple. The inputs of the network are passed through different layers to the output layer. In the output layer, the results are compared with the target outputs. The error is found and used in the update of the different layers weights. The update function is based on the idea of error minimization that implies the derivation of the output signal and error functions. Small variations in the weights are calculated and applied to the actual weights. The new weights are then used in doing a new iteration to find the new error value. This process continues in a loop until the error is minimal and acceptable. It is important to notice that the error back propagation is done in each layer of the network separately. It means that the propagation starts from the last layer were the error between targets and actual outputs is propagated back to the output layer. The gradient of the last layer transfer function is used with the final error to find the error of the previous layer. This error is then propagated to the previous layer. Another error is calculated in series until the propagation reaches the first layer of ANN (A.D.Dongare, R.R.Kharde, & D.Kachare, 2012; Seiffert, 2002).
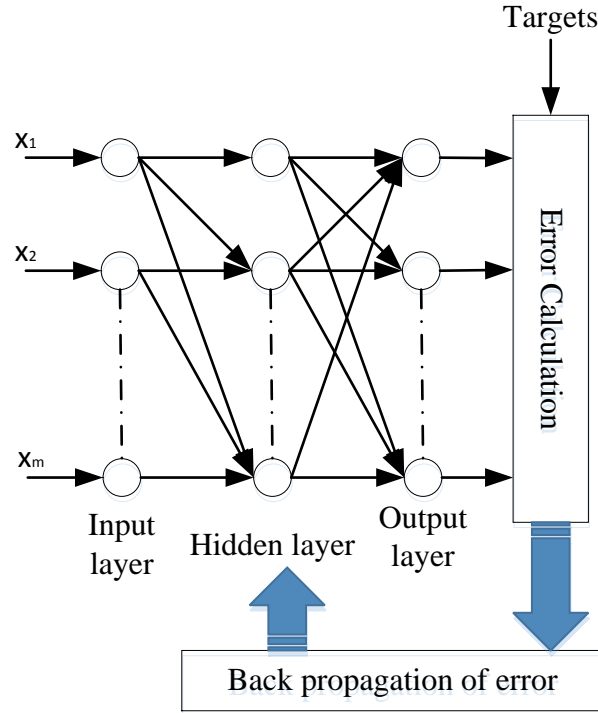
Figure 2.10: Error back propagation process

2.6.1 Model of back propagation algorithm

The back propagation algorithm is designed after the theory of gradient descent that search for the minimal mean squared error. In order to reach the minimum error value, the gradient or derivative of error is to be calculated prior to every step of weight update. This creates a constraint on the function of the error; that is, it should be continuous and derivable. Sigmoid functions represent the best choice for training of the ANN using back propagation networks due to the ease of finding their error gradient and derivative. The logarithmic sigmoid is defined by (Zurada, 1992):

$$f(O) = \frac{1}{1 + e^{-aO}} \tag{2.10}$$

The derivative of this function is given by (Zurada, 1992):

$$f^{\backslash}(x) = O(x)(1 - O(x)) \tag{2.11}$$

The input of the transfer function can be defined by (Zurada, 1992):

17

$$O = \sum x_n \omega_n + b_n \qquad (2.12)$$

Where the vector x represents the input values; the matrix w is the weights matrix, and the vector b is a special bias vector. The activation function is applied to the calculated output before being transmitted to the next layer. The tangent activation function and its derivative are both defined by (Zurada, 1992):

$$f(O) = \frac{e^O - e^{-O}}{e^O + e^{-O}} \qquad (2.13)$$

$$f'(O) = 1 - \frac{(e^O - e^{-O})^2}{(e^O + e^{-O})^2} \qquad (2.14)$$

When applying the transfer function of the output layer, its output is the output of the neural network that should be equal to the desired output. The error can be calculated at this step to start propagating back the update values of the network.

$$E = \sum_{j=1}^{n} (T_j - f_j)^2 \qquad (2.15)$$

The term T of this equation is referred to the desired outputs and the term f is output of the transfer function. Based on this equation, the gradient of the error can be calculated using:

$$\Delta_j = (T_j - f_j) f_j (1 - f_j) \qquad (2.16)$$

And the gradient of error for the hidden layers is given by (Zurada, 1992):

$$\Delta_h = f_h (1 - f_h) \sum_{j=1}^{n} \omega_{jh} \Delta_j \qquad (2.17)$$

Where; f refers to the output of the layer defined by the index of f, index i refers to the input layer while index h refers to the hidden layers. Each gradient is propagated back to the previous layer and a new weight values are generated. The new hidden weight values are generated using (Zurada, 1992):

$$\omega_{jh}(t+1) = \omega_{jh}(t) + \eta \Delta_h f_h + \alpha(\omega_{jh}(t-1) - \omega_{jh}(t-2)) \qquad (2.18)$$

The weight of the output layers are updated using the equation (Zurada, 1992):

$$\omega_{jh}(t+1) = \omega_{jh}(t) + \mu \Delta_j f_j + \alpha(\delta \omega_{ji}(t)) \qquad (2.19)$$

$$\Delta_j = f_j(1-f_j)\sum \omega_{ji} \Delta_j \qquad (2.20)$$

Where; the indexes h and i refers to the hidden and input layers; the index j refers to the weight number in the layer, the term Δ defines the error of the jth neuron in the considered layer, the parameters μ and α are the learning rate and momentum factor respectively. $\delta \omega_{ji}(t)$ is the weight variation of the last iteration, $\omega_{jh}(t-1)$ is the weight value in the previous iteration and $\omega_{jh}(t-2)$ is the weight value before the previous iteration.

CHAPTER 3

HUMAN COMMUNICATION AND THE EYE GAZE

3.1 Introduction

The eye is used for more than looking at things and collecting visual information about the environment around us. Human eye is a very important communication part of our body. In social studies, the eye is considered one of the key stones of revealing the personality and making communication between people. Specialists can read the signs emitted from the eyes and translate it into useful information about criminals. Different kinds of people are able to transfer special codes and information throughout the movements and signs of their eyes.

Suppose being in a place with 5 people namely, Anne, James, Malcolm, Munem, and Samet. If someone asks what sport you prefer while looking at Malcolm; everybody will understand that the question is forward to Malcolm. If the same question with the same tune and verbal expressions was posed while looking at Anne, everybody will understand that the question is for Anne although her name was not pronounced. This is due to the fact that humans can detect the gaze easily and understand it as a mean of communication.

When people talk to a person they generally address him or her just by gazing at him or her. When we gaze at somebody without reason it can be considered impolite and creates troubles. These ideas show how the human beings can communicate through their eyes without even moving their lips. Generally, people need to keep controlled their eye gaze because they know that others are aware of their gaze directions (Drewes, 2010) (Mehmood, 2010).

The interest to study the eye gaze activities in human to human discussions began around the 1960's. The researches have become more intensive and accurate in the past two decades due to the increasing needs of developing efficient and accurate eye gaze detection systems. Some researchers have shown that gaze can be used to produce synchronisation between people (Hugot, 2007). When coming to comparison between human eyes and animal eyes, it is noticed that the white part of the human eye is distinct.

Figure 3.1 shows the distinction between the human eye and chimpanzee's eye. Eyes of mammals, that are the most similar of human eyes, doesn't have visible white eyeball around their pupils. This makes more difficult to find the direction of animal gaze compared to human. The eye gaze is now accepted to be very important means of communication between humans. Scientists have proved that looking at someone starts a communication channel with him; however, most of cultures consider gazing at somebody as an aggressive action (Fu, 2015).

(a) Human's eye

(b) Chimpanzee's eye

Figure 3.1: Eye's of human vs. eye of Chimpanzee (Drewes, 2010)

It is clear that the humans use the eye for more than looking at things. They are used for communication between each other. This leads to the fact that humans can use their eyes as output devices to communicate with humans or machines. When looking at a person while speaking that indicates that we are speaking to him. The fact that eyes can be used as output can be implemented for communication between humans and machines. For long time, eye was used as input that reads or sees the information produced by the machines. An eye tracking process is important to ensure the transmission of information from humans to machines or computers. In digital mobile phones, this is used actually as an energy saving property that can help the mobile to determine whether the user is gazing at the screen or elsewhere. The mobile then turns on or off the screen to reduce power consumption.

Computers also can determine whether the speech control is pointed toward it or the person is speaking with someone else (Drewes, 2010).

3.2 Eye Tracking History

Eye tracking systems are used to track the eye gaze for different purposes. The science of eye movement detection and eye tracking is more than 100 years old. Eye tracking method based on light reflection from the eye cornea was developed by Cline and Dodge. The eye position was recorded onto a photo plate.

In the 1930s, Tinker and his group started applying imaging techniques in the study and analysis of eye movements while reading. The group has noticed different responses of the eye and reading speed upon changing the size of font, page size, and other properties (Hartridge & Thompson, 1948). The major step in the development of eye tracking systems was the creation of head fixed eye tracker system (Moran & Card, 1983). This type of eye trackers are still widely implemented and developed for eye tracking purposes. A Russian psychologist called Yarbus has made important contributions in the field of gaze surveying. He studied complex images during the 1950s and 1960s and registered the eye movements of a person gazing on different objects (Mehmood, 2010).

Nowadays, more and more interest is given to the subject of eye and gaze tracking methods. They are being studied to replace the actual human machine interface devices for disabled people and handicaps. Many researches are being generated and published yearly presenting new ideas on the use of different digital algorithms using the gaze as an interaction method with computers and machines. This work itself is a contribution to the efforts for finding efficient gaze detection and tracking method.

3.3 Eye Gaze as Computer Input

An eye based human machine interface is believed to be a main player within the modern interface techniques. It can be easier and more flexible than the actual interface methods we use nowadays. It is also very useful to help disabled people to interact easier and better with machines. Handicaps who cannot move their body parts implement their eye gaze to interact with people around them. They can also use human gaze controlled systems to interact with computers and other machines. Such systems offer a very suitable and easy solution for these

people to use computers and smart devices. However, for the moment, these systems are less accurate and slower than other interaction methods. The efficiency of these methods is still a challenging issue in the field that is under study and development. This is mainly caused by the difficulties of capturing the images and detecting the eye movements accurately (Drewes, 2010). Many systems that analyze eye movements implement specialized hardware and software. The use of infrared light source is one example of such structures. The eye gaze interaction system can present many advantages such as:

It is easy to be used as it makes use of the eye movements and implies no hand movement. This will reduce the stress on the muscles of the hand and arm. The eye movement is further fast and easier than hand movement. The eye movement doesn't cause any extra loading of its muscles as they are in continuous movement.

The interaction between the eye and the machine is pretty faster and can offer new dimensions for the interaction speed. Keyboard typing can be very fast compared with the actual typing speeds. Clicking, paging, and moving inside the screen can be very much fast and easier using eye gaze.

The use of such systems reduces the maintenance costs and decrease the hygienic demands, like an operation room, an eye gaze interaction system can be very beneficial as it offers free of contact interaction means.

With the existence of high resolution cameras, the eye gaze tracking and detection algorithms can offer the possibilities of remote controlling devices through eye contact. The detection of eye movements over distances of many meters is possible with high density and high zoom cameras. The control of TV and air conditioning devices through the eye seems to be a very nice application of eye detection.

The safety is another benefit that can be guaranteed through the use of eye detection. The eye tracking implies that the user works with his attention focuses of the work being done. The interaction will be disconnected once the user's attention is not detected. This can be very useful while driving a car, or controlling a heavy machine.

On the other hand, the eye detection has its own disadvantages and drawbacks in terms of interaction possibilities. One of these drawbacks is the lack of controllability of the eye movements. The eyes are moving continuously with or without the conscious control of

person. They can disturb the function of interaction when moving unconsciously in any direction. Profound researches need to be done to evaluate the degree of interaction that can be done using the eyes.

Another drawback of this system can be the fatigue of eye muscles if used too much as an interface with machines. Some special injuries can happen to these muscles as a result of the repetitive use of them. This problem needs to be explored in more details before using an eye based interaction system.

The last problem of the eye interaction system is the conflict that can happen between the vision function and interaction function of the eye. The eye is known for its role as a visual sensor that captures the surrounding events and presenting them to the brain. This function implies the movement of the eye in all directions. This makes difficult for an algorithm to decide whether the person is interacting with the system or just doing an observation task. In such situations, an exact separation between the natural movements of the eye and the intentional movements is a must.

3.4 Technology of Eye Tracking

In general, three methods of eye tracking can be distinguished. In the first approach, a sensor can be fixed at the eye to detect its movements. This problem is a bit risky and can cause fatigue of the eye muscles due to the contact with the sensor during movements. Eye lenses can also be used for this purpose with less risk and good performance. This method is very accurate and implemented in medical applications. **Figure 3.2** shows an example of the sclera contact lenses used to detect the eye movements.

Another method of eye movement detection is to fix electromagnetic sensors on the skin around the eye. These sensors can detect the dipole caused by the eye movement. **Figure 3.3** presents this method and the used sensors.
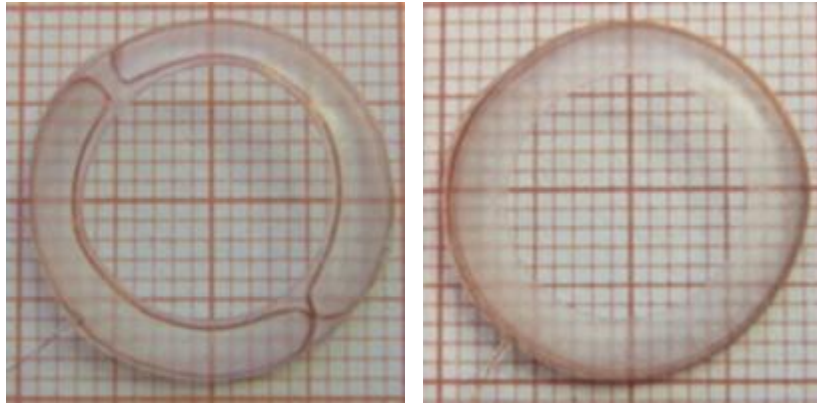
Figure 3.2: Scleral contact lenses used for eye movement detection

Figure 3.3: Electromagnetic sensors around the eye (Mehmood, 2010)

The above mentioned methods are accurate and able to detect every movement of the eye. However, these two methods are not suitable to be used for communication between human and machine. The need for fixing sensors on the eye Iris or around the eye is a major drawback of such solutions. The third revolution idea for human machine communication is the video. A real time image processing connected to the device's camera is the core idea of this method. The algorithm will be responsible to distinguish the eye and its pupil. Then the algorithm will find the gaze direction and translate it to an action or instruction. This idea is advantageous over the other methods because the user will not need to fix any parts on his body. For that reason, it is the best choice for human machine interaction.

3.5 Eye tracking in the video stream

The purpose of a video stream eye tracking is mainly the detection of the direction of gaze in a picture. The picture is captured using a camera that shots camera stream. This can be done by detecting the iris based on the contrast between the iris and the white of the eye. This method suffers the position of the eyelids that cover the iris from up and down. The effect of the

eyelids reduces the detection accuracy of the vertical position. The detection of the eye pupil is considered as promising method as it can offer better results. The pupil of the eye can be detected via two extinctive methods; these are the dark pupil and the bright pupil method. In the dark pupil method, the position of a black pupil in a camera image is detected. The bright pupil detection idea uses reflected IR light from the retina. This reflected infrared make the pupil appear white in the image of the camera. This effect is defined by the term red eye in photography. This method requires an infrared light that comes from the direction of the camera.

3.6 Anatomy of the Eye

Before studying the eye gaze, any research needs to be aware of the construction of the human eye. The knowledge of the eye is a must to build an accurate and efficient gaze detection system.

In an easy inspection and from the point of view of engineering, the human eye and the eye muscles can be considered as digital camera with stabilising system. The eye is connected to the head through six muscles as illustrated in **Figure 3.4**. These muscles are connected in a way that gives the eye three degrees of freedom in its movement. Each pair of muscles is responsible for one axis and two directions. The whole structure of eye and muscles offer the possibility of absorbing the effect of head movement. The control of these muscles is connected to the brain and equilibrium system of the human.
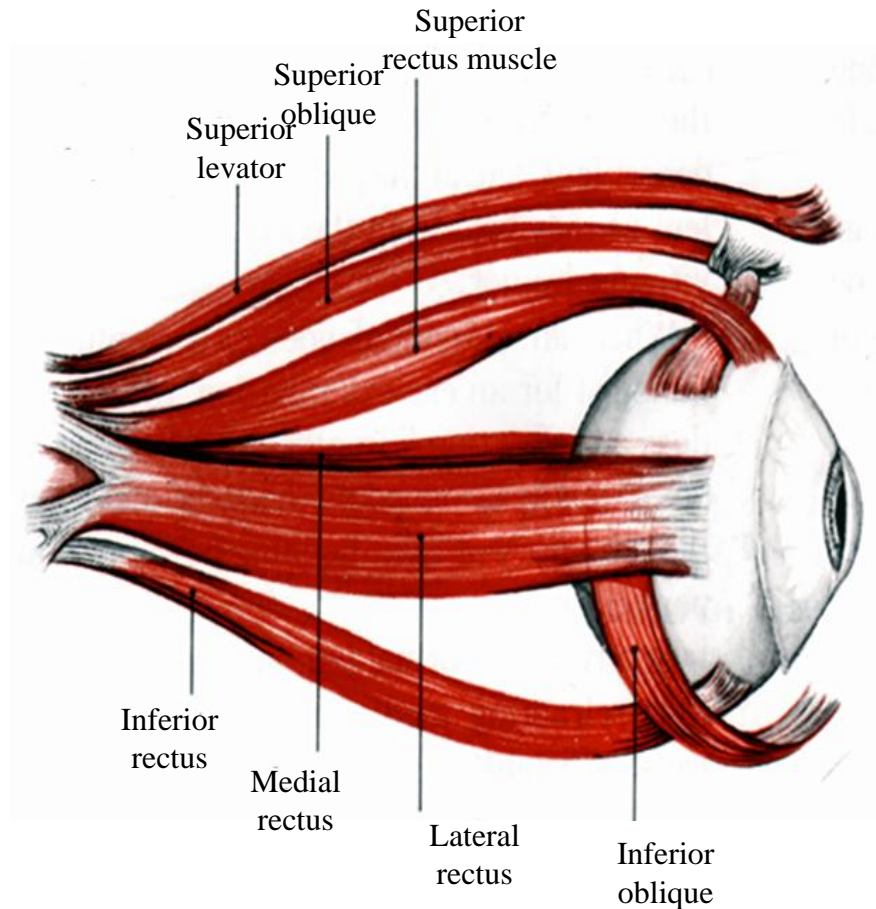
Figure 3.4: Eye and its stabilization muscles

The study of the eye reveals that it is exactly a digital camera. Or for more realistic description, the digital camera has been designed in similar way of the eye. The lens in the camera is fixed and changing its position offers the zoom possibility. However, in the biological eye, the position of the lens is fixed and the zoom is achieved by playing with the shape of lens itself. This is done with the help of lens muscles. A simplified description of the eye is illustrated in **Figure 3.5**. The outer part of the eye is the cornea. It has a spherical shape that offer the possibility to collect the light from wide angle. The iris contains the pupil that controls the amount of light entering to the eye. Inside the iris we can see the flexible lens fixed with the help of muscles. The muscles offer the possibility of changing the shape of the lens to make zoom on subjects. The light reflecting from the objects falls on the cornea and pas through the pupil to the lens. The lens focuses on the subject and concentrates the light on the retina. The retina contains a huge number of light sensors that convert the light into nervous signal.
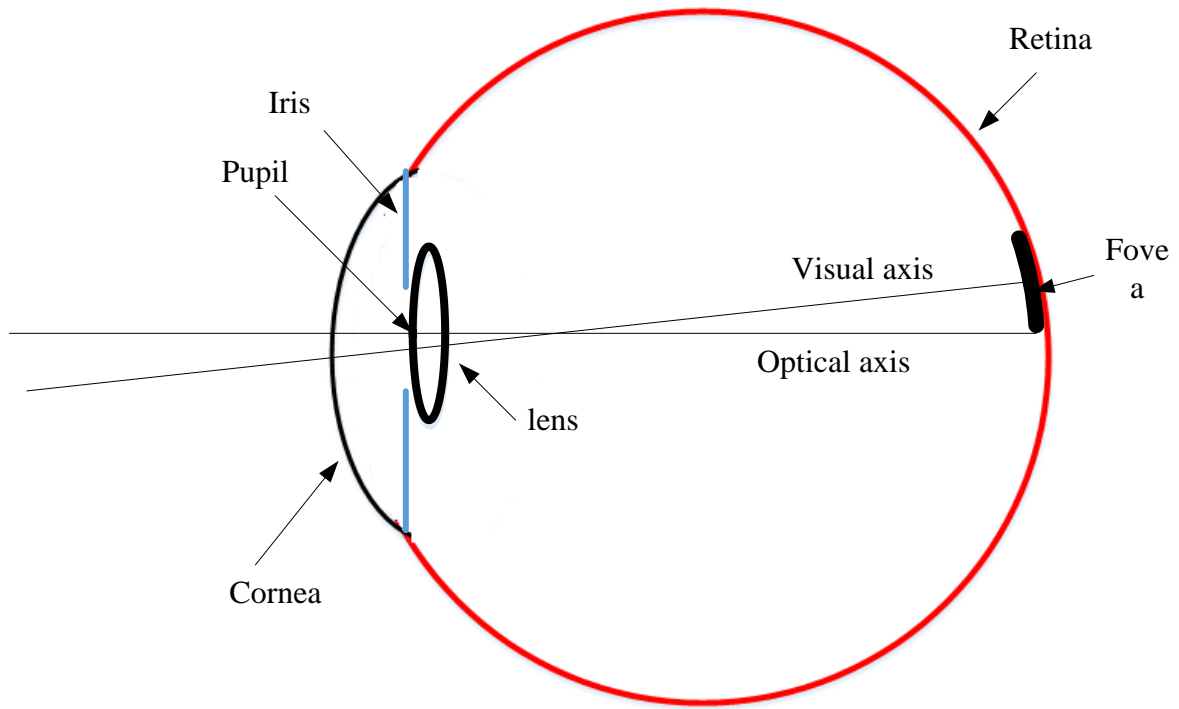
Figure 3.5: Illustration of human eye parts

CHAPTER 4

DATABASE DESCRIPTION AND PROCESSING

4.1 Introduction

This chapter of the work discusses the different processes of data base collection and processing of the collected database until the phase of training the artificial network to perform the gaze estimation task. The gaze estimation process implies the implementation of different techniques to enhance the image visual properties and to find the main features of the image. Also special attention must be paid for the segmentation process of the images that can be done manually or automatically using specific algorithms designed for this purpose. The different image processing techniques that were used in this research are going to be presented and discussed also in this chapter.

4.2 Database Collection Procedure

In this work, gaze estimation process is applied based on back propagation artificial neural networks. The process uses data base collected from the site of Max Planck Institute of Informatik MPIIGaze. The data base constitute of the gaze of many participants. The eye gaze dataset was collected using laptop over wide range of time that extended for more than 7 months as stated in the source of dataset. It is important to mention here that database was collected under normal daily life conditions (Zhang et al., 2015). The dataset was downloaded from their website and used in our work. Their database was collected under realistic conditions. MPIIGaze contains originally 213659 images. This database was collected from 15 different laptop users over a long time period. **Figure 4.1** shows the images of 12 different laptop users during some of the data collection sessions. The main advantage of MPIIGaze data set is that it represents a variety of the illumination levels and appearance. This can be seen clearly from the figure.
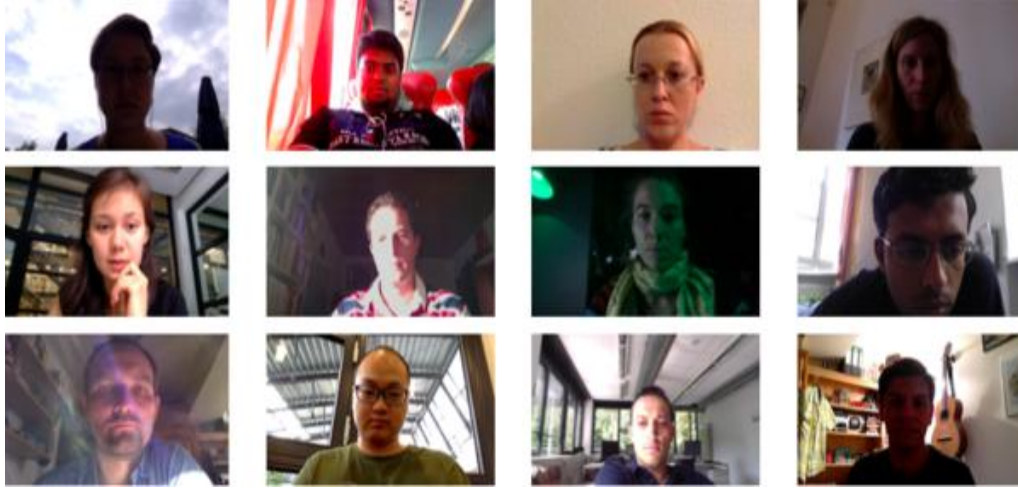
Figure 4.1: Appearance based gaze image capturing (Zhang et al., 2015)

The original database was collected by implementing custom software that runs in the background service of the laptop operation system. The program's mission is to generate a random sequence of 20 points on the laptop screen and ask the user to look at these points. These points appear in the form of gray circle that shrinks gradually and contains a white dot in the centre. This service was executed automatically each 10 minutes on the user's laptop during the experiment. User's were asked to press a key when the point disappears confirming that they were looking at it. No any constraints were given to the users limiting their immovability or position during the sessions. As different laptops and screen sizes were used in the experiment, the dot positions were converted to 3D position in camera coordinate system. Manual annotation of about 10848 images was carried out by human annotators. These images were annotated with 12 face landmarks (Zhang et al., 2015). The data base contains the images for 15 users taken over a long period of time in a daily basis. It also contains the target gaze in screen coordinates as a vector of two elements. The gaze of each image was also given in 3D coordinates referred to the camera coordinates. This one can be translated into two angles coordinates as illustrated by (Zhang et al., 2015). The converted gaze positions are given in 2D angle positions that describe the yaw and pitch angles of the gaze. The database also contains the 3D head position of each one of the users in the camera coordinate system.

In this work, images of 15 persons are going to be used for the examination and training of the ANN for the task of gaze detection. 119 different images were chosen for each person to be

implemented in the proposed training of ANN. These images were extracted from the original data base after being treated through different processes as mentioned in (Zhang et al., 2015). A sample of the used database is presented in **Figure 4.2**. The figure shows the database of the second person gaze images. The target data of the gaze detection system is composed of five parameters for each image per each direction. The five parameters are composed of two parts; the first part represents the position of the dot or circle at which the user gazed in the screen coordinates. The other part is the 3D camera coordinates of the same point.

Figure 4.2: Sample of the gaze database images

The **Figure 4.2** shows the gaze direction and position of the gaze point at the laptop screen coordinate for the first image of each person of the training set as provided in the database. The eye zone detection was carried out through different steps explained in the source of the database among which a manual annotation is presented. After finding the images, all the images will be treated and passed through the artificial neural networks to train and evaluate its capability to detect the correct gaze position and direction. In this work, the position and direction of the gaze are going to be given in the form of yaw and pitch angle given in the camera coordinate system.

In order to be able to use the database images for the training and test of the structure of artificial neural network, it is important to apply some image processing on the images so that

they will be suitable for the implementation of the ANN. The next part is discussing the image processing of images before being fed to the neural networks. The steps are resumed also by the flowchart presented in Figure 4.3.

4.3 Images Reading and Processing

The images were stored in the database files in gray scale format before the start of training of the ANN structure. These images were read using MATLAB software and processed preparing them for the training phase. The gray scale image is composed of a 2D matrix containing the pixel values of the image. These values are represented in 8bit coding with 256 levels of white and black colour intensity. **Table 4.1** presents the pixel values of a portion of one of the gray scale images used in our work. It can be seen that the values of pixels are in the range 0 to 255 which is the 8bit numeric range. The original image size and the resized image are presented in **Figure 4.4** below. The image resizing is used to reduce the size of data that is going to be fed to the ANN. The use of original image will load the computer and make more difficult the training and testing processes. It also will imply the use of very huge memory that may not be available for ordinary computer with no gain in performance or detection efficiency. Resizing the image doesn't cause any data loss from the point of view of computer although it looks loss from the point of view of the human eye.
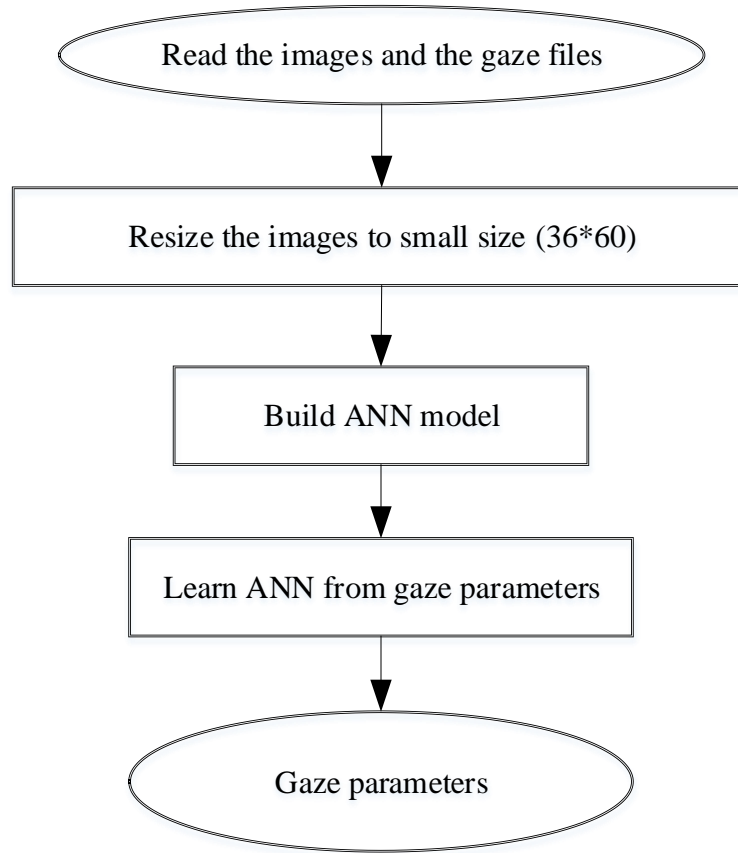
Figure 4.3: Flowchart of the image processing steps

Table 4.1: Gray scale image matrix representation

Pixel values in the gray scale image									
116	140	109	77	65	90	67	64	64	68
110	125	106	70	64	87	75	72	69	82
107	112	113	68	63	79	79	76	67	85
113	110	135	82	63	66	74	72	68	85
119	111	158	106	65	61	67	67	71	96
124	109	167	127	65	66	64	62	64	100
132	106	162	136	62	75	60	55	67	97
140	104	155	139	58	81	55	47	88	99
133	109	136	131	80	93	63	67	91	93
126	107	138	139	95	104	80	89	83	89

Original image size

Shrinking
process

New image size
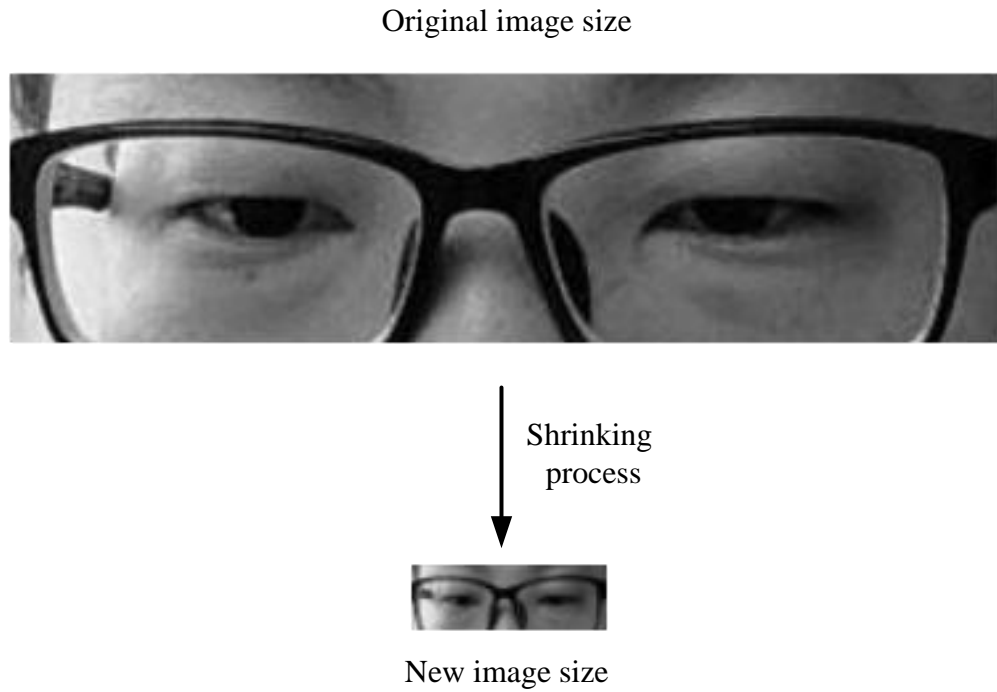
Figure 4.4: Original size of the eye square image vs. resized image

4.3.1 Image normalization

As it was presented earlier in Table 4.1, the image pixels are presented in the form of 8bit unsigned integers. This representation is difficult – but acceptable – to be used with artificial neural networks. In different applications of ANN the use of normalized data that is enveloped in the period [0, 1] is the preferred.

Table 4.2: Normalized pixel values of a portion of the image

Pixel values of the image after normalization									
0,45	0,55	0,43	0,30	0,25	0,35	0,26	0,25	0,25	0,27
0,43	0,49	0,42	0,27	0,25	0,34	0,29	0,28	0,27	0,32
0,42	0,44	0,44	0,27	0,25	0,31	0,31	0,30	0,26	0,33
0,44	0,43	0,53	0,32	0,25	0,26	0,29	0,28	0,27	0,33
0,47	0,44	0,62	0,42	0,25	0,24	0,26	0,26	0,28	0,38
0,49	0,43	0,65	0,50	0,25	0,26	0,25	0,24	0,25	0,39
0,52	0,42	0,64	0,53	0,24	0,29	0,24	0,22	0,26	0,38
0,55	0,41	0,61	0,55	0,23	0,32	0,22	0,18	0,35	0,39
0,52	0,43	0,53	0,51	0,31	0,36	0,25	0,26	0,36	0,36
0,49	0,42	0,54	0,55	0,37	0,41	0,31	0,35	0,33	0,35

Range 0-1 is preferred. For that reason, a normalization process was applied on all the database images such that the input pixels will all be in the range 0-1. **Table 4.2** illustrates the use of normalized pixel data as inputs for the ANN structure. The normalization of images data is generally simple and acquired by dividing all pixel values by the value 255.

4.3.2 Input matrix construction

After resizing all images, converting resized images into one dimensional vector, and normalizing the images; these images or vectors are going to be arranged in a suitable form to be fed to the neural network structure before starting the training and test process. Vectors are arranged one beside the other in a 2D matrix that will contain all training inputs. Target positions are also arranged in an output matrix in the same order of the correspondent inputs. The training process will pick one input vector and one output vector at a time and process them. **Figure 4.5** illustrates the vector arrangement in the input matrix of the ANN. each person's image is converted to a vector. The vectors are then arranged in one matrix as shown in the next figure. The first column represents the image of first person while the second column represents the image of the second person and vice versa.
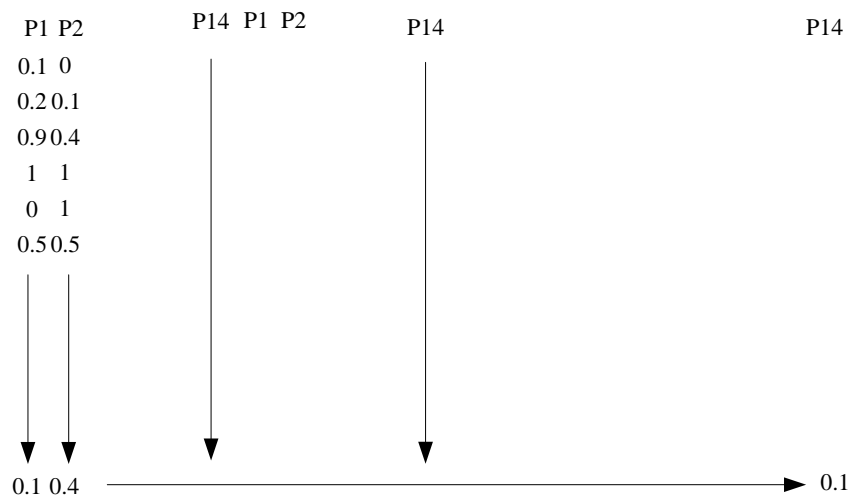
Figure 4.5: Input matrix structure during the ANN training

4.4 Training Procedure of the Artificial Neural Network

The proposed system consists of multilayer artificial neural network trained through the use of back propagation algorithm. The back propagation algorithm uses a gradient based minimization algorithm to ensure the convergence of outputs toward the wanted target. The flowchart of the proposed training algorithm is presented in the flowchart of Figure 4.6.
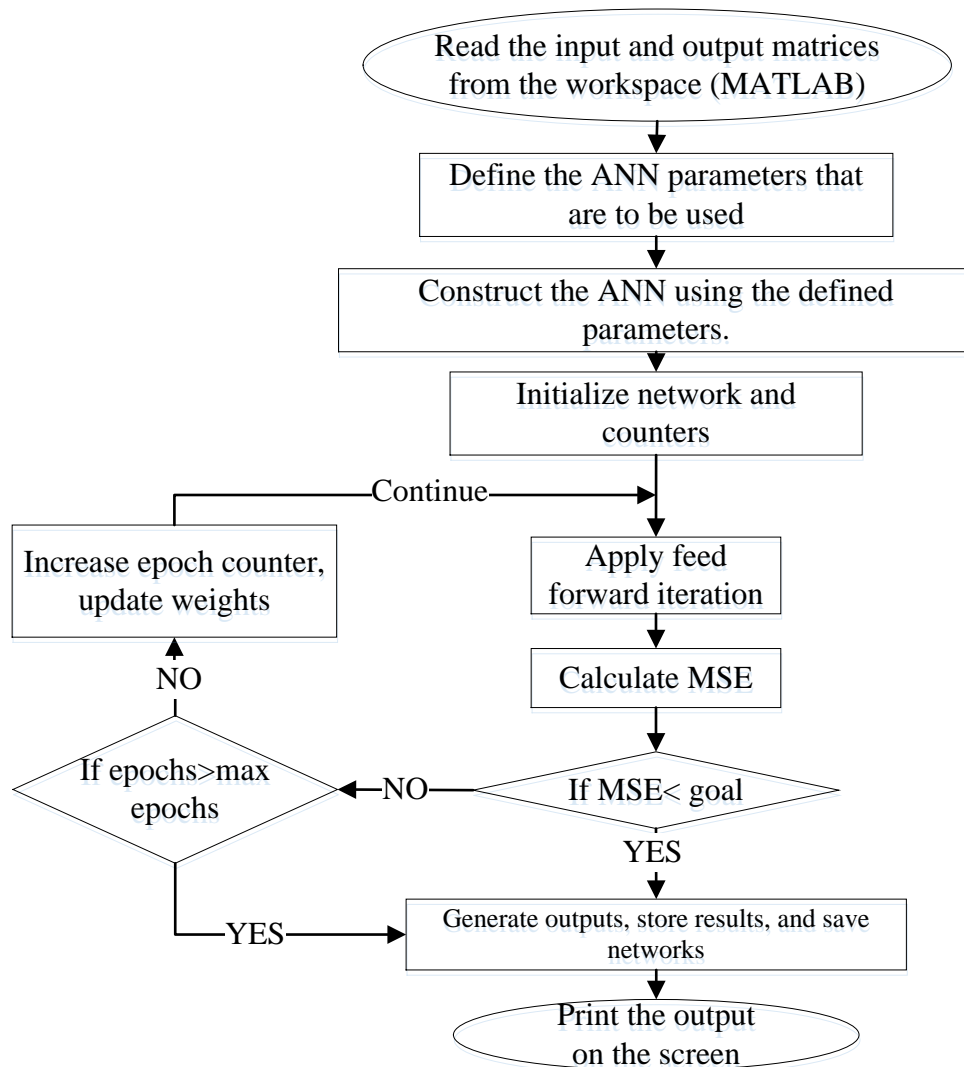
Figure 4.6: Flowchart of the training process of the ANN

The flowchart shown above illustrates the steps of training a neural network using back propagation algorithm. The algorithm uses the inputs in a feed forward iteration to generate outputs of the network. The generated outputs are then compared with the desired outputs. The

comparison is stored in the form of mean squared error function. If the MSE is small enough, the network is generalized and said to be learned neural network. If the MSE is still greater than the MSE goal; the weights are updated based on the MSE values. Another epoch or iteration is initiated until a suitable MSE value is found or until a maximum number of iterations is found.

4.5 Practical Application of ANN in MATLAB

The above mentioned training procedure was applied and executed on MATLAB using a light version of the MATHWORK software. The artificial neural networks tool of MATLAB was used to implement the proposed system. Different training and test parameters are going to be investigated in order to experiment the performance of the system under different configurations. The database was initially divided into two parts to realize the objective of training and testing of the ANN. The first part was provided to the network with its expected targets while the other part was provided without the targets and used to test the capability of the network.

CHAPTER 5

## RESULTS AND DISCUSSIONS

5.1 First ANN Configuration

In the application of the neural network for gaze detection of our database, we have chosen to use the method of leave one person out test. The images of 14 subjects or person are going to be used for the training of our neural network while the eye images of the 15th subject is kept out of the training process. This means that 14*119 images are going to be used for training the network while 119 other images will be implemented for the test of the trained network. In the next stage of the work; another implementation two persons will be kept out during the training of the network. In this stage two different configurations will be used in which the first two persons or the last two persons are going to be omitted in the training of the system. The experiment is done to examine the capability of the network to learn with fewer examples and to generalize the learned information. It is important to mention here that this application is very different from classification and recognition systems. It is very sensitive for the position value and needs for higher accuracy. Although test results are very important in neural networks, they can be with little importance here as a better system can be built around the idea of training with all possible cases. This can be done by setting a first configuration session where all possible positions are captured and trained using the system.

The parameters of the implemented neural network structure are illustrated in **Table 5.1** below. After training the network using these parameters training mean error of 6.2 degrees was found while the test mean error of 8.0 degrees was detected. The training of the back propagation neural network took approximately 5 minutes and a half to reach a MSE of 0.001. The MSE curve during the training is presented in **Figure 5.2**. The figure illustrates that the MSE curve is decreasing gradually and that the network was converging successfully. The networks have given accurate outputs of the gaze direction with performance of 92% for the gaze with error of less than 6 degrees. The test of the trained network has revealed that it was able to detect the gaze direction with a performance of 83% for an error angle of 8 degrees. The overall training and test performance of the system was found to be 90%. **Figure 5.1**

presents the MATLAB interface for the artificial neural networks tool during the training of the ANN. The mean error in degrees is found by calculation of the mean of the error of all images gaze targets in degrees. All errors between actual output and desired outputs or real outputs are calculated; the mean error is the average of all these errors.

Table 5.1: ANN parameters of first configuration

Hidden layers	1	Time	5 min and 32 sec
First hidden size	100	Epochs	1025
Transfer functions	tansig, tansig	Learning rate	0.1
MSE	0.00174	Momentum factor	0.8
Training performance	92%	Input / output	2160 / 2
Test performance	83%	Mean train error	6.2 degree
Overall performance	90%	Mean test error	8.0 degree
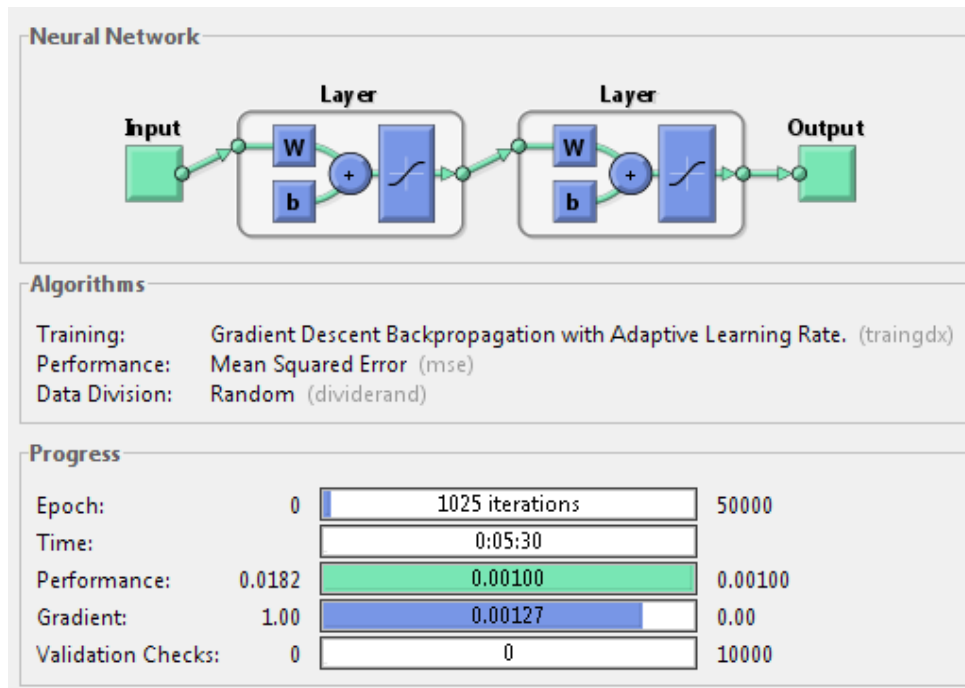
Figure 5.1: Training tool of ANN network
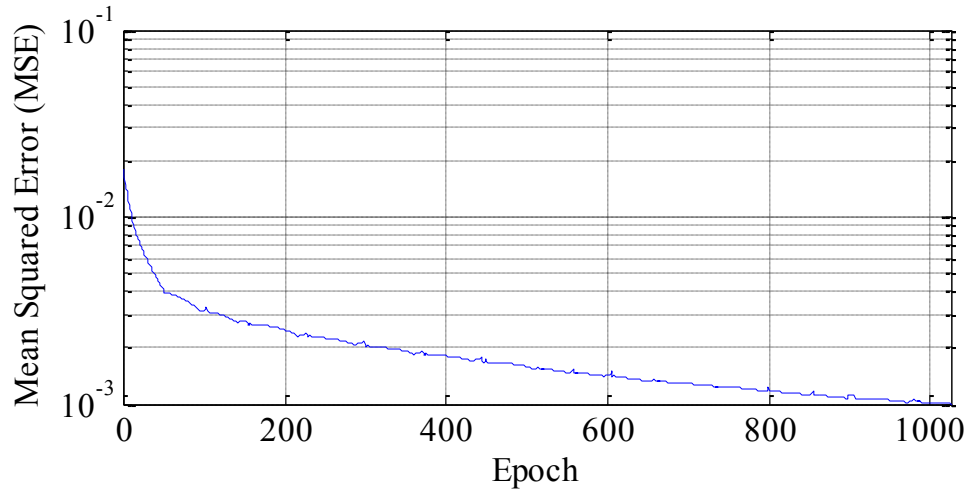
Figure 5.2: MSE curve during the training of first configuration

5.2 Second ANN Configuration

In this configuration the first person images were reserved for test purpose while the images from 2 to 15 were used in the training. This was done by shifting the first person to the last position and applying another training process. Results of training of this configuration are present in **Table 5.2** below. The MSE was reduced to 0.0009 and the mean error degree to 4.8 degree during the training and 6.5 during the test of the network. **Figure 5.3** shows the MSE curve during the training of the network.

Table 5.2: ANN parameters of the second configuration

Hidden layers	1	Time	6 min and 33 sec
hidden size	100	Epochs	2004
Transfer functions	tansig, tansig	Learning rate	0.1
MSE	0.0009	Momentum factor	0.8
Training performance	97%	Input / output	2160 / 2
Test performance	80%	Mean train error	4.8 degree
Overall performance	95.8%	Mean test error	6.5 degree
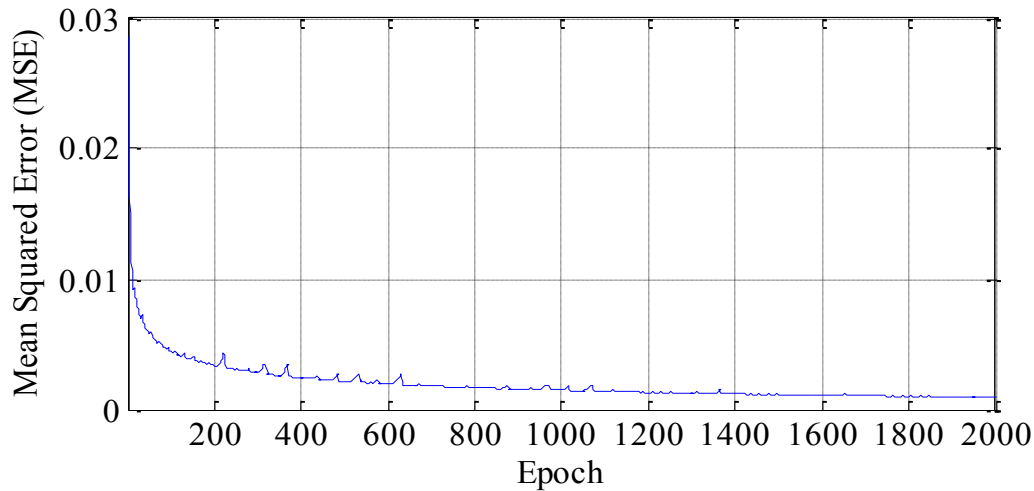
Figure 5.3: MSE curve of the training set

5.3 Third ANN Configuration

In this configuration, the training of the neural network was carried out with the exclusion of 2 persons from the training. After the end of the training, the system has given the next results tabulated in Table 5.3. 200 elements of the hidden layer were used in this configuration. The training took around 18 minutes to finish and reach MSE of 0.0007. The performance of the training of 99.4% was achieved while the test performance was 74.36%. Overall performance of the system was 96%. The mean error of the training was 3.8 degrees while that of test was found to be 9.4 degrees. **Figure 5.4** presents the curve of the training mean squared error obtained in this configuration.

Table 5.3: ANN parameters of the 3rd configuration

Hidden layers	1	Time	18 min and 4 sec
hidden size	200	Epochs	3228
Transfer functions	tansig, tansig	Learning rate	0.25
MSE	0.0007	Momentum factor	0.8
Training performance	99.4%	Input / output	2160 / 2
Test performance	74.36%	Mean train error	3.8 degree
Overall performance	96%	Mean test error	9.4 degree

41
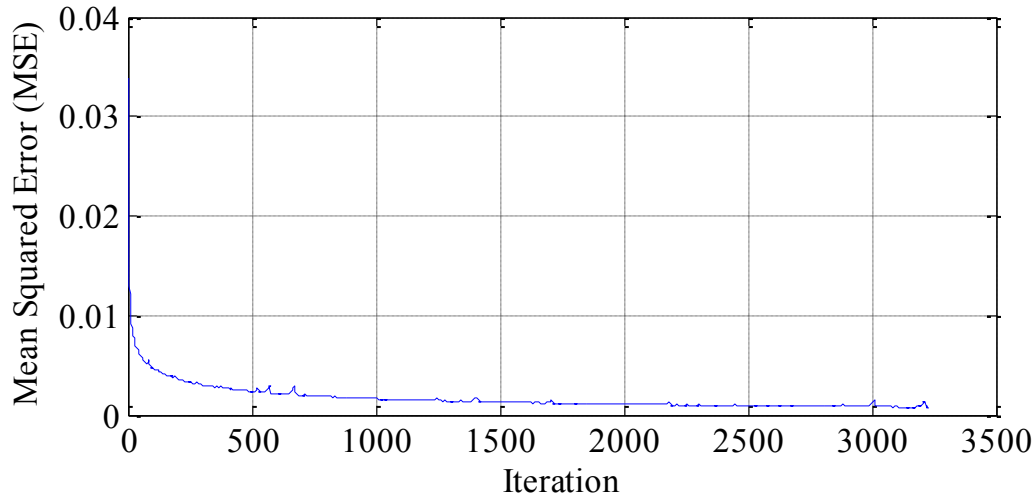
Figure 5.4: MSE curve for the 3rd configuration

5.4 Fourth ANN Configuration

The last two persons of the dataset were eliminated from the training dataset and the training of the neural network was applied using the parameters shown in Table 5.4. The training took 11 minutes and 28 seconds to achieve MSE of 0.00047. The training performance achieved in this experiment was 99.8% which is better than the previous ones. The test performance of the experiment was 70.1% with 10% less than the previous result. The total performance was found to be 95.9% for all data base.

Table 5.4: ANN parameters of the 4th configuration

Hidden layers	1	Time	11 min and 28 sec
hidden size	120	Epochs	3328
Transfer functions	tansig, tansig	Learning rate	0.25
MSE	0.00047	Momentum factor	0.8
Training performance	99.8%	Input / output	2160 / 2
Test performance	70.1%	Mean train error	4.2 degree
Overall performance	95.9%	Mean test error	9 degree

Figure 5.5 shows the evolution of the MSE curve during the training in this configuration. The curve is decreasing slowly with the training development as the figure illustrates. The minimum registered MSE was 0.00047.
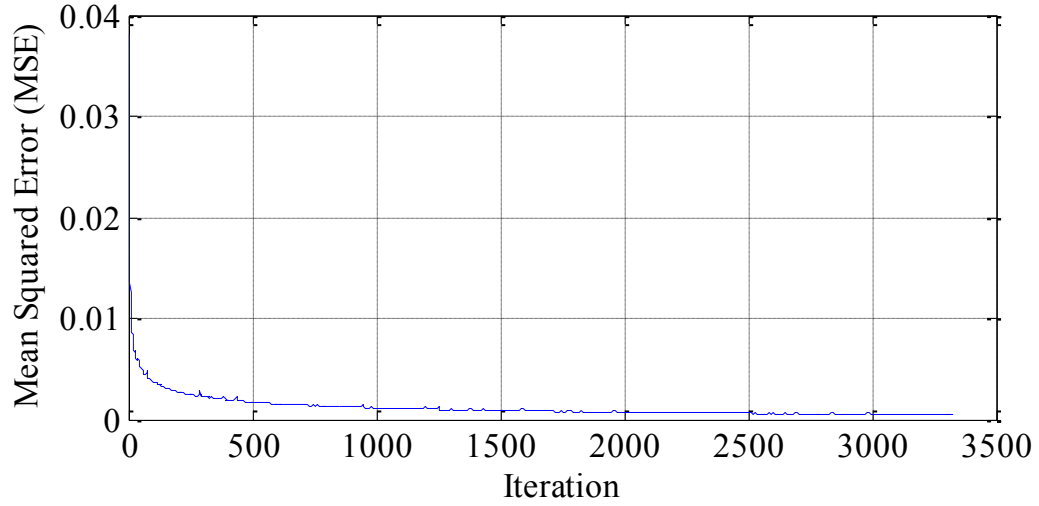
Figure 5.5: MSE curve of the 4th configuration

Next table shows a comparison between the obtained results in the two different cases. It presents the MSE, training performance, test performance, and mean error of the test and training results between the two cases. The table shows that the training error of 3.8 degrees was achieved in the 3rd configuration experiment. Test of the datasets has given different mean error values as shown in the table.

Table 5.5: Comparison of the obtained results

Train/test	Mean error (training)	Mean error (test)	Train perf	Test perf.	MSE
1-14 / 15	6.2 degree	8 degree	90%	82%	0.001
2-15/ 1	4.8 degree	6.5degree	97%	80%	0.0009
3-15/1-2	3.8	9.4	99.4%	74.36%	0.0007
1-13/14-15	4.2	9	99.8%	70.1%	0.00047

5.5 Comparison of Different Configurations of ANN

In order to show the effect of parameters' variation of artificial neural networks on the performance of our system different experiments were carried out under various configurations. The results of these experiments were collected and presented in the **Table 5.6** below. The results show how the variation of learning rate and number of hidden layer

neurons affect the obtained results. A good performance was obtained for number of neurons between 80-140 neurons in the single hidden layer.

Table 5.6: Training and test results of ANN under different configurations

Conf.	LR	MF	Hidden neurons	Time (s)	Train perf.	Test perf.	MSE	Train error	Test error
1	0.01	0.2	30	434	85%	75%	0.003	7.5	8
2	05	0.1	20	1000	80%	72%	0.0026	8	9
3	0.1	0.1	30	1028	81%	75%	0.002	7.8	8.2
4	0.01	0.2	50	1200	89%	78%	0.002	7.4	8
5	0.05	0.2	80	100	90%	82%	0.002	7	7.6
6	0.05	0.2	90	85	91%	82%	0.002	6.8	7.4
7	0.1	0.2	150	1250	99%	85%	0.0005	5	6.8
8	0.05	0.3	200	155	91%	88%	0.0005	6.5	8.5
9	0.05	0.3	250	140	89%	88%	0.0008	6	8
10	0.05	0.3	300	2600	92%	89%	0.0005	5	7
11	0.05	0.3	60	350	91%	85%	0.0009	6.5	7.3
12	0.05	0.3	120	500	90%	82%	0.001	7	10
13	0.05	0.3	500	3500	96%	86%	0.0006	5.3	7.4

In the next table, the parameters of training of the neural network for 15 different experiments that covers all the samples of the database are presented. In each experiment one person out of the training was applied. The results of the training of the neural network are presented in **Table 5.7** below. In each experiment, different person was excluded from the training and used for the test of network results.

Table 5.7: Parameters of the training of the neural network

Hidden layers	1	Time	
hidden size	200	Epochs	
Transfer functions	tansig, tansig	Learning rate	0.25
MSE	0.002	Momentum factor	0.8
		Input / output	2160 / 2

Table 5.8: Results of the training for different test person

Test	Train	Test error (degree)
1st person	2-15	6.4
2nd person	1, 3-15	7.1
3rd person	1-2, 4-15	6.8
4th person	1-3, 5-15	6.3
5th person	1-4, 6-15	6.3
6th person	1-5, 7-15	7
7th person	1-6, 8-15	6.1
8th person	1-7, 9-15	6.5
9th person	1-8, 10-15	5.8
10th person	1-9, 11-15	6.4
11th person	1-10, 12-15	6.2
12th person	1-11, 13-15	6.3
13th person	1-12, 14-15	5.9
14th person	1-13, 15	6.1
15th person	1-14	6
Average		6.5

CHAPTER 6

CONCLUSIONS OF THE WORK

In this work, the study of gaze estimation based on the artificial neural network was carried out. Gaze estimation is gaining more importance due to the increasing needs for more intelligent and faster interface between humans and computers and automated machines. The regular used interface devices are offering good performance under some conditions of function; however, they look unable to offer high performance for some groups of people who has special types of disability. They are also slow in terms of response time and efficiency. Typing using normal methods or surfing the internet using the computer mouse is limited by the speed of the hand. The researchers are trying to find new effective, fast, and easy methods to accomplish the interface between the human and the computer devices. The use of gaze position detection to interact with machines is one of the promising ideas in this domain. It is based on the detection of the gaze of the eye to move a cursor or an object in the computer screen. This method can offer very high speed due to the high speed of the eye movement. Different algorithms are being implemented to perform gaze estimation. This work studies the implementation of back propagation neural networks in the gaze estimation process. A single hidden layer neural network was used in the training and test of the gaze dataset. The gaze dataset obtained from MPIIGAZE database contained a huge number of gaze images collected from 15 participants in the period of seven months. It also contains the gaze direction for each one of these images as registered at the capturing time. The gaze direction was given in the form of two values showing the pitch and yaw angles of the gaze. 119 different images for each participant were chosen arbitrarily to be implemented in our system. This means a total of 1785 different images with their correspondent gaze targets were used in the training and test of the system.

After arranging the images and their correspondent target gaze, two different configurations were used in the training and test of the neural network. The first is the use of one person out of training and the second is the use of two persons out of the training. Each one of these

configurations was also applied twice; once with excluding the first person or two persons, the other was obtained by excluding the last one or two persons. The training of the neural network based on MATLAB platform and the neural network tools of MATLAB has illustrated that the use of artificial neural networks is effective and can offer good results. They can draw the future for this type of applications. The obtained results showed that the use of back propagation artificial neural network had a training performance of 90-99.8% while the test performance was between 70.1-82%. The total performance of the system was between 90-96%. The mean training error of the system was between 3.8 and 6.2 degrees while the mean test error was between 6.5-9.4 degrees. These results and error values show that the neural network is suitable to be applied for gaze detection in computer applications.

Real time application of this type of gaze detection based on artificial neural networks and processing techniques of digital images is recommended for future works. Such a work can explore more and investigate in the real time application and configuration of gaze detection systems.

REFERENCES

A.D.Dongare, R.R.Kharde, & D.Kachare, A. (2012). Introduction to Artificial Neural Network. *International Journal of Engineering and Innovative Technology*, *2*(1), 2277–3754.

Alnajar, F., Gevers, T., Valenti, R., & Ghebreab, S. (2013). Calibration-Free Gaze Estimation Using Human Gaze Patterns. In *2013 IEEE International Conference on Computer Vision* (pp. 137–144). https://doi.org/10.1109/ICCV.2013.24

Anderson, D., & McNeill, G. (2010). *Artificial Neural Networks Technology A DACS State-of-the-Art Report*. NewYork.

Chen, J., & Ji, Q. (2008). 3D gaze estimation with a single camera without IR illumination. In *19th International Conference on Pattern Recognition* (pp. 1–4). https://doi.org/10.1109/ICPR.2008.4761343

Clabaugh, C., Myszewski, D., & Pang, J. (2000). Neural Networks. Retrieved May 1, 2017, from http://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/index.html

Colin, F. (1996). *Artificial Neural Networks* (1.1). University of Paisly.

Drewes, H. (2010). *Eye Gaze Tracking for Human Computer Interaction*. Ludwig-Maximilians-University.

Fu, Y. (2015). *Regression Based Gaze Estimation with Natural head movement*. Concordia university.

Gupta, C. (2006). *Implementatıon of Back Propagatıon Algorıthm (of Neural Networks) In VHDL*.

Hartridge, H., & Thompson, L. C. (1948). Methods of investigating eye movements. *British Journal of Ophthalmology*, *32*, 581–591.

Hugot, V. (2007). *Eye Gaze Analysis in Human-Human Interactions*. Royal Institute of

Technology, Sweden.

Kiran, K. (2009). *Parallelized Backpropagation Neural Network Algorithm using Distributed System*. Thapar University. Retrieved from http://dspace.thapar.edu:8080/jspui/bitstream/10266/760/1/ME THESIS.pdf

Mehmood, M. I. S. S. (2010). *Control of Articulated Robot Arm by Eye Tracking*. Blekinge Institute of Technology.

Moran, T., & Card, S. (1983). No Title. In L. E. Associates (Ed.), *The Psychology of Human-Computer Interaction*. New Jersey.

Pomerleau, D., & Shumeet, B. (1993). *Non Intrusive Gaze Tracking Using Artificial Neural networks*.

Roberts. (2015). Neural Networks. Retrieved February 25, 2017, from http://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/history2

Schneider, T., Schauerte, B., & Stiefelhagen, R. (2014). Manifold Alignment for Person Independent Appearance-based Gaze Estimation. In *22nd International Conference on Pattern Recognition* (pp. 1167–1172).

Seiffert, U. (2002). Artificial Neural Networks on Massively Parallel Computer Hardware. In *European Symposium on Artificial Neural Networks Bruges* (pp. 319–330). Belgium.

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015). Appearance-Based Gaze Estimation in the Wild. In *IEEE International Conference on Computer Vision and Pattern Recognition* (pp. 4511–4520). Germany. Retrieved from https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/gaze-based-human-computer-interaction/appearance-based-gaze-estimation-in-the-wild-mpiigaze/

Zurada, J. (1992). *Introduction to Artificial Neural Systems*. (P. W. Company, Ed.). Minnesota.