

**NO-REFERENCE IMAGE QUALITY ASSESSMENT  
USING NEURAL NETWORKS**

BY

**AMER BIN ZIYAD**

A Thesis Presented to the  
DEANSHIP OF GRADUATE STUDIES

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the  
Requirements for the Degree of

**MASTER OF SCIENCE**

In

**ELECTRICAL ENGINEERING**

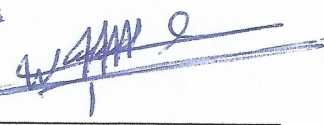
**FEBRUARY 2013**

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS


DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by AMER BIN ZIYAD under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

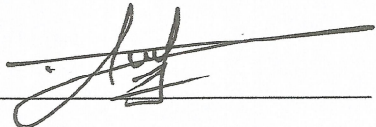
PP  


Dr. Ali A. Al-Shaikhi  
Department Chairman


  
Dr. Salam A. Zummo  
Dean of Graduate Studies



7/7/13  
Date



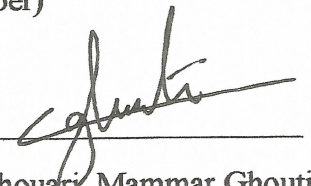
Dr. Mohamed A. Deriche  
(Advisor)



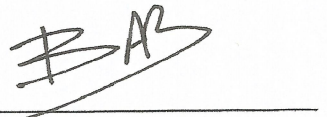
Dr. Mohamed Mohandes  
(Co-Advisor)



Dr. Abdelmalek Zidouri  
(Member)



Dr. Lahouari Mammam Ghouti  
(Member)



Dr. Azeddine Beghdadi  
(Member)

© Amer Bin Ziyad

2013

*This thesis is dedicated to my beloved parents, sisters, & brothers,  
for their endless love, support and encouragement.*

## ACKNOWLEDGMENTS

*In the name of Allah, the Most Gracious and the Most Merciful*

All Praise and thanks be to Allah! We praise and thank Him, ask Him for His Help and Forgiveness, and we seek refuge in Allah from the evils of our souls and the mischiefs of our deeds. He whom Allah guides will not be misled, and he whom Allah misleads will never have a guide. I testify that there is no deity but Allah alone, without any partners, and that Muhammad is His 'Abd (worshiper) and Messenger.

I offer my profound gratitude to my thesis advisor, Dr. Mohamed A. Deriche, for his sustained enthusiasm, creative suggestions, motivation, and exemplary guidance throughout the course of my thesis research. I submit him my honest and humble appreciation, for helping me in achieving the research goals.

I extend my sincere thanks to my thesis committee members, Dr. Azeddine Beghdadi, Dr. Mohamed Mohandes, Dr. Abdelmalek Zidouri, and Dr. Lahouari M. Ghouti, for their useful feedback and insightful comments, to improve the quality of my research work.

My sincere and heartfelt thanks to my friends, Syed Abdul Baqi, Mohammed Fahham, Ashhad Imam, Maaz Farooqui, Abdul Rehman Aravind, Abdul Azeem Siddiqui, Akbar Ali, Javed Khan, Sameer Arastu, Abdul Hai, Mohamed Tamim, Mohamed Khalifa, Abdul Malik, Rizwan, and many others for their help, support, and prayers.

Last but not the least, I express my deepest gratitude to my parents, for their unconditional love, patience, and supplications for me. My brothers Abdul Khader, Mohsin, Amar, Fahad, and my sisters too, share the same acknowledgements for their

support and motivation. I ask Allaah (‘Azza wa Jalla) to reward them all with the best in this world and the hereafter, aameen.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS .....</b>	<b>V</b>
<b>LIST OF TABLES.....</b>	<b>X</b>
<b>LIST OF FIGURES.....</b>	<b>XI</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>XIII</b>
<b>ABSTRACT .....</b>	<b>XV</b>
<b>CHAPTER 1. INTRODUCTION.....</b>	<b>1</b>
<b>1.1 Motivation.....</b>	<b>1</b>
<b>1.2 Image Quality Measures.....</b>	<b>3</b>
<b>1.2.1 Subjective Image Quality Measures .....</b>	<b>4</b>
<b>1.2.2 Objective Image Quality Measures.....</b>	<b>6</b>
<b>1.3 Thesis Objectives .....</b>	<b>8</b>
<b>1.4 Contributions.....</b>	<b>9</b>
<b>1.5 Thesis Outline .....</b>	<b>10</b>
<b>CHAPTER 2. HUMAN VISUAL SYSTEM .....</b>	<b>11</b>
<b>2.1 Human Visual System.....</b>	<b>11</b>
<b>2.2 Physical Structure of Human Visual System.....</b>	<b>11</b>
<b>2.2.1 The Human Eye .....</b>	<b>11</b>
<b>2.2.2 The Retina.....</b>	<b>13</b>
<b>2.2.3 The Visual Cortex .....</b>	<b>16</b>
<b>2.2.4 The Physiological Pathways Hypothesis .....</b>	<b>18</b>
<b>2.3 Visual Perception.....</b>	<b>19</b>

2.3.1	Adaptation of Visual Perception.....	20
2.3.2	Ambiguous Perception of Objects.....	20
2.3.3	Visual Completion of the Environment .....	22
2.3.4	Perception of Impossible Objects.....	23
2.3.5	Classification of Objects.....	24
2.3.6	Attention and Consciousness .....	24
2.4	Summary.....	25
<b>CHAPTER 3. LITERATURE REVIEW.....</b>		<b>26</b>
3.1	Full-Reference Image Quality Measures.....	26
3.2	Reduced-Reference Image Quality Measures .....	36
3.3	No-Reference Image Quality Measures.....	39
3.3.1	Measures used for distortion due to JPEG compression .....	40
3.3.2	Measures used for distortion due to JPEG2000 compression .....	45
3.3.3	Measures used for distortions due to blur and noise .....	50
3.3.4	Non-Distortion-Specific Quality Measures .....	54
3.4	Performance of Objective Quality Measures .....	57
3.4.1	Subjective Quality Evaluation .....	57
3.4.2	Performance Evaluation Criteria .....	64
3.5	Summary.....	66
<b>CHAPTER 4. THE PROPOSED NO-REFERENCE IMAGE QUALITY ASSESSMENT ALGORITHM USING ARTIFICIAL NEURAL NETWORKS.....</b>		<b>68</b>
4.1	Introduction.....	68
4.2	Artificial Neural Networks .....	69
4.2.1	Artificial Neuron .....	70
4.2.2	Activation Function.....	72



4.2.3	Back-Propagation Algorithm/ Delta Rule.....	77
4.3	The Proposed Methodology .....	83
4.3.1	Feature Extraction .....	83
4.3.2	Neural Network Training.....	91
4.4	Summary.....	95
<b>CHAPTER 5. PERFORMANCE EVALUATION AND COMPARISON .....</b>		<b>97</b>
5.1	Introduction .....	97
5.2	Performance evaluation.....	98
5.2.1	Optimum feature set .....	98
5.2.2	Optimum number of hidden neurons .....	104
5.2.3	Performance comparison with traditional algorithms .....	112
5.3	Summary.....	115
<b>CHAPTER 6. CONCLUSTION AND FUTURE WORKS .....</b>		<b>116</b>
6.1	Conclusion .....	116
6.2	Future Works.....	117
<b>REFERENCES .....</b>		<b>118</b>
<b>VITAE .....</b>		<b>127</b>

## LIST OF TABLES

Table 1.1: Mean Opinion Score .....	5
Table 5.1: LIVE database .....	97
Table 5.2: Symbolic representation of the features .....	99
Table 5.3: Optimum features selection for JPEG compression .....	99
Table 5.4: Optimum features selection for JPEG2000 compression.....	100
Table 5.5: Optimum features selection for blur distortion.....	101
Table 5.6: Optimum features selection for noise distortion.....	102
Table 5.7: Optimum features selection across all distortions .....	103
Table 5.8: Selection of optimum number of hidden neurons.....	104
Table 5.9: Performance evaluation and comparison of the proposed model.....	113

# LIST OF FIGURES

Figure 1.1: Degradation of visual quality .....	2
Figure 1.2: Hierarchy of IQMs.....	4
Figure 1.3: Quality Comparison of the images .....	6
Figure 1.4: Objective IQMs .....	8
Figure 2.1: A cross section of the human eye [4].....	14
Figure 2.2: The human visual system [3] .....	15
Figure 2.3: The human retina [3].....	16
Figure 2.4: Receptive field structure of ganglion cells.....	19
Figure 2.5: Schematic diagram of the visual pathways hypothesis [3] .....	21
Figure 2.6: Ambiguous figures.....	22
Figure 2.7: An impossible object. ....	23
Figure 3.1: White noise contaminated images .....	59
Figure 3.2: Blurred images .....	60
Figure 3.3: JPEG compressed images .....	61
Figure 3.4: JPEG2000 compressed images.....	62
Figure 3.5: Reference images used in LIVE database .....	63
Figure 4.1: Block diagram of the proposed algorithm.....	69
Figure 4.2: Artificial Neuron Model .....	71
Figure 4.3: Linear activation function .....	73
Figure 4.4: Sigmoid Transfer Function .....	74
Figure 4.5: Hyperbolic Tangent Sigmoid Function.....	75
Figure 4.6: Derivative of binary sigmoid and hyperbolic tangent function.....	76
Figure 4.7: Elliot Symmetric Sigmoid Function .....	77
Figure 4.8: Back-Propagation of network's error .....	78
Figure 4.9: Blocking artifact .....	85
Figure 4.10: Blurring artifact due to poor spatial activity .....	87
Figure 4.11: Image contaminated with Gaussian noise .....	91
Figure 4.12: Feedforward Network .....	92
Figure 5.1: MSE vs. varied number of hidden neurons for JPEG compression .....	105
Figure 5.2: MSE vs. varied number of hidden neurons for JPEG2000 compression.....	106
Figure 5.3: MSE vs. varied number of hidden neurons for blur distortion.....	107
Figure 5.4: MSE vs. varied number of hidden neurons for noise distortion.....	108
Figure 5.5: MSE vs. varied number of hidden neurons across all distortions .....	109
Figure 5.6: Scatter plot of Subjective scores vs. Objective scores for JPEG compression .....	110
Figure 5.7: Scatter plot of Subjective scores vs. Objective scores for JPEG2000 compression .....	110
Figure 5.8: Scatter plot of Subjective scores vs. Objective scores for blur distortion ....	111

Figure 5.9: Scatter plot of Subjective scores vs. Objective scores for noise distortion ..111  
Figure 5.10: Scatter plot of Subjective scores vs. Objective scores across all distortions  
.....112

## LIST OF ABBREVIATIONS

ANN	:	Artificial Neural Network
BIQI	:	Blind Image Quality Index
BP	:	Back Propagation
CSF	:	Contrast Sensitivity Function
DCT	:	Discrete Cosine Transform
DTF	:	Distortion Transfer Function
DMOS	:	Differential Mean Opinion Score
DWT	:	Discrete Wavelet Transform
FR	:	Full Reference
FWQI	:	Foveated Wavelet image Quality Index
FWD	:	Foveated Wavelet image Distortion
HVS	:	Human Visual System
IT	:	Infero Temporal
IQM	:	Image Quality Measure
IQA	:	Image Quality Assessment
LIVE	:	Laboratory for Image and Video Engineering
LGN	:	Lateral Geniculate Nucleus
MAE	:	Mean Absolute Error
MLP	:	Multilayer Perceptron
MOS	:	Mean Opinion Score
MSE	:	Mean Square Error
MT	:	Medial Temporal

NN	:	Neural Network
NR	:	No Reference
OR	:	Outlier Ratio
PCA	:	Principal Component Analysis
PCC	:	Pearson's Correlation Coefficient
PSNE	:	Peak Signal to Noise Ratio
RAS	:	Ringing Annoyance Scores
RR	:	Reduce Reference
SVM	:	Support Vector Machine
SVR	:	Support Vector Regression
SROCC	:	Spearman's Rank-Order Correlation Coefficient
VSNR	:	Visual Signal to Noise Ratio
VQEG	:	Video Quality Experts Group

## ABSTRACT

Full Name : Amer Bin Ziyad  
Thesis Title : No-Reference Image Quality Assessment using Neural Networks  
Major Field : Electrical Engineering  
Date of Degree : February 2013

*In recent years, we have witness a remarkable progress in the field of digital information processing. The digital images/videos are rapidly proliferating. Despite the advancements in digital communications, the images/videos persist to suffer from major impairments that degrade the visual quality of the signal. Thus, some objective quality measures are needed to monitor these impairments, and optimize the control parameters for quality improvement. However, the formulation of the objective image quality measures is very challenging, especially in the absence of the original medium. In this thesis, we propose a learning-based No-Reference Image Quality Assessment (NR-IQA) system using Artificial Neural Networks (ANN). The aim is to develop a computational model for the quality assessment of the images degraded by following distortions: blur, noise, JPEG compression, JPEG2000 compression, and across all distortions. The major artifacts observed in these distortions are quantified by a set of characterizing features extracted from the distorted medium. These active features are then used as an input to the neural network for quality prediction. The adaptive neural network learns the highly non-linear relationship between the statistical features and the overall quality rating, and approximates the quality score close to human perception. The experiments are performed on the images taken from the standard LIVE database. The performance of the proposed*

*algorithm is evaluated using the criteria recommended by the Video Quality Expert Group (VQEG). The experimental results show that the proposed quality measure outperforms the traditional logistic regression models, with an excellent correlation between the predicted and the subjective quality scores. The proposed machine learning approach is a powerful technique, and can be implemented for images/videos suffering from any kind of distortion.*



## ملخص الرسالة

الاسم الكامل: **عامر بن زياد**  
عنوان الرسالة: **التقييم بلا مرجع لجودة الصور باستخدام شبكة الأعصاب**  
التخصص: **هندسة اتصالات**  
تاريخ الدرجة العلمية: **فبراير ٢٠١٣**

في السنوات الأخيرة، شهدنا تقدماً ملحوظاً في مجال معالجة المعلومات الرقمية. الصور والفيديو الرقمية تنتشر بسرعة. وعلى الرغم من التقدم في مجال الاتصالات الرقمية، فإن الصور / الفيديو مازالت تعاني من عاهات رئيسية والتي تؤدي إلى تدهور جودة الصورة. وبالتالي، هناك حاجة إلى بعض التدابير النوعية الموضوعية لرصد هذه العاهات، وتحسين المعاملات المسيطرة لتحسين الجودة. ومع ذلك، فإن الهدف من صياغة تدابير جودة الصورة صعب للغاية، خاصة في غياب الصورة الأصلية. في هذه الأطروحة، نقترح صورة ذات نوعية التقييم نظام (لا مرجع تعلم) قائم على استخدام الشبكات العصبية الاصطناعية. والهدف هو تطوير حسابي نموذج لتقييم جودة الصور التي تدهورت بفعل التشوهات التالية: الطمس، الضوضاء، ضغط JPEG، ضغط JPEG2000، وعبر كل التشوهات. إن الملاحظة الرئيسية الاصطناعية التي في هذه التشوهات من سببها مجموعة من الميزات التي تميز المستخرجة من الوسيط الناقل مشوهة. يتم استخدام هذه الميزات بالموقع كمدخل إلى الشبكة العصبية للتنبؤ بالجودة. تتعلم الشبكة العصبية التأقلمية العلاقة الغير الخطية بين الميزات الإحصائية وتصنيف الجودة الشاملة، والتي تقارب درجة جودة قريبة من الإدراك البشري. تجرى التجارب على الصور المأخوذة من قاعدة بيانات حية قياسية. يتم تقييم أداء الخوارزمية المقترحة باستخدام المعايير التي أوصى بها فريق الخبراء لجودة الفيديو. تظهر النتائج التجريبية أن التدبير المقترح للجودة يتفوق على نماذج الانحدار اللوجستي التقليدية، مع وجود علاقة ممتازة بين المتوقعة ونقاط الجودة الشخصية. إن نهج التعلم الآلي المقترح هو تقنية قوية، ويمكن تطبيقها على الصور / الفيديو التي تعاني من أي نوع من التشويه.

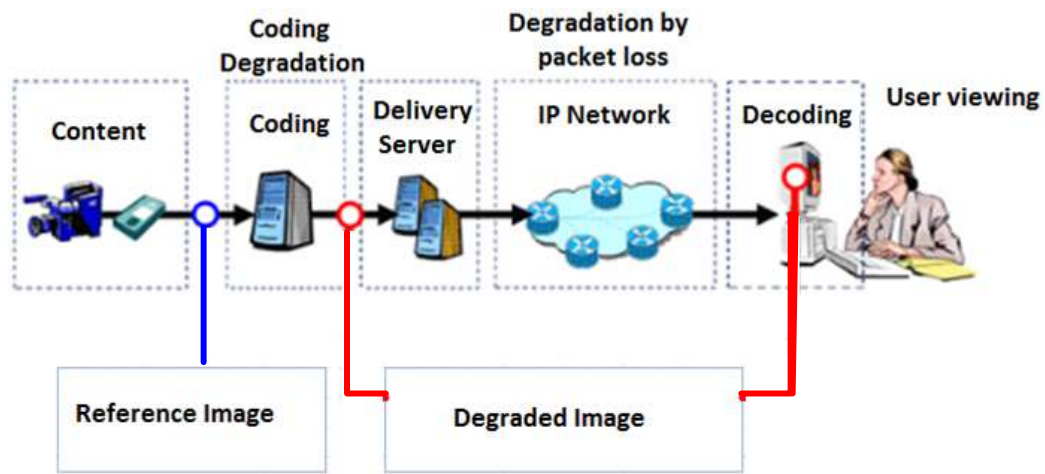
# CHAPTER 1

## INTRODUCTION

### 1.1 Motivation

Human beings rely highly on visual information to perceive the world. The advancements in multimedia technologies had made it possible to capture, compress, send, and display various kinds of visual information. A great deal of effort has been made by experts in the 2D image/video transmission industry, to guarantee a satisfactory quality of the signal to the end user [1]. In current bandwidth famine era, coupled with increasing multimedia traffic, perceptual optimization of the multimedia services is promising to provide similar/improved quality of service to the consumer. The produced visual content is transmitted to the consumer through the communication channels, by the network provider. Various impairments due to compression, channel noise, packet loss etc. are introduced in the signal from the chain of operations from the signal acquisition till its transmission. These impairments mar the viewing experience, and thus, degrade the visual quality of the signal. It is very important to recognize and quantify the quality degradation of these images and improve visual content. The compression techniques used for the signal bandwidth reduction is one of the important sources of image degradation. For instance, images which are compressed by lossy compression techniques experience different artifacts upon reconstruction. Thus, it is essential to evaluate the visibility of compression artifacts, so as to optimize the parameter settings of the related

system. Moreover, images are subjected to various losses, errors, or decays during transmission through communication networks. All these transmission impairments can lead to poor quality of received images. It is important for the network server to identify the image degradation and control the streaming resources in transmission. The reduced visual quality of the signal due to coding and transmission is represented in Figure 1.1.



**Figure 1.1: Degradation of visual quality**

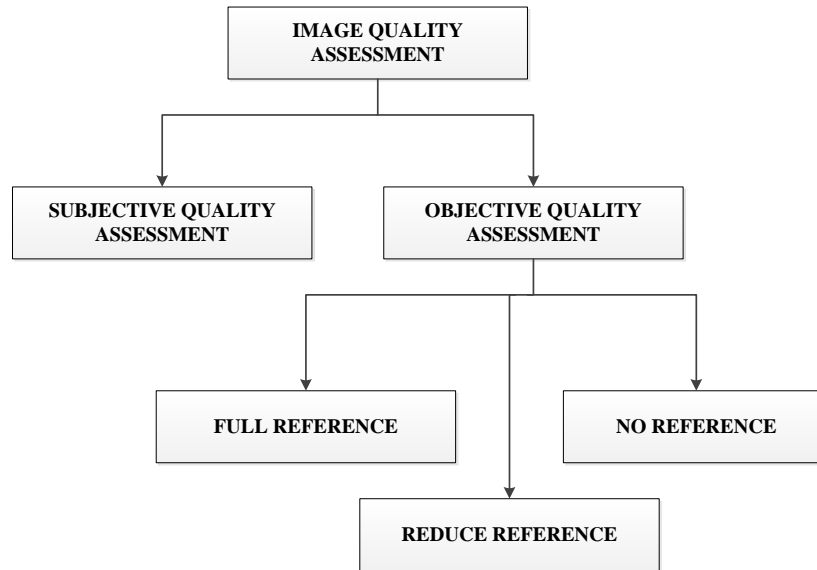
Humans are the ultimate receivers of most of the visual information; hence, the *Human Visual System* (HVS) is the system we rely on for image quality assessment. The techniques involving the use of human observers for quality assessment are known as subjective quality measures. However, such subjective quality assessment is not suitable for real time-applications due to the huge implementation cost and time consumption. It is desirable to develop computational models, which could automatically estimate the quality of perceived image. This is the basic motivation behind developing objective quality measures whose, ultimate aim is to estimate the quality of the received image the way HVS does.

In objective quality assessment, it is required to have information about three important aspects: information of the original image, information about the distortion process, and information about the HVS. There are many real-time applications wherein the information about the original image is not always available. Under such scenarios, the quality of the distorted images is to be estimated only by analyzing the distorted medium itself. The ability of HVS to perceive image quality without any reference motivates the kind of image quality assessment, without referring to distortion-free image. Thus, both the practical requirement and the working mechanism of the HVS inspired to develop quality assessment models using only the distorted medium; these are referred to as No-Reference Image Quality Assessment (NR-IQA) techniques.

## **1.2 Image Quality Measures**

In any digital imaging system, it is very important to analyze the quality of the image it receives before any processing. This is needed to control, retain or improve the quality before processing, storage or transmission. Thus, Image Quality Measures (IQM) is important as a first processing stage. The IQMs are highly reliable and commonly used in the development of image and video processing systems. In general, Image Quality Assessment (IQA) measures are classified into two types: subjective quality measures and objective quality measures. The subjective measures involve the use of human observers for image quality evaluation. However, this method has many inherent drawbacks, due to which, the objective quality measures are preferred over the subjective quality measures. The ultimate aim of objective image quality assessment techniques is to

analyze the perceived image and approximate its quality close to subjective scores [2]. The hierarchy of different image quality metrics is shown in Figure 1.2.



**Figure 1.2: Hierarchy of IQMs**

### **1.2.1 Subjective Image Quality Measures**

The concept of subjective quality measurement was originally introduced in the area of signal compression mainly for image and speech signals. Traditionally, the value which specifies the signal quality after processing is the Mean Opinion Score (MOS). It is a numerical indication specifying the quality of perceived multimedia signal after compression/decompression. The MOS is a single numerical value on a scale of a predefined range given by the observer upon observing the image. For instance, a five level scale representing 1 as the lowest quality and 5 as the excellent quality of an image. The in-between values represent 2-poor, 3-acceptable, and 4-good. The grading scale of MOS is given in Table 1.1.

**Table1.1: Mean Opinion Score**

Score	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible/not annoying
3	Fair	Slight annoying
2	Poor	Annoying
1	Bad	Very annoying

The MOS value is an average of quality scores given by a group of observers for a given image and is expressed as:

$$MOS = \sum_{i=1}^N X_i P(X_i) \quad (1.1)$$

In the above equation,  $X_i$  is the quality score of  $i^{th}$  image,  $P(X_i)$  is the image score probability and  $N$  is the total number of observers. An equi-probable score is considered by all observers. The quality comparison of the images is shown in Figure 1.3

The main advantages of subjective quality measures include:

- Reliable results are obtained for both conventional and compressed television systems.
- Scalar Mean Opinion Score (MOS) is produced, which works effectively for a wide range of still and motion picture applications.

The disadvantages of subjective testing on the other hand include:

- Large range of possible factors and test parameters need to be considered.

- Scrupulous setup and mechanism is needed.
- Good number of observers must be employed and screened.
- Complex and time consuming process.

However, the subjective tests are used only for the purpose of development, not for operational monitoring, production line testing, or trouble shooting.



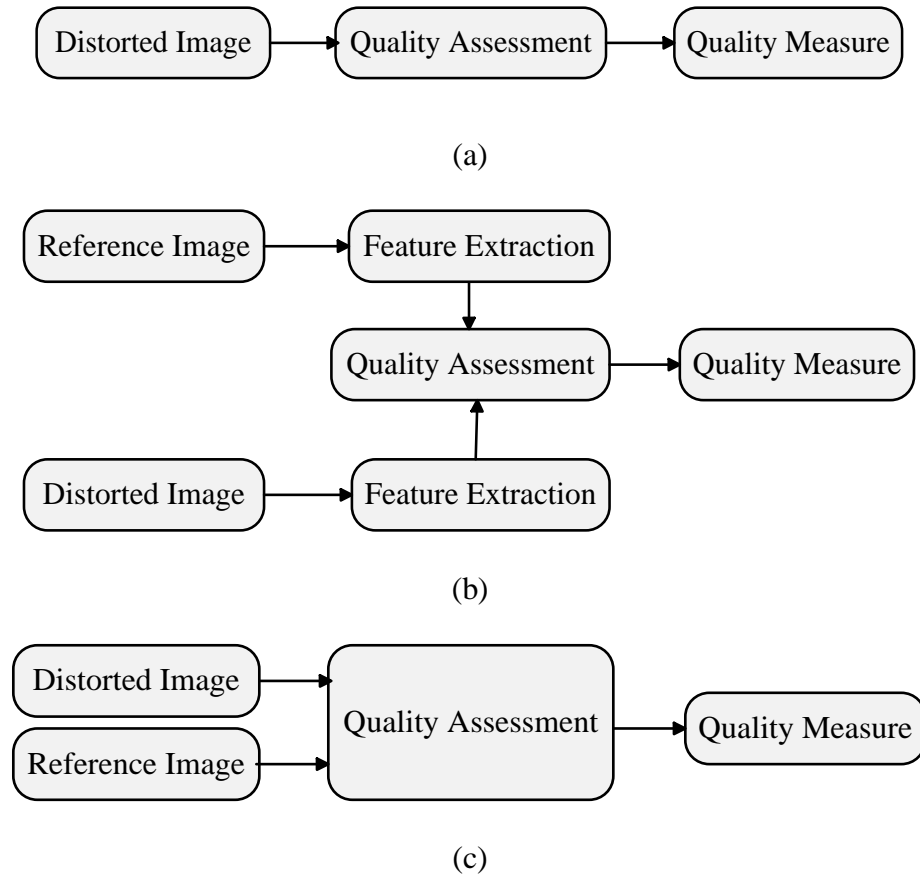
Figure 1.3: Quality Comparison of the images

## 1.2.2 Objective Image Quality Measures

Objective IQMs have been developed to quantitatively estimate perceived image and video quality. They are classified based on the availability of the reference data at the quality assessment system. The different classifications are shown in Figure 1.4. The HVS doesn't experience any difficulty in quality assessment of distorted image/video signals, even in the absence of the reference medium. However, the task which HVS can

do with such an ease seems to be very complex for a machine to perform. The metrics which follow the quality prediction of the perceived image without the prior knowledge of reference data are referred to as No-Reference (NR) or “blind” methods. These methods are heavily used in communication systems as quality estimators. On other hand, metrics involving the use of some degree of knowledge about the reference medium are known as Reduce-Reference (RR) methods. Such information is provided to the receiver along with the transmitted signal for quality prediction of the received image. This makes the task of RR methods easier compared to the NR methods. A set of distinct features are extracted from the original image and transmitted to the receiver through an auxiliary channel. At the receiver end, the features are retrieved and utilized as a base for quality estimation. In applications where the complete reference medium is available at the evaluation system, the Full-Reference (FR) methods are used. These methods utilize the entire original image as a reference for quality prediction. The availability of the reference image facilitates the task of quality prediction and provides a higher degree of reliability. Most of the proposed metrics are based on FR methods [2]. The major drawback of the FR/RR methods is the unavailability of the reference signal at the receiver at all times. Therefore, NR methods are preferred over FR/RR methods in communication systems. The field of NR-IQA remains largely unexplored and is still far from being a mature research area. The intrinsic complexity and limited knowledge of the human visual perception pose major difficulty in the development of robust NR-IQMs. Despite such substantial challenges, the field of NR-IQM is rapidly evolving and is presently an active and rapid emergent area.





**Figure 1.4: Objective IQMs**

**(a) No-reference methods, (b) Reduced reference methods, and (c) Full-reference methods**

### 1.3 Thesis Objectives

The task of no-reference image quality assessment is very challenging due to its intrinsic complexity and the unavailability of the reference signal. NR-IQMs depend solely on the robustness of the features used from the distorted medium. These features must correlate well with the perceived image quality for the development of the robust model.

The main objectives of the thesis are as follows:

- Extraction of robust features from the distorted images for the no-reference image quality assessment.
- Development of an algorithm based on these features and an ANN regressor for quality prediction.
- Intensive testing of the algorithm on the standard LIVE database.
- Comparison with different existing techniques.

## **1.4 Contributions**

Major contributions achieved in thesis are as follows:

- An artificial neural network based no-reference quality assessment algorithm, applicable for JPEG compression, JPEG2000 compression, blur and noise is developed.
- Neural networks with different activation functions were studied, which approximate the image quality by quality-score prediction.
- Several networks are trained and tested for each distortion individually, and also across all the distortions.
- The formulation of the problem of IQA as a classification problem.

## **1.5 Thesis Outline**

The rest of the thesis is organized as follows:

Chapter 2 gives an overview of the anatomical structure and the perceptual behavior of the HVS.

Chapter 3 reviews some representative work reported in the fields of the FR, RR, and NR image quality assessment. It also discusses the performance evaluation criteria used to evaluate the performance of the objective quality measures.

Chapter 4 discusses the proposed NR image quality assessment algorithm using ANN. Under this, we discuss the extraction stage of the features from different distortions; also, the neural network's training stage for quality prediction.

Chapter 5 discusses the performance evaluation of the proposed method on the standard LIVE database, and its comparison with existing objective models.

Chapter 6 gives a summary of the thesis, and some potential future research directions.

## **CHAPTER 2**

### **HUMAN VISUAL SYSTEM**

#### **2.1 Human Visual System**

The concept of Human Visual System (HVS) continues to be an inspiration for the development of the majority of the perception based approaches in computer graphics. In this chapter, some of the important properties of the HVS are discussed. The physical structure of the HVS is first discussed, which is well established. It forms the foundation to understand the complex characteristics of the perceptual behavior [3].

#### **2.2 Physical Structure of Human Visual System**

In this section, the basic visual anatomy and physiology is discussed. This would help in understanding the kind of information that can be coded by human visual mechanisms.

##### **2.2.1 The Human Eye**

The shape of the human eye is almost spherical, except for a bulge at the front [3]. Eyes are placed in almost hemispherical holes in the skull, known as the eye sockets, which are positioned at the horizontal midline of the head. The eye movement is possible by the coordinate use of six small, but strong muscles, called the extraocular muscles, which are

controlled by specific areas in the brain. The eye movement is used for scanning various sections of the visual field, or to focus on the objects located at various distances, without even turning the entire head. The basic function of the eye is to gather light, reflected from objects in the world, and focus it in a clear image on the back of the eye. Various optical functions [3], accomplished by different parts of eye are shown in Figure 2.1. For a human to see an object, the reflected light first enters the cornea, a transparent bulging structure in the front part of an eye, in arrears to which lies a cavity filled with a clear liquid known as the aqueous humor. The light then propagates through the pupil, a variably sized opening in the opaque iris, which gives an eye its external color. Just in arrears to iris lies the lens. The light then passes through the lens, whose shape is controlled by ciliary muscles. By changing the shape of the lens, its optical properties can be altered. This process is known as accommodation. The central chamber of the eye is filled by the clear vitreous humor. The photon then travels through the vitreous humor, and finally, it reaches the retina, a curved surface at the back of an eye. The retina is compactly covered with around 100 million light-sensitive photoreceptors, which convert light into neural activity [3].

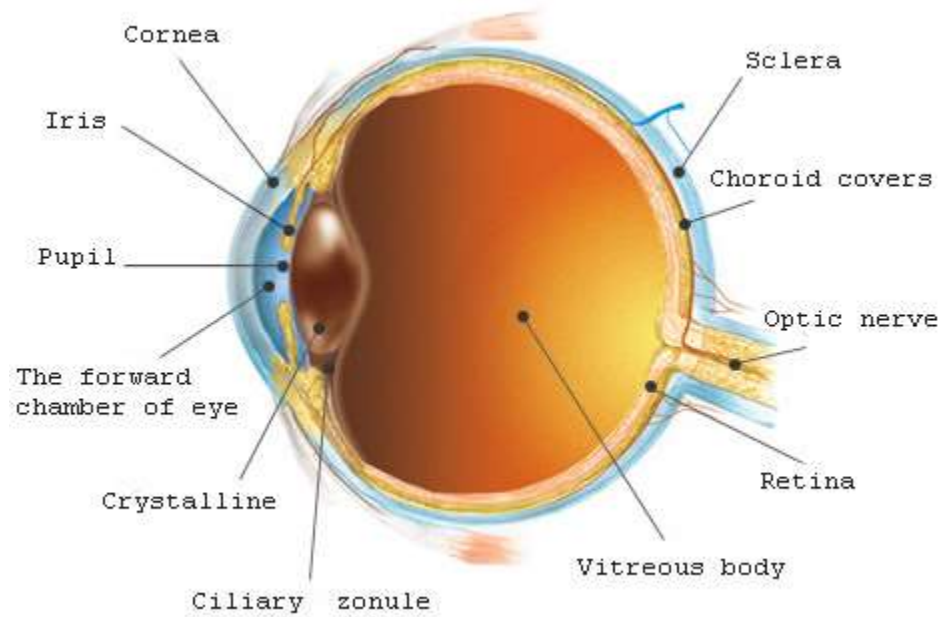
When light strikes the retina, the information about striking is transmitted to the primary visual cortex, available in the occipital lobe, located at the back of the head, shown in Figure 2.2. Most of the brain, as well as the eyes, include the complete visual system. Thus, for an organism to reliably extract the information in the environment, the whole eye-brain system must function properly.

### **2.2.2 The Retina**

Once the optical processing is performed, the eye converts the light into a neural activity, to make it compatible for the brain to perform further processing. This conversion is carried by photoreceptors available in the retina. These photoreceptors are specialized retinal cells, which get stimulated by the light energy. In general, photoreceptor cells are classified into two types [3], rods and cones. The rods are highly abundant (about 120 million), very sensitive to light, and are positioned all over the retina, except at its very center. The main purpose of the rods is to provide vision at very low light levels, known as scotopic conditions. On the other hand, the cones are less copious (about 8 million), much less sensitive to light. They are totally concentrated at the center of the retina, with a few dispersed through the periphery. The main purpose of the cones is to provide vision under normal lighting conditions, known as photopic conditions. Cones are even responsible for all the colors we experience. There is a small region at the center of retina, called the fovea, which contains nothing but heavily stacked cones. The fovea can cover only about 2 degrees of visual angle. Another region, known as the optic disk (or the blind spot), exists where the axons of the ganglion cells leave the eye at the optic nerve. This region contains no receptor cells at all. Blindness though, is not experienced there, but only under very special circumstances.

After the optical information is coded into neural responses, some preliminary processing is carried inside the retina itself. This processing is done by various other types of neurons, which include horizontal, bipolar, amacrine, and ganglion cells [3]. All these neurons integrate the responses from nearby cells, shown in Figure 2.3. The inputs are

received by the retinal ganglion cells, organized antagonistically, and are concentric pattern composed of a center and a surround region. The region of the retina, from which the ganglion cell receives the input, is known as the receptive field. The background signal is constantly emitted from the ganglion cell.



**Figure 2.1: A cross section of the human eye [4]**

The response from the retinal ganglion cell is drastically increased when the light strikes the photoreceptors in one region (on-response), while on the other hand, it generates the reduce response (off-response) when light falls on the other regions. In general, the ganglion cells are classified into two types, namely, the on-center cells and the off-center cells. The former are the cells where the center region is stimulated by an on-response, and the later are the cells where the center region is stimulated by an off-response (see Figure 2.4).

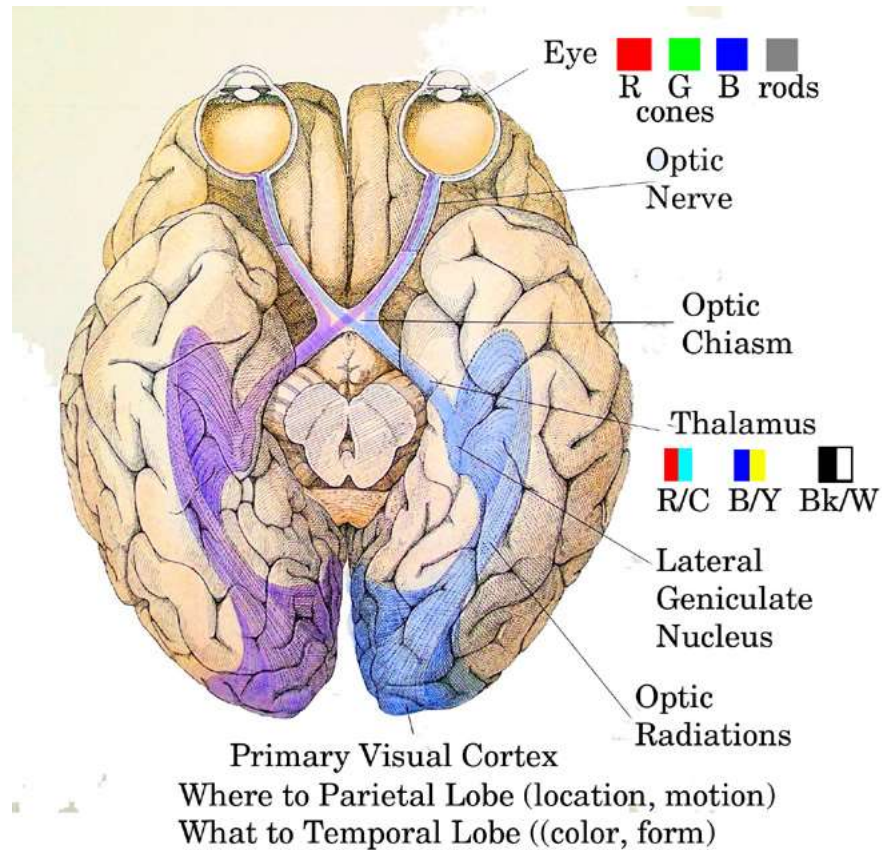


Figure 2.2: The human visual system [3]

The ganglion cells are provided with the axons, which convey information out of the eye, through the optic nerve, and into the optic chiasm [3]. In the optic chiasm, the fibers in each eye, positioned at the nasal side of the fovea, cross over to the opposite side of the brain, while the other fibers remain on the same side. This leads to the complete swapping from external visual fields to the cortex; i.e., the information from the left half of the visual field goes to the right half of the brain, whereas the entire information from the right visual field goes to the left half of the brain. There are two discrete passage ways from the optic chiasm into the brain, on either side. The smaller one leads towards the superior colliculus, a nucleus in the brain stem. This visual center is basically involved in



the control of eye movements, and also appears to process information about where things are in the world. The larger pathway leads initially to the Lateral Geniculate Nucleus (LGN) of the thalamus, and then to the occipital cortex (or primary visual cortex).

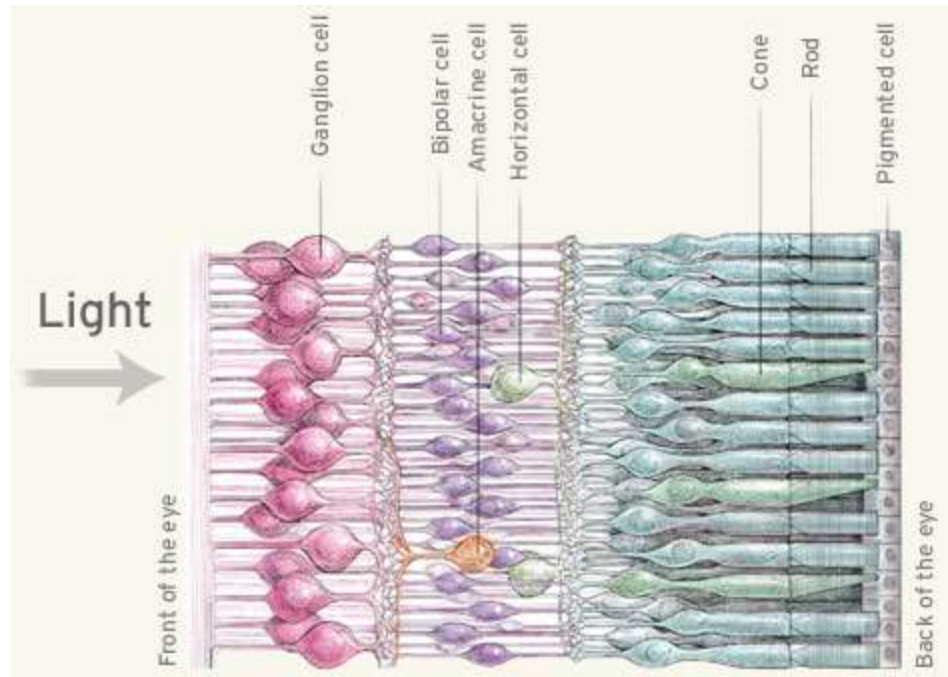


Figure 2.3: The human retina [3]

### 2.2.3 The Visual Cortex

The human cortex is separated into two halves, known as cerebral hemispheres, which are nearly symmetrical. Numerous neuropsychological studies [3] state that for the visual information, the occipital lobe is the chief cortical receiving area. It would be pure speculation to say that vision scientists understand all about how the visual cortex works. They are in fact starting to get some glimmerings of what the assorted pieces might be

and how they might fit together. Striate cortex, which is sometimes known as primary visual cortex or area V1, is the section, where the first steps in cortical processing of visual information takes place [3]. It covers the larger section of the occipital lobe, and even handles the most complex visual processing tasks. The input to striate cortex is from the LGN on the same side of the brain. Just like that of LGN, the visual input of striate cortex is completely crossed. Both sides are triggered by the thin central vertical strip, which measures about 1 degree of visual field. The cells on one side of the brain, which are sensitive to this strip, are connected to the corresponding cells on the other side of the brain, through the large fiber tract known as corpus callosum. The function of corpus callosum is to allow communication between the two cerebral hemispheres. The nearby regions on the retina, project to nearby regions in striate cortex. This makes the mapping from the retina to the striate cortex, topological in nature. The cortical magnification factor is the process in which central area of the visual field receives a much greater proportion in the cortex than the periphery. This area of the visual field falls on or near the fovea.

There are two pathways which are often referred as the “what” system and the “where” system. The “what” system represents the lower (ventral) system, wherein the inferior temporal centers seem to be involved in identifying objects, whereas the “where” system represents the upper (dorsal) system, wherein the parietal centers seem to be involved in locating objects [3]. It appears almost inevitable that these two diverse classes of information must get together at some place in the brain, so that the “what-where” connection can be established, but where this happens is not known yet.

It is clear that a good amount of visual processing is performed across different areas in parallel. Every section is projecting fibers to various other sections, but by no means to all of them. There is a bidirectional communication among these, that is, if area A projects to area B, then area B also projects back to A.

#### **2.2.4 The Physiological Pathways Hypothesis**

In recent years, the connection between anatomical structure and physiological function has started to unveil. Hypothesis is made, that, there are discrete neural pathways to process information regarding distinct visual properties which includes shape, color, motion and depth. Livingstone and Hubel [5] proposed that from the retina onwards itself, the afore mentioned four different types of information are processed in different neural pathways. The reports and evidences show that color, form, motion, and stereoscopic depth information are processed in separate sub-regions of visual cortical areas V1 and V2 as shown in Figure 2.5. These areas are further projected to different higher-level areas of cortex; i.e., color projected to area V4, movement and stereoscopic depth information to area V5 (also referred as MT, Medial Temporal cortex), and form to area IT (Infero-Temporal cortex) via various intermediary centers (including V4) [3]. The depth and motion pathways from these areas projects to the dorsal “where” system for object localization, whereas the form and color pathways to the ventral “what” system for object identification.



A. On-center, Off-surround

B. Off-center, On-surround

**Figure 2.4: Receptive field structure of ganglion cells**

On-center, off-surround cells (A) fire to light onset, and stop at offset in their excitatory center. But they stop firing to light onset, and begin firing at offset in their inhibitory surround. Off-center, on-surround cells (B) exhibit the opposite characteristics [3]

## 2.3 Visual Perception

Visual perception is defined as the process of extracting knowledge about the events happening in the environment and its objects. This is achieved by extracting information from the light they emit or reflect. Visual perception cares only about acquiring knowledge. This implies that, vision is fundamentally a cognitive activity [3], which is different from pure optical processes like photographic ones. When the eye is compared with a camera, significant similarities are found between these in terms of optical phenomena, but no similarities whatsoever in terms of perceptual phenomena. For instance, perceptual capabilities are totally absent in camera. Perception is not just about subjective visual experiences of the observer. The knowledge obtained through visual

perception specifies about the objects and events in the environment. Optical information is the basis of all vision. It is the information that is processed in visual perception coming from the light emitted or reflected by different objects.

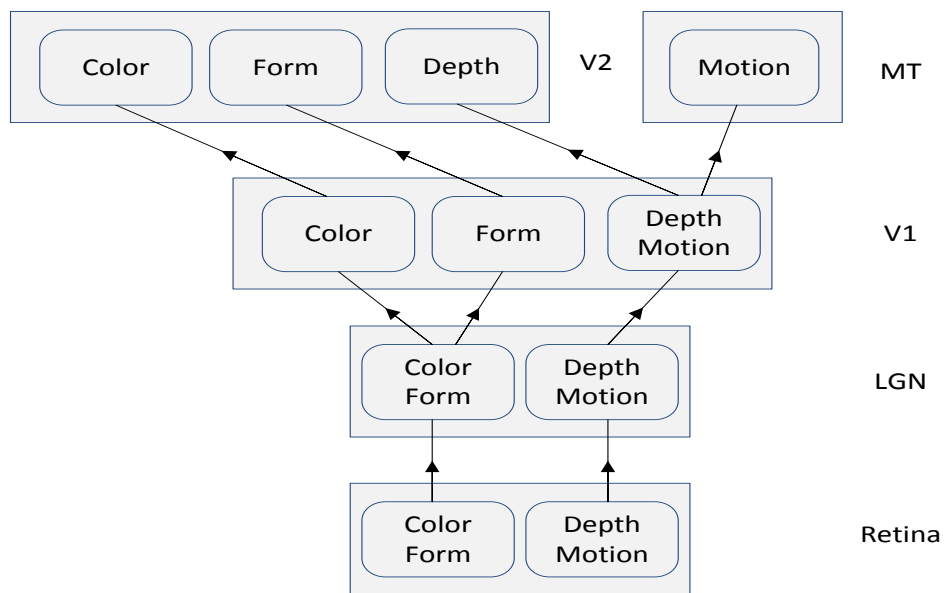
### **2.3.1 Adaptation of Visual Perception**

Visual perception adapts to certain conditions, and hence, changes over time. For instance, upon entering a very dark room during an afternoon daylight, we don't have a good vision. But after we spend some time in that ambience, say around 20 minutes, we can see very well inside. This attribute of sensitivity to light is known as dark adaptation [3]. In visual perception, adaptation is considered to be a very general phenomenon, and due to the protracted exposure of visual experience to varied range of stimulation: color, orientation, size, motion, etc., it may become less intense. The visual perception is not always a clear window onto reality. It is evident from the above mentioned changes in visual experience, that, for same physical environment, we have distinct visual experiences at different stages of adaptation.

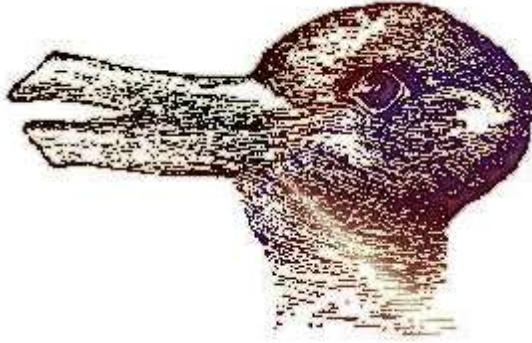
### **2.3.2 Ambiguous Perception of Objects**

Vision is an interpretive process, which provides us the information [3]. It somehow transforms the moving two-dimensional patterns of light, into stable perceptions of three-dimensional space, moving at the back of the eyes. The objects we perceive are not in fact the direct registrations of physical reality, but the interpretations based on the

structure of images. This aspect of vision exhibiting the interpretive nature comes from ambiguous figures. Ambiguous figures are single images, but can exhibit two or more distinct perceptions (for example, observe Figure 2.6). The interpretations out of such ambiguous figures are mutually exclusive; i.e., there can be only one perception at a time: a duck or a rabbit, not both. This is logical with the idea that perception leads to the formation of an interpretive model, because out of all models, only one such can be fit to the sensory data at one time. There would be no ambiguous figures, if perception was absolutely determined by the light stimulating the eye. This is because each pattern of stimulation would map onto a unique percept.



**Figure 2.5: Schematic diagram of the visual pathways hypothesis [3]**



**Figure 2.6: Ambiguous figures.**

Figure on the left can be seen as a duck (facing left) or a rabbit (facing right).

### **2.3.3 Visual Completion of the Environment**

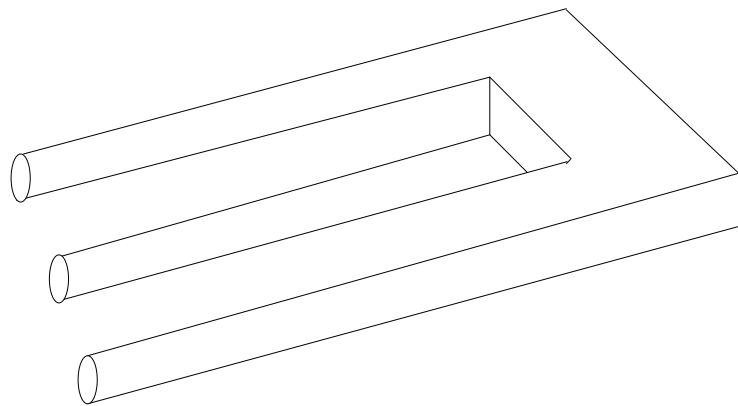
Perceptions of the people do not correspond to the sensory stimulation on which the models are based, but correspond to the models that their visual systems have constructed [3]. As a result, perceptions can sometimes be ambiguous and illusory, in spite of the non-illusory and unambiguous status of the raw optical images on which they are based. Perceptual models should be coupled to the information in the projected image of the world, and should be able to deliver fairly precise interpretations of the information.

Our perceptions contain portions of surfaces that we actually cannot see. This is the most substantial evidence that, visual perception involves the construction of environmental models. This perceptual filling-in, of parts of objects, concealed from our vision, is known as visual completion. It takes place automatically and smoothly every time we perceive the environment. Visual perception also fills in the information about the

surfaces of an object that is entirely hidden from view by its own visible surfaces. Such surfaces are referred to as self-occluded surfaces.

### 2.3.4 Perception of Impossible Objects

A two-dimensional line depiction in the beginning, gives a clear perception of coherent three-dimensional object, but can physically be impossible, as shown in Figure 2.7. This is a clear demonstration which backs the idea that, vision actively constructs environmental models instead of merely registering what is present. The physically impossible objects merely could not be perceived, if visual perceptions were simply a reliable reflection of the world. This indicates that, some visual processes work initially at a local level, and only later fit the results into a global framework.



**Figure 2.7: An impossible object.**

The figure above gives the perception of coherent three-dimensional object, but it is impossible to achieve physically.



### **2.3.5 Classification of Objects**

Our perceptual constructions go beyond filling-in the unseen surfaces [3]. They can even encompass information regarding the functional significance or the meaning of objects and situations. Once we are able to classify objects as members of known classes, we can reply to them in appropriate ways, based on the information collected from past experiences with similar objects. The past experience with members of a given class helps us in predicting what new members of the same class will do or will look like, with reasonable certainty. In a nutshell, we are able to address most new objects at a more abstract level of their class, even if we have never seen these particular objects before.

### **2.3.6 Attention and Consciousness**

The information present in the visible environment is much more than what one can fully perceive [3]. Therefore, we must be careful in what we attend to, because, what we opt will greatly depend on our plans, desires, goals, and needs. Perceptions are not driven merely by the nature of the optical information present in the sensory stimulation. Our perceptions are also influenced by higher-level goals, expectations, and plans, known as cognitive constraints. Based on what we are trying to achieve, we look at different things in our environs, as a result, we perceive them differently.

One of the functions of attention is to bring visual information to consciousness. Some attributes of objects are not experienced consciously, but they are attended. On the other hand, objects which are not attended are often processed completely out of

consciousness, to attract our attention. We become conscious of the detailed properties of the object as soon as it is attended. We can also identify it and distinguish its meaning in the present situation. Unlike lower levels of perception, the higher levels seem to be accessible to, or can be modified, by conscious knowledge and expectations. However, there is not much information about the role of consciousness in perception.

## **2.4 Summary**

In this chapter, some basic properties of HVS are discussed. This chapter gave an insight of the anatomical and the physiological properties of the visual system. We focused mainly on those aspects of the physiology, on which most of the vision quality models are based. An overview of the physical structure and several vital parts of the eyes, and their functioning in the visual system was given. The various aspects of the visual perception were also discussed. Such understanding is fundamental in linking the concepts of objective and subjective image quality assessment.

## **CHAPTER 3**

### **LITERATURE REVIEW**

In last few years, a great deal of research has been performed in the field of quality assessment, to develop effective objective quality metrics for both image and video. This chapter gives an overview of various algorithms proposed for image quality evaluation. The FR-IQA algorithms are discussed in section 3.1, followed by RR-IQA measures in Section 3.2, and NR image quality measures in Section 3.3. Various validation measures recommended by the Video Quality Experts Group (VQEG) [6], for the performance evaluations will be reviewed in Section 3.4.

#### **3.1 Full-Reference Image Quality Measures**

FR-IQA algorithms take both the reference and the distorted image as input, and yield an estimate of the quality of the distorted image as an output. The reference image is assumed to be a perfect image/distortion-free image. In the last decade, the classical approaches of FR-IQM received a lot of attention, due to their practical applications in the multimedia and communication area. A significant part of the literature discusses the development of FR image quality measures and various efforts and contributions made in this field. The major challenge is to develop “simple” but robust quality metrics which

can estimate visual quality of images with high reliability and high correlation with human observers.

The conventional FR image quality measures like the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) have found widespread use, given their simple mathematical foundations. These measures were among the very first FR-IQMs used in the literature.

Let an image  $I$  represent the original image, and  $\hat{I}$  be its distorted version under evaluation. Both the images are assumed to be 8 bits/pixel images. The concept has also been extended to color images [7]. The error between images  $I$  and  $\hat{I}$  is given by,  $E = I - \hat{I}$ . The MSE is simply expressed as:

$$MSE = \frac{\|E\|^2}{N} = \frac{1}{N} \sum_{i=1}^N E_i^2 \quad (3.1)$$

The PSNR on the other hand is expressed as:

$$PSNR = 10 \log_{10} \left( \frac{(2^n - 1)}{MSE} \right) = 20 \log_{10} (2^n - 1) - 20 \log_{10} \frac{\|E\|}{\sqrt{N}} \quad (3.2)$$

Here,  $\|\cdot\|$  represents the  $L_2$ -norm,  $E_i$  is the distortion value corresponding to pixel  $i$ , and  $N$  is the total number of pixels. The PSNR is effective, when the images being compared have different range of pixel values, but contains no new information regarding the MSE. Equations (3.1) and (3.2) show that, both the MSE and the PSNR are functions of the energy of the pixel-wise distortions  $\|E\|$ . Though, the MSE and the PSNR has been

widely used for a long time, they are considered ineffective and unreliable measures for visual quality estimation as they don't correlate well the human perceptual visual quality [8–11].

To develop visual models comparable to the HVS, intense efforts have been made to mathematically model the functional components of the HVS, related to the visual quality assessment. The HVS is a very complex system, and most of its functional properties are still not well understood. In the past few decades, serious research has been carried to understand the functioning of the HVS, and its other abilities related to vision [12], [13]. Some of the traditional image quality measures based on the HVS are discussed below.

The Lubin's model [14], [15] predicts image quality by estimating the probability of the difference between the images being compared. In order to estimate the probability map, a group of filters used to resample the image. This is done to simulate the functioning of eye optics and photoreceptor sampling process of the retina. The image is decomposed using a Laplacian pyramid [16], followed by band-limited contrast calculations [17]. The processing image is further decomposed by a set of steerable filters [18] to mimic the orientation selection of the HVS. The results are then normalized using the contrast sensitivity function. Finally, the convolution of normalized error signal and disk-shaped kernels is performed using a Minkowski pooling across scales. The error signal obtained after the pooling stage across the spatial space is represented as the probability of detection map.

Teo-Heeger *et al.* [19][20] proposed a model based on two components: a steerable pyramid transform [21] and a contrast normalization. Usually, the decomposition using a

steerable pyramid is done to mimic the channel decomposition in the HVS. The large number of neurons in the primary visual cortex is tuned to visual stimuli, with specific orientation, frequency, and orientation. The normalization scheme is inspired by the models that have widely been used to explain the physiology data in early visual systems. The detection method is a simple squared-error norm given by:

$$\Delta R = \|R_{ref} - R_{dist}\|^2 \quad (3.3)$$

Here,  $R_{ref}$  and  $R_{dist}$  are the vectors of normalized responses from the corresponding regions in the reference and distorted images respectively.

Watson *et al.* [22] developed a Discrete Cosine Transform (DCT) model, which was first designed to optimize JPEG compression. In this, the image is divided into a number of non-overlapping blocks. For every block, a visibility threshold is calculated for each DCT coefficients. The visibility threshold is calculated using three factors, which include baseline contrast sensitivity, masking of contrast/texture, and masking of luminance. This is done to simulate the properties of the HVS. The purpose of determining the visibility threshold is to normalize the error between the reference and the distorted image signals. Finally, the error is pooled spatially and across frequency to estimate image quality. These threshold units are commonly referred as ‘Just Noticeable Differences’ (JND), represented by  $d_{ijk}$ . In spatial error pooling, the JNDs for a specific frequency  $\{i, j\}$  over entire blocks  $k$  are pooled to give a perceptual error metric.

$$P_{ij} = \left( \sum_k |d_{ijk}|^{\beta_s} \right)^{1/\beta_s} \quad (3.4)$$

Here,  $\beta_s$  is the exponent in the Minkowski metric.

In frequency error pooling, the perceptual error metric is optimized by Minkowski metric of different exponent  $\beta_f$  and is given by:

$$P = \left( \sum_{ij} p_{ij}^{\beta_f} \right)^{1/\beta_f} \quad (3.5)$$

Karunasekera *et al.* [23] proposed a model, based on the HVS sensitivity to horizontal and vertical edge artifacts due to DCT-based compression. The edge error is calculated by averaging the absolute transform error over the entire image and is given by:

$$E_{vedge} = mean(|e_t|) \quad (3.6)$$

In the above equation,  $e_t$  is the non-linearly transformed masked error and  $E_{vedge}$  is the vertical edge error.

Miyahara *et al.* [24] developed a picture quality scale, which combines a group of properties associated to the HVS, for both global features and localized distortions. This includes the adaptation to light, contrast sensitivity, and visual masking of the HVS. For color images, Winkler *et al.* [25] proposed a distortion metric based on the following properties of visual perception: the theory of opponent colors and color perception, the response properties of neurons in primary visual cortex, the contrast masking, and the contrast sensitivity of HVS. The distortion measure was given by:

$$\Delta s = \sqrt[\beta]{\sum |s_0 - s_1|^\beta} \quad (3.7)$$

Here,  $s_0$  and  $s_1$  are the sensor outputs from the reference and processed image respectively. The amount of perceptual distortion measure  $\Delta s$  is transformed using equation (3.8), to quantify the measure of perceptual distortion on a scale of 0 to 5 (low to high quality)

$$Q = \frac{5}{1 + K\Delta s} \quad (3.8)$$

Here,  $K$  is a model parameter, which is properly chosen to ensure proper mapping [25].

Damera-Venkata *et al.* [26] developed an algorithm which models degradation as a linear frequency distortion and additive noise. Thus, two distinct measures were developed to compute both the distortions separately. Frequency distortion is calculated based on the model of frequency response of the HVS over visible frequencies. The additive noise distortion is calculated using the following properties of the HVS: the variation in the contrast sensitivity, variations in the mean of local luminance, contrast interactions between spatial frequencies, and the contrast masking effect of the HVS. The frequency distortion measure is given by:

$$DM = \int_0^{f_{\max}} \left[ 1 - DTF \left( \frac{f_r}{f_N} \right) \right] CSF(f_r) df_r \quad (3.9)$$

In the above equation,  $f_r$  and  $f_{\max}$  are the radial frequency and the maximum radial frequency respectively, and  $f_N$  is the Nyquist frequency. The terms DTF and CSF are the Distortion Transfer Function and the Contrast Sensitivity Function respectively.



Wang *et al.*[27] developed an image quality metric in the wavelet domain, known as the foveated wavelet image quality index. The metric is based on certain functional properties of the HVS; i.e, the space variance of contrast sensitivity function, space variance of local visual cut-off frequency, variance of human visual sensitivity in different wavelet sub-bands, and the influence of the viewing distance on display resolution. The Foveated Wavelet image Quality Index (FWQI) is represented as a function of Foveated Wavelet image Distortion (FWD) and is given by:

$$FWD = \left( \frac{1}{M} \sum_{n=1}^M [S(v, x_n) \cdot |c(x_n) - c'(x_n)|] \right)^{1/q} \quad (3.10)$$

In the above equation,  $M$  represents the number of wavelet coefficients,  $c(x_n)$  and  $c'(x_n)$  are the  $n^{th}$  wavelet coefficients of the reference and the compressed images at position  $x_n$  in the wavelet domain respectively. The foveated wavelet image quality index is given by:

$$FWQI = \exp(-FWD) \quad (3.11)$$

Lin *et al.* [28] presented a distortion metric, based on the noticeable local contrast changes as perceived by the HVS. It is achieved by discriminately analyzing the influence of the pixel difference on visual quality. A novel formula for the adjustment to luminance adaptation is proposed, which uses a block classification for contrast masking of the HVS. The proposed bottom-up visual distortion metric is given by:

$$D = \alpha_1 \overline{c_e^-} - \alpha_2 \overline{c_e^+} + \alpha_3 \overline{c_{ne}} \quad (3.12)$$

Here,  $\overline{c_e^-}$ ,  $\overline{c_e^+}$  are the averages of local contrast decrease and increase of edge pixels respectively and  $\overline{c_{ne}}$  is the average of local contrast change of non-edge pixels.

Chandler *et al.* [29][30] developed a metric which quantifies visual fidelity of natural images based on the near-threshold and the supra-threshold properties of the HVS. The metric is known as the Visual Signal to Noise Ratio (VSNR). First, the metric identifies whether the distortion is visible by comparing it with a contrast threshold. The contrast threshold is computed via wavelet based models of visual masking and visual summation. On the other hand, if the distortions are less than the threshold values of visual detection, the distorted image is considered to be of good visual fidelity. If the distortions are observed to be supra-threshold, the distortions are quantified by the low-level visual properties of observed contrast and the mid-level visual properties of global preference. The VSNR is given by:

$$VSNR = 10 \log_{10} \left( \frac{C^2(I)}{VD^2} \right) = 20 \log_{10} \left( \frac{C(I)}{\alpha d_{pc} + (1 - \alpha) \frac{d_{gp}}{\sqrt{2}}} \right) \quad (3.13)$$

Here,  $d_{pc}$  is the perceived contrast of the distortion,  $d_{gp}$  is the measure of extent of global precedence that is disrupted and  $C(I)$  is the RMS contrast of the original image  $I$ .

In contrast to the approaches based on “bottom-up” HVS, the “top-down” image quality measures assume the HVS as a black box. For some input to the black box, some output response is produced. So, the only concern is the input-output relationship of the HVS. It is based on the postulates related to the functionality of the HVS. The computational

models of top-down approaches are much simpler than bottom-up models for image quality assessment.

Sheikh *et al.* [31–34] developed a visual information fidelity approach based on the statistical source modeling, channel distortion, and a receiver. The model relates the quality of an image to the amount of information that is shared between the reference and the distorted image. A reference image (the source) is modeled by a Gaussian scale mixture in the wavelet-domain [35]. The degradation between a reference and a distorted image (the channel distortion) is modeled as the amalgamation of uniform energy attenuation in wavelet-domain with the additive noise. Finally, the visual distortion process (the receiver) is modeled as a stationary, zero-mean, additive white Gaussian noise process in the wavelet domain. The image quality is predicted by quantifying the information shared between the reference and the distorted mediums.

Weken *et al.* [36] developed an algorithm based on the concept of fuzzy logic. The model resorts to similarity measures using fuzzy set theory. For image quality prediction, the disjoint patches between the reference and the distorted images are compared. The weighted average of the local similarities between the patches gives an approximation to image quality. Thirteen fuzzy similarity measures were discussed in [36], and were then used to measure local similarities. The weight is defined as a similarity between the homogeneities of corresponding patches of the image. The homogeneity is calculated as the similarity between the maximum and minimum intensities of an image patch. The weighted average of the similarities between the disjoint parts of an image is given by:

$$S^n(A, B) = \frac{1}{N} \sum_{i=1}^N w_i S(A_i, B_i) \quad (3.14)$$

In the above equation,  $S(A_i, B_i)$  represents the similarity between the image parts  $A_i$  and  $B_i$  of the images  $A$  and  $B$  respectively.

Shnayderman *et al.* [37] developed an image quality measure based on the Singular Value Decomposition (SVD) of the image. The SVD of an image can either be used as a scalar or a graphical measure. In [38], Han *et al.* used the LU factorization to represent the structural information of an image. A two-dimensional distortion map is obtained using LU factorization of the reference and the distorted images. This map forms the basis for quality prediction. The resulting image quality measure is given by:

$$MLU = \frac{1}{\left(\frac{W}{N}\right) \cdot \left(\frac{H}{N}\right)} \sum_{j=1}^{\left(\frac{W}{N}\right) \cdot \left(\frac{H}{N}\right)} |D_j^U - D_{med}^U| \quad (3.15)$$

Here,  $W$  and  $H$  represents the image size, and  $N$  represents the size of image block.  $D_j^U$  is the distortion map of  $j^{th}$  block and  $D_{med}^U$  is just its median.

Bouzerdoun *et al.* [39] presented a neural network-based IQA algorithm for JPEG/JPEG2000 compression distortions. In this, a set of key features are extracted from the reference and test images to train neural network for image quality prediction. The image is first divided into blocks of size  $16 \times 16$ , and six statistical features are extracted from each block as an input to the network. The network is trained to estimate the image quality of corresponding block, and the average of estimated qualities of all individual blocks gives the overall quality of an image. The six statistical features extracted from the

original and images include: the two means, the two standard deviations, the covariance, and the mean-squared error between the reference and the test blocks.

### **3.2 Reduced-Reference Image Quality Measures**

Reduced-Reference IQA methods are preferred in applications where the reference image is not fully accessible. Instead of the entire reference image, a fractional part of some features are used for quality prediction. The partial information is the set of parameters extracted from the reference image. The RR measures provide a compromise between FR and NR quality measures. They are easier than NR measures at the trade-off of transmitting additional information. In conventional RR image quality measures, the side information is transmitted through a separate data channel, which makes it cost ineffective in real time applications. An alternative solution is to transmit the side information through the same channel, along with the image being transmitted. The parameters being extracted from the reference image are selected based on certain specific criteria. The features must provide enough information about the reference medium, they should be sensitive to image distortions, and should even provide good perceptual relevance. At the receiver, the quality score is predicted based on the features extracted from both the reference and the distorted images.

Based on the natural image statistics in the wavelet transform domain, Wang *et al.* [40] developed a robust algorithm for RR image quality assessment. The concept of the Kullback-Leibler distance [41] was applied between the marginal distributions of the wavelet coefficients of a reference and a distorted image. The marginal distribution was

represented by the Generalized Gaussian Density (GGD) model [42]. The quality measure is obtained by fitting the error between the wavelet coefficients of the distorted image and the Gaussian distribution of the reference image. This quality model motivated to develop a subsequent quality-aware model in [43], which combines the techniques of information data hiding, robust image communication, information data decoding, and RR quality measures. The basic idea of such a quality-aware image system is to implant the features extracted from an original image into the image data as a hidden message. At the reception end, the received distorted image is decoded to extract the hidden message, and then predict image quality using different RR quality measures. The distortion between the original and the distorted image is quantified using following equation:

$$D = \log_2 \left( 1 + \frac{1}{D_0} \sum_{k=1}^K \left| \hat{d}^k (p^k \| q^k) \right| \right) \quad (3.16)$$

In the above equation,  $D_0$  is the constant used to control the scale of distortion measure,  $k$  is the total number of sub-bands,  $p^k$  and  $q^k$  are the probability density functions of the sub-band  $k$  of the original and the distorted image, and  $\hat{d}_k$  is the KLD's estimation between  $p^k$  and  $q^k$

Chetouani *et al.* [44] proposed a RR image quality assessment algorithm based on feature extraction and neural networks training. In the first stage, an original image and its degraded version are decomposed using 3-level wavelet decomposition, and an edge map is derived at each decomposition level. The mean and the standard deviation are extracted as the statistical features from each decomposition level, giving a total of 6 features from each image. Overall, 12 features (6 from original and 6 from distorted image) are

extracted for image quality assessment. In the second stage, the neural network is trained using these extracted features against their human observer scores for quality prediction.

The edge map of an image at each decomposition level is given by:

$$EM_{ap}(k) = \sqrt{CH(k)^2 + CV(k)^2 + CD(k)^2} \quad (3.17)$$

Here,  $CH, CV$ , and  $CD$  are the horizontal, vertical, and diagonal details of an image respectively.

Gao *et al.* [45] presented a RR framework which incorporates various concepts like multi-scale geometric analysis, the contrast sensitivity function, and the Weber-Fechner law of Just Noticeable Difference (JND). The multi-scale geometric analysis is used to decompose the image for feature extraction, and to mimic the multichannel structure of the HVS. The multi-scale geometric analysis offers a series of transforms, to extract different types of image geometric information. The contrast sensitivity function is used to apply weights to the coefficients obtained by the multi-scale geometric analysis. This is performed to simulate the nonlinearities observed in the HVS. Finally, the JND is used to produce a noticeable variation in sensory experience. The quality metric of the distorted image is given by:

$$Q = \frac{1}{1 + \log_2 \left( \frac{S}{Q_0} + 1 \right)} \quad (3.18)$$

Here,  $Q_0$  is a constant, which controls the distortion measure scale, and  $S$  represents the city block distance given by:

$$S = \sum_{n=1}^L P_R(n) - P_D(n) \quad (3.19)$$

Here,  $P_R(n)$ ,  $P_D(n)$  are the normalized histograms of the reference and distorted image respectively.

### 3.3 No-Reference Image Quality Measures

Unlike FR and RR image quality measures, NR algorithms attempt to evaluate quality without the use of any information from reference image. The absence of any information about the reference image makes the task of the NR-IQA very challenging. Despite such a complex task, NR quality measures are widely preferred for real time applications.

In the last few decades, the field of NR-IQA has rapidly emerged. A large number of algorithms have been proposed that widely vary in performance. Most of the existing NR-IQA models are provided with the prior knowledge of the distortion type. Fortunately, in many multimedia applications the distortion process is known, thus, the task of modeling a distortion-specific NR quality measures becomes feasible.

The most commonly observed distortions in multimedia applications are blur and noise. These distortions are generated during image acquisition and in display systems. Thus, the conventional NR quality measures were designed to quantify blur and noise distortions in visual signals. The use of compression techniques (lossy/lossless), for bandwidth reduction too, resulted in reduced visual quality of reconstructed images. In lossy compression techniques, the lost data during coding cannot be recovered. Due to



which, some unwanted artifacts are observed in the retrieved image after reconstruction. In fact, the compression artifacts and the errors during transmission are the two major degradations observed in the experiments conducted by the Video Quality Experts Group (VQEG) [6]. This group focuses on providing industrial standards to the video quality assessment community. The remaining part of this section gives an in-depth analysis of some traditional NR image quality measures.

### **3.3.1 Measures used for distortion due to JPEG compression**

In block DCT-based JPEG compression [46], the image is divided into  $8 \times 8$  non-overlapping blocks. The discrete cosine transform is applied to every block, and the DCT coefficients in each block are quantized independently. Finally, the quantized coefficients are coded using entropy coding. The coarse quantization of the block-based DCT coefficients, especially at low bit rates, leads to the occurrence of unwanted artifacts in the reconstructed image. These artifacts are the blurriness within blocks and the appearance of the blockiness at the block boundaries [47]. The blurriness is due to the loss of high frequency components during quantization within each block. The independent quantization within each coding block, leads to the formation of periodic horizontal and vertical discontinuities at block boundaries. The blocking artifact is considered to be the most bothersome distortion in JPEG compression [48]. Thus, most of the NR quality measures are modeled based on the quantification of blocking artifacts, either in the spatial domain [49–53] or in the frequency domain [54–56]. Besides this, some other NR quality assessment measures of JPEG compressed images use machine

learning techniques to model the relationship between image features and subjective quality ratings using a training process [57].

Wu *et al.* [53] developed an algorithm based on the luminance masking effect, and designed a quality metric known as the generalized impairment metric. The quality metric quantifies the blocking artifacts observed at the block boundaries. Based on the subjective experiments performed in [49], it was stated that this single distortion is sufficient to assess the quality of the JPEG compression images. Based on this notion, Meesters *et al.* [49] proposed a new image quality metric based on the blocking artifact itself. The Hermite transform [58] is a signal decomposition technique, which approximates the signals by polynomials within a Gaussian window. The Hermite transform was used to measure the low-amplitude edges due to blocking artifacts, and the edge amplitudes were estimated as the representative of blockiness.

Wang *et al.* [51] proposed a model based on some distinct features from the distorted medium. These features include: the average luminance differences across block boundaries, and the measure of the activity within each block. The activity of an image is measured using the average absolute difference between in-block pixels and by the zero-crossing rate. All these factors are pooled to approximate a quality score for JPEG compressed images. The non-linear equation used to predict quality is given by:

$$S = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3} \quad (3.20)$$

In the above expression,  $B, A, Z$  are the blocking factor, activity and zero-crossing rate, and  $\alpha, \beta$  and  $\gamma$  represent the model parameters that are estimated through a simple regression technique.

In [52], the authors claimed that only the edge detection around block boundaries is sufficient to quantify blocking artifacts. This is based on the assumption that occurrence of the horizontal or vertical edge exactly at the block boundary is less in natural images. Thus, Li [52] proposed to use the Prewitt operator for the detection of horizontal and vertical edges across the image. The edges due to the blocking artifacts are usually weak edges, and thus, all the gradients below a pre-defined threshold are selected to quantify the blocking effect.

The quality metric based on both blockiness and blurriness was discussed in [50]. The blockiness was evaluated based on inter-pixel differences at and near the block boundaries, while, the blurriness/flatness was measured using the zero-crossing rate within each  $8 \times 8$  regions. The contrast and spatial masking effects were also incorporated for quality assessment. The quality map obtained was a function of inter-block difference  $B_{BLK}$  and inter-block flatness measure  $Z_{BLK}$  across the blocks of the test image and was given by:

$$Q_{BLK} = \begin{cases} \max(B_{BLK}, Z_{BLK}) & \text{if } \max(B_{BLK}, Z_{BLK}) > T_{JND} \\ 0 & \text{otherwise} \end{cases} \quad (3.21)$$

Here,  $T_{JND}$  is the threshold of local pixel activity based on average local Just Noticeable Distortion (JND). The mean of  $Q_{BLK}$  for all the blocks over an entire image gives the overall quality measure given by:

$$Q_{image} = average(Q_{BLK}) \quad (3.22)$$

Wang *et al.* [54] proposed a model to assess the blocking effect by evaluating the energy of the blocky signal in the frequency domain. Since, the analysis is carried in the frequency domain, the Fast Fourier Transform (FFT) was used to evaluate the power spectra of the absolute difference signals, in either horizontal or vertical direction. The blockiness was represented as the power of blocky signal, which was computed after the smoothing the power spectrum using median filter. The luminance and texture masking effects were also incorporated in the model design. The blockiness of an image given by:

$$\begin{aligned}
 A &= \left( P \left[ \frac{N}{8} \right] - P_M \left[ \frac{N}{8} \right] \right) + \left( P \left[ \frac{N}{4} \right] - P_M \left[ \frac{N}{4} \right] \right) \\
 B &= \left( P \left[ \frac{3N}{8} \right] - P_M \left[ \frac{3N}{8} \right] \right) + \left( P \left[ \frac{N}{2} \right] - P_M \left[ \frac{N}{2} \right] \right) \\
 M_{Bv} &= \frac{8}{7} \{A + B\}
 \end{aligned} \tag{3.23}$$

In the above equation,  $N$  represents the column size of an image and  $P$  denotes the power spectrum. The blockiness over the entire image is given by:

$$M_B = \frac{M_{Bv} + M_{Bh}}{2} \tag{3.24}$$

Here,  $M_{Bv}$  and  $M_{Bh}$  represent the vertical and horizontal blockiness measure of an image.

Liu *et al.* [55] also developed a DCT based algorithm to quantify the blocking artifact. The blocking effect is represented as a 2-D step function within the shifted blocks, constructed across two adjacent coding blocks. The amplitude to the 2-D step function represents the strength of the blockiness. It also takes into account the luminance and

texture masking effects for quality measurement. The locally measured blockiness is pooled to provide the overall quality measure given by:

$$\Theta = \sqrt[p]{\frac{1}{N} \sum_{k=1}^N \eta_k^p} \quad (3.25)$$

Here,  $\Theta$  is the overall blocking artifact measure of the test image,  $N$  is the total inter-block boundaries, and  $\eta$  represents the perceptibility of blocking artifact across blocks.

Brandão *et al.* [56] proposed a IQA model using the Natural Scene Statistics (NSS) of the DCT coefficients. The distribution of the DCT coefficients is modeled by the Laplace probability density function. The resulting coefficients distribution is used to estimate the local error due to JPEG encoding. Watson's model [22] is used to quantify these local errors, which are finally pooled to give an image quality score. The distortion measure over the entire image is given by:

$$D_w = \sqrt[4]{\frac{1}{M} \sum_{k=1}^N \sum_{i=0}^7 \sum_{j=0}^7 \hat{\varepsilon}_{pk}(i, j)^4} \quad (3.26)$$

In the above expression,  $D_w$  is the global metric,  $N$  represents the number of coefficients, and  $M$  represents the total number of coefficients under analysis

Venkatesh *et al.* [57] developed an image quality measure based on learning algorithms. The algorithm uses the Growing And Pruning Radial Basis Function (GAP-RBF) network for quality prediction. Like any other learning networks, this network is also trained to approximate the functional relationship between the extracted features and the subjective quality scores. The features extracted were edge map, background activity

mask, and background luminance weight. The score predicted by the quality model is given by,

$$\hat{Q} = \sum_{i=1}^K \alpha_i \exp\left(-\frac{1}{\sigma_i^2} \|U - \mu_i\|\right) \quad (3.27)$$

In the above expression,  $U$  is the feature set extracted from the test image,  $K$  is the number of Gaussian neurons,  $\mu_i$  is the center vector of  $i^{th}$  Gaussian neuron,  $\sigma_i$  represents the width of the neuron, and  $\alpha_i$  represents the weight of the  $i^{th}$  Gaussian neuron connecting the hidden layer and the output neuron.

### 3.3.2 Measures used for distortion due to JPEG2000 compression

In the wavelet-based JPEG2000 compression algorithm [59], the blurring and the ringing artifacts are introduced due to coarse quantization of the Discrete Wavelet Transform (DWT) coefficients. These artifacts are considered to be highly substantial in JPEG2000 compressed images. The coarse quantization truncates the high frequency DWT coefficients, which leads to the visible discrepancies across the edges in spatial domain. These discrepancies are referred as ringing artifacts due to their natural appearance. The ringing artifact results in appearance of ripples and oscillations across contours and sharp edges in an image. These artifacts can go from an imperceptible range to the extremely annoying level depending on the compression rate. In contrast to blocking artifacts, the ringing artifacts significantly depend on the content of image and degree of compression,

and follow no regularity. Some NR image quality measures for JPEG2000 compressed images are discussed in the following section.

In [60], the authors proposed to isolate those regions of an image where the ringing artifacts are visually prominent, while preserving the genuine edge information and fine details. The traditional binary morphological operators were used for this purpose. The image intensity variance was computed to evaluate the effects of ringing artifacts around the vicinity of the edges.

Li *et al.* [52] proposed an algorithm to measure the ringing effect by evaluating the noise spectrum filtered out by anisotropic diffusion [61]. The ringing artifacts are commonly integrated into the noise spectrum, and the noise spectrum can be colored when the image contains ringing artifacts. The percentage of energy across high frequencies gives the strength of the ringing effect.

Tong *et al.* [62] proposed an algorithm based on the concept of Principal Component Analysis (PCA). The local features of the compressed image are extracted using PCA, by assuming that all the edge points are either “distorted” or “un-distorted”. The relationship between the local features and the local distortion metric is modeled based on the probability of an edge point being “distorted” or “un-distorted”. The overall distortion metric of the test image is given as a function of local distortion metric:

$$Dm = \frac{1}{N_{edge}} \sum_{(i,j) \text{ edge points}} Ld(i, j) \quad (3.28)$$

In the above equation,  $N_{edge}$  is the total number of edge points in the test image, and  $Ld, Dm$  represent the local and overall distortion metric of the image under analysis. The

quality metric is modeled to represents the quality score on an image, and is given by the equation (3.37), where  $s$  is replaced by  $Dm$ .

In [63], the blurring measure is modeled as the ratio of edge activity weighted by the probability of edge occurrence in the middle/low frequencies, while the ringing measure is modeled as the ratio of the activity in the middle low over middle high frequencies in the ringing regions around high strong edges. Finally, all the measures: the ringing measure, blurring measure and the percentage of strong edges, are combined to give an image quality score given by:

$$PQ = a_1 + a_2.BM + a_3RM.BM + a_4.EM \quad (3.29)$$

Here,  $BM, RM, EM$  represent the blurring, ringing and image features respectively, and weights  $a_i, i = 1 to 4$  are estimated from training set using minimum MSE estimate between predicted and subjective scores.

Sheikh *et al.* [64][65] proposed a model for quality assessment of JPEG2000 compressed images. The model operates in the wavelet domain, and incorporates the natural image scene statistics model and an image distortion model for the quality prediction. The idea was to analyze how the quantization process of JPEG2000 compression influences the statistics of the wavelet coefficients. The distribution of the wavelet coefficients is described by using NNS model [66]. A distortion model related to quantization is incorporated to quantify the departure from the natural image statistic model. The sub-band probabilities of all the bands are calculated to give an image quality metric. The quality metric is modeled to approximate the quality score, and is given by:



$$q_i = K_i \left( 1 - \exp \left( - \frac{(p_{ss,i} - u_i)}{T_i} \right) \right) \quad (3.30)$$

Here,  $q_i$  is the predicted quality score and  $p_{ss,i}$  is the  $p_{ss}$  probability of  $i^{th}$  sub-band,  $u_i$   $T_i$   $K_i$  are the model parameters obtained after curve-fitting. The predicted quality score of all the sub-bands,  $q = \{q_i | i \in 1 \dots 6\}$ , are pooled to approximate the quality score of entire image, and is given by:

$$Q = q'^T w \quad (3.31)$$

where  $q' = \begin{bmatrix} \frac{(q_1 + q_2)}{2} \\ q_3 \\ \frac{(q_4 + q_5)}{2} \\ q_6 \end{bmatrix}$  and  $w$  is the learned weights to reduce prediction error.

Sazzad *et al.* [67] proposed a quality assessment model based on the pixel distortions and edge information. The pixel distortions are estimated using the local standard deviation ( $S$ ) and the absolute difference measure ( $A$ ) of a central pixel from the second closest neighborhood pixels. The edge information is estimated using the zero-crossing rate ( $Z$ ) and a histogram measure, with and without edge preserving filters. Finally, both the measures are pooled to assess the image quality. The model used to pool all the features is given by:

$$\begin{aligned}
A &= [\gamma_1 \log(S+1) + \gamma_2 \log(A+1) + \gamma_3 \log(Z + \gamma_4)] \\
B &= [\gamma_5 \log(H_f+1) + \gamma_6 \log(V_f+1) + \gamma_7 \log(H+1) + \gamma_8 \log(V+1) + \gamma_9] \\
C &= A.B
\end{aligned} \tag{3.32}$$

In the above equation,  $H_f, V_f, H, V$  are horizontal and vertical histogram measures with and without edge preserving filters, and  $\gamma_1$  to  $\gamma_9$  are model parameters estimated using test data and optimization algorithm. The quality metric is modeled to approximate the quality score, and is given by:

$$MOS_p = \frac{b_1}{1 + \exp[-b_2(C - b_3)]} + b_4 \tag{3.33}$$

Here,  $b_1, b_4$  are the model parameters of the logistic function obtained through simple curve fitting.

The algorithm proposed in [68], used the ringing artifacts to assess the quality of JPEG2000 compressed images. To quantify the ringing effect, the ringing region detection method [69] was used to identify those regions that are likely to be affected by the ringing artifacts. A Ringing Annoyance Score (RAS) is assigned to every detected ringing region, which is calculated by estimating the local visibility of ringing artifacts and comparing it with local background activity. The RAS over the entire image is pooled to give an overall quality metric. The ringing region detection model [69] consists of two steps: extraction of edges related to the ringing artifacts, and detection of possible ringing regions. For edge extraction, a bilateral filter-based advance edge detector [70] is used to extract the possible edges related to ringing regions. The Canny edge detector

[71] is applied on the filtered image to obtain the most relevant edges. The mean of the  $RAS$  is given by:

$$MRAS = \frac{1}{T} \sum_{n=1}^N RAS(RO_n) \quad (3.34)$$

Here,  $RO$ 's represent the ringing objects,  $N$  is the total number of  $RO$ 's and  $T$  is the total number of pixels in each of the  $N$   $RO$ 's.

### 3.3.3 Measures used for distortions due to blur and noise

Wu *et al.* [72] presented a quality metric for out-of-focus blurred images. For the quality assessment, the idea was to extract a point spread function (PSF) from the line spread function (LSF) of the blurred image. The sharp edges in the image are identified and the LSF is extracted from these edges. The radius of the PSF parameters is evaluated from the LSF, which can be used as a criterion to measure degree of blur. The radius of the PSF is given by:

$$R = 2\sigma = 2\sqrt{\sum_1^N (x - \bar{x})^2 d_l(x)} \quad (3.35)$$

Here,  $d_l(x)$  is the LSF distribution and  $R$  is the radius of the PSF function.

The blur measures reported in [73], [74] are based on the measurement of edge spread. Edge spread is defined as “the number of pixels with monotonically changing intensities along the gradient orientation at an extracted edge pixel”. Both blur measures differ in one simple way. In [73], the edges are detected using the Sobel operator, while in [74],

the Canny operator [71] is used. In [73], the edge spread is measured along the horizontal direction of the vertical edge, while in [74], it is measured along the gradient orientation at a general pixel. Finally, the edge spread over entire image is given by,

$$s = \frac{\text{sum of all edgewidths}}{\text{Number of edges}} \quad (3.36)$$

The quality of an image is approximated using a non-linear equation given by:

$$Q = \alpha + \beta s^\gamma \quad (3.37)$$

Here,  $\alpha, \beta$  and  $\gamma$  are the model parameters obtained using simple regression techniques.

Chong *et al.* [75] developed an eigenvalues-based image sharpness metric. In this approach, the effect of image contrast is first minimized by normalizing the image with its energy. The covariance matrix is then obtained from the normalized image, which is analyzed using SVD to obtain its eigenvalues. The first six largest eigenvalues are computed to get the sharpness score of an image.

$$M_E = \text{trace}[\Lambda_k] \quad (3.38)$$

Here,  $M_E$  is the sharpness metric and  $\Lambda_k$  represents the  $k$ -dimension eigenvalue matrix.

Vu and Chandler [76] proposed a sharpness metric based on the spectral and spatial properties of the image. The image is first divided into blocks, and the slope of the magnitude spectrum and the total spatial variance, for each block are measured. These measures are tuned to account for visual perception. The tuned measures are then combined to yield an overall perceived sharpness map given by:

$$S_3(X) = S_1(X)^\gamma \times S_2(X)^{1-\gamma} \quad (3.39)$$

Here,  $S_1(X)$  and  $S_2(X)$  are the spectral and spatial- based sharpness maps, and  $S_3(X)$  is the overall sharpness map of the test image.

Chen and Bovik [77] developed an IQA model for blurred images based on the Natural Scene Statistics (NNS) and a multi-level wavelet decomposition. A probabilistic SVM classifier was used to classify the image as either “blurred” or “sharp”. The image is then decomposed to obtain the *detail map*, which is a combination of horizontal and vertical responses in high frequency band. The detail map improves the quality assessment process. The sharpness score of the blurred image is given by:

$$BlurQualityScore = (QS - SVM)^{\gamma_0} \prod_{i=1}^N (DS_i)^{\gamma_1} \quad (3.40)$$

In the above equation,  $DS_i$  is the detail score evaluated from the detail map and  $QS - SVM$  is the score from probabilistic SVM .

Vu and Chandler [78] also proposed a wavelet-based quality model to estimate the global and local sharpness in an image. The image was first decomposed using the three-level DWT and the log-energies of each sub-band are measured. The weighted average of all the log-energies is measured to give an overall sharpness metric of the image. The resulting Fast Image SHarpness (FISH) index is given by:

$$FISH = \sum_{n=1}^3 2^{3-n} E_n \quad (3.41)$$

Here,  $n$  is the number of levels in the wavelet decomposition and  $E$  is the log-energy of the image.

Hassen *et al.* [79] presented another wavelet transform-based image sharpness metric. The image sharpness is represented as the Local Phase Coherence (LPC), which is measured in the complex wavelet transform domain. The image being processed is first decomposed using the 3-level complex wavelet transform, then, LPC is used to measure the sharpness index ( $SI$ ), given by:

$$SI = \frac{\sum_{i=1}^N W_i P_{(i)}}{\sum_{i=1}^N W_i} \quad (3.42)$$

Here,  $P_{(i)}$  are the LPC values and  $W_i$  are the weights assigned to the LPC values.

Choi *et al.* [80] developed an algorithm for IQA based on the blur and noise in the image. The distorted image is first analyzed to evaluate following parameters: the blur mean, blur ratio, noise mean, and the noise ratio. These parameters are modeled using regression technique to approximate a quality score, represented by:

$$QM = 1 - (w_1 Blur_{mean} + w_2 Blur_{ratio} + w_3 Noise_{mean} + w_4 Noise_{ratio}) \quad (3.43)$$

Here,  $w_i$  is the parameters determined using simple regression.

Narvekar and Karam [81][82] proposed a probabilistic model-based image blur metric. The test image is computed to obtain an edge map, and the probabilistic model estimates

the probability of blur detection at every edge pixel. The overall sharpness metric is obtained by computing the Cumulative Probability of Blur Detection (CPBD), given by:

$$CPBD = P(P_{blur} \leq P_{JNB}) = \sum_{P_{blur}=0}^{P_{blur}=P_{JNB}} P(P_{blur}) \quad (3.44)$$

Here,  $P_{JNB}$  is the probability of the Just Noticeable Blur (JNB), and  $P_{blur}$  is the probability of blur detection given by:

$$P_{blur} = P_{blur}(e_i) = 1 - \exp\left(-\left|\frac{w(e_i)}{w_{JNB}(e_i)}\right|\right)^\beta \quad (3.45)$$

In the above equation,  $w(e_i)$  is the measured width of edge  $e_i$  and  $w_{JNB}(e_i)$  is the JNB width depending on the local contrast in neighborhood of edge  $e_i$ . At JNB,  $w(e_i) = w_{JNB}(e_i)$ , which corresponds to  $P_{blur} = 63\% = P_{JNB}$ . It implies that blur is not detected at an edge when  $P_{blur} \leq P_{JNB}$ . Thus, CPBD represents the percentage of edges at which probability of blur detection is below JNB. The lower metric value represents a blurred image.

### 3.3.4 Non-Distortion-Specific Quality Measures

Most of the existing NR-IQA measures are distortion specific. The models are provided with the prior knowledge of the type of distortion in an image. There are few algorithms which don't require the knowledge of the distortion type and operates on the images directly. These are briefly discussed below.

Chetouani *et al.* [83] proposed a novel NR image quality metric using neural networks. This approach aimed at estimating the most annoying distortions such as, blocking, ringing, and blurring effect observed in an image. The first stage of this framework is to quantify the following artifacts: blocking effect [84], blurring effect [85], and ringing effect [64]. These extracted features are provided as an input to the artificial neural network to estimate these distortions. The output on a trained network is a single value corresponding to the quality level of the test image.

Moorthy *et al.* [86][87] developed a framework, which doesn't require the knowledge of distortion type for quality assessment. The algorithm is based on the NNS [88] of an image. The framework is a combination of two systems: a Support Vector Machine (SVM) based classifier [87] and a quality estimator. The classifier analyzes the test image, and gauges the probability of being each distortion from predefined distortions set. The quality estimator estimates the quality index for every distortion within the distortion set. Finally, a probability-weighted summation of measured probabilities and the quality indexes are pooled to give an overall image quality score. The framework incorporates the following distortions: JPEG compressions, JPEG2000 compression, Gaussian blur white noise and fast fading. A Support Vector Regression (SVR) approach was used in [89] to estimate the quality index of distorted images. The model didn't perform well for JPEG compressed images, thus, the algorithm proposed in [51] is employed for JPEG compression. The image is first transformed using Daubechies 9/7 wavelet basis over three scales and three orientations. Then each sub-band coefficients are parameterized using Generalized Gaussian Distribution (GGD). The two parameters of the distribution; namely, the variance ( $\sigma^2$ ) and the shape parameter ( $\gamma$ ) are extracted



at 3 scales and 3 orientations in each scale. This makes the feature set an 18-D vector ( $\vec{f}_i$ ). The predicted quality score is given by:

$$BIQI = \sum_{i=1}^5 p_i \cdot q_i \quad (3.46)$$

Here,  $p_i$  is the probability of each distortion and  $q_i$  is the quality scores measured by all the five models, each trained for a specific type of distortion.

Saad *et al.* [90][87] developed a probabilistic prediction model. Based on the statistics of the DCT coefficients, the model extract certain distinct features from the distorted image to form the feature vector. The feature vector comprises of following constraints: kurtosis of DCT coefficients histogram, DCT based contrast, and anisotropy in DCT domain. Then, a probabilistic prediction model, following the multivariate Gaussian distribution and the multivariate Laplacian distribution, used to predict quality. The performance of the model is tested on the following distortions: white noise, Gaussian blur, fast fading channel distortions, and JPEG and JPEG2000 compression. The probabilistic model is given by:

$$P(X_i, DMOS_i) = P(DMOS_i / X_i) P(X_i) \quad (3.47)$$

In the above equation,  $X_i$  represents the feature set extracted from the image  $i$  and  $DMOS_i$  is the subjective score.

### **3.4 Performance of Objective Quality Measures**

Validation is necessary for the successful development of the objective image quality measure. The main aim of any objective quality measure is to estimate the perceived visual quality by human observers. The standard approach of validation, is to compare the model output (objective quality score) with the subjective quality scores from the human observers. This section discusses the most commonly used validation approaches for subjective quality evaluation, and also the evaluation criteria recommended by VQEG [6] for objective quality evaluation.

#### **3.4.1 Subjective Quality Evaluation**

Subjective quality evaluation is determined by complicated experiments. It involves many aspects of human psychology and viewing conditions, such as vision ability of the observer, translation of the perceptual quality into a score, stimulus content, surrounding light, display devices, etc. The two widely used approaches for subjective quality evaluation are: single-stimulus method (single-stimulus continuous quality evaluation) and double-stimulus method (double-stimulus continuous quality scale). These methods have been standardized by the International Telecommunications Union.

In single-stimulus method, the observers express their impression of quality on a linear quality scale, divided into equally spaced sections. These segments are labeled as “Bad”, “Poor”, “Fair”, “Good”, and “Excellent”. When the image is exposed to the observers, they move the slider and place it at a point that best reflects the subject’s impression of

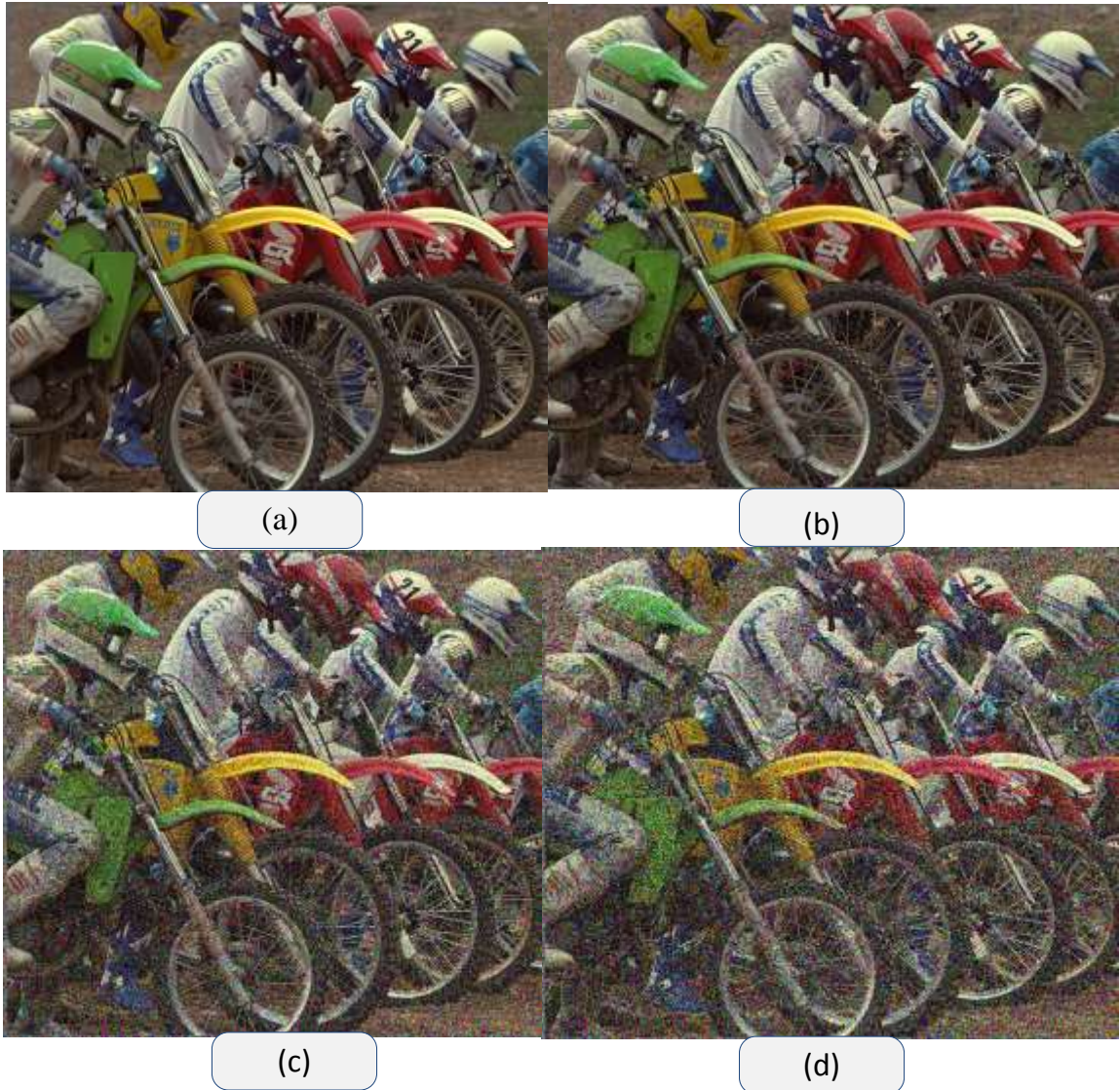
quality. The position of the slider is later converted to a numerical score known as Mean Opinion Score (MOS).

The double-stimulus method is a discrimination-based method, in which the subjects are exposed to both distortion-free and distorted image, one after the other within a small time gap of few seconds. Subjects evaluate the quality of both the images by moving the slider as discussed in the single-stimulus approach. The difference between the scores of the reference and the distorted image gives the subjective distortion score known as Difference Mean Opinion Score (DMOS).

Sheikh *et al.* [91] conducted an intense subjective image quality assessment study and developed an image database, widely known as LIVE database. This database has commonly been preferred as a benchmark in the literature to evaluate the performance of image quality measures.

The LIVE image database has 779 distorted images, which are evaluated by more than 12 observers in a row. The subjective scores are in terms of DMOS that are obtained from about 25000 individual human quality judgments. There are 29 color reference images (768 × 512 pixels size), with diverse image contents including human faces, natural sceneries, animals, monuments, statues, flowers, etc. The reference images are exposed to various distortions of different levels, to generate a group of distorted images. Some reference images are shown in Figure 3.5. The images were subjected to following distortions: JPEG compression, JPEG2000 compressions, Gaussian blur, and white noise. Each of these distortions is briefly discussed below:

- **White noise:** An additive white Gaussian noise is added to the RGB components of an image with a standard deviation ( $\sigma_N$ ) ranging from 0.012 to 2.0 as shown in Figure 3.1. This gives a total of 145 distorted images.



**Figure 3.1: White noise contaminated images**

(a) Reference image; (b), (c), (d) noisy images with varied Gaussian noise levels

- **Gaussian blur:** Every image is passed through a filter having a circular-symmetric 2D Gaussian kernel, with a standard deviation ( $\sigma_B$ ) ranging from 0.42 to 15 pixels as shown in Figure 3.2. This gives a total of 145 distorted images.

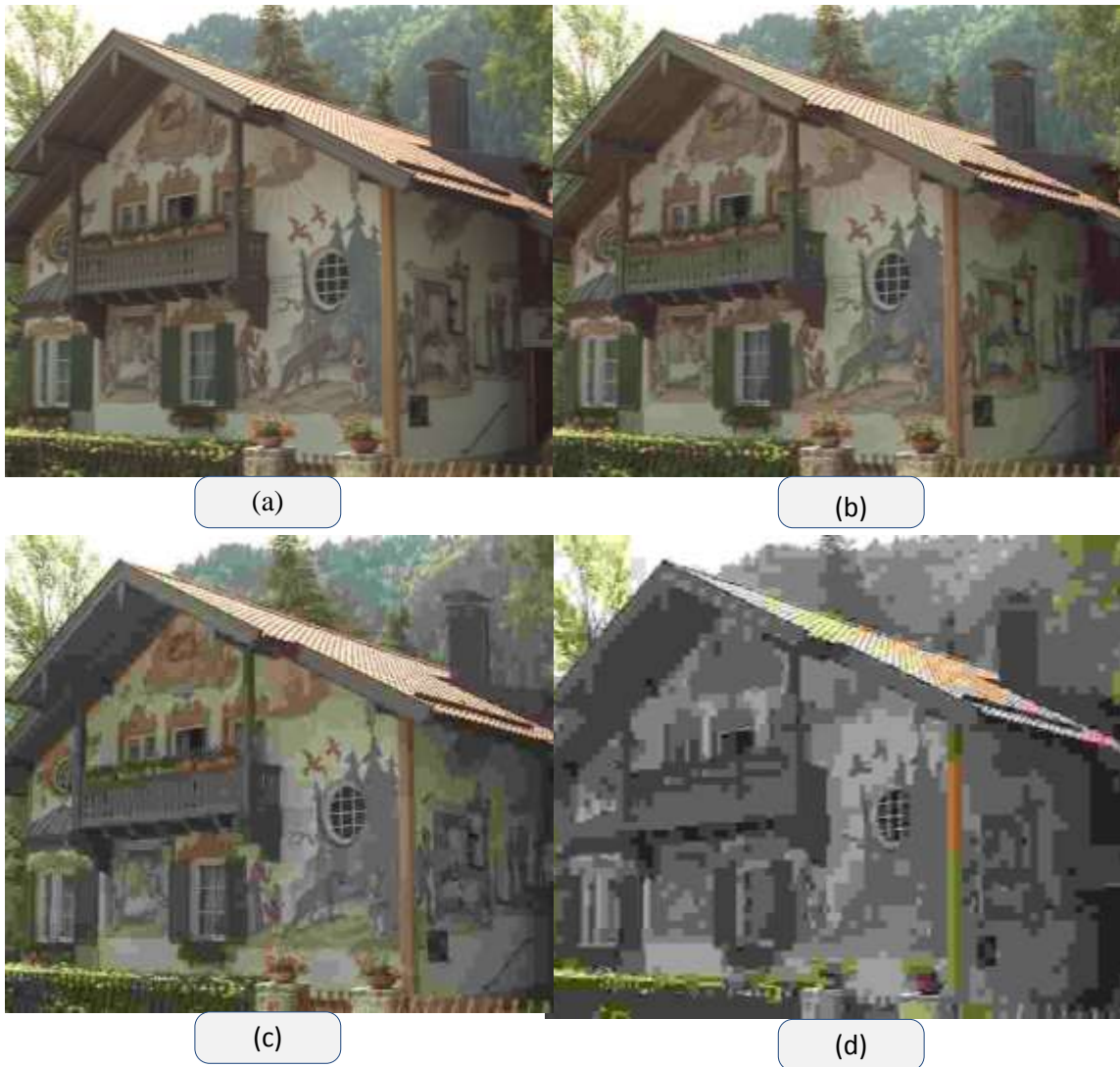


**Figure 3.2: Blurred images**

(b) Reference image; (b), (c), (d) blurred images with varied blur levels



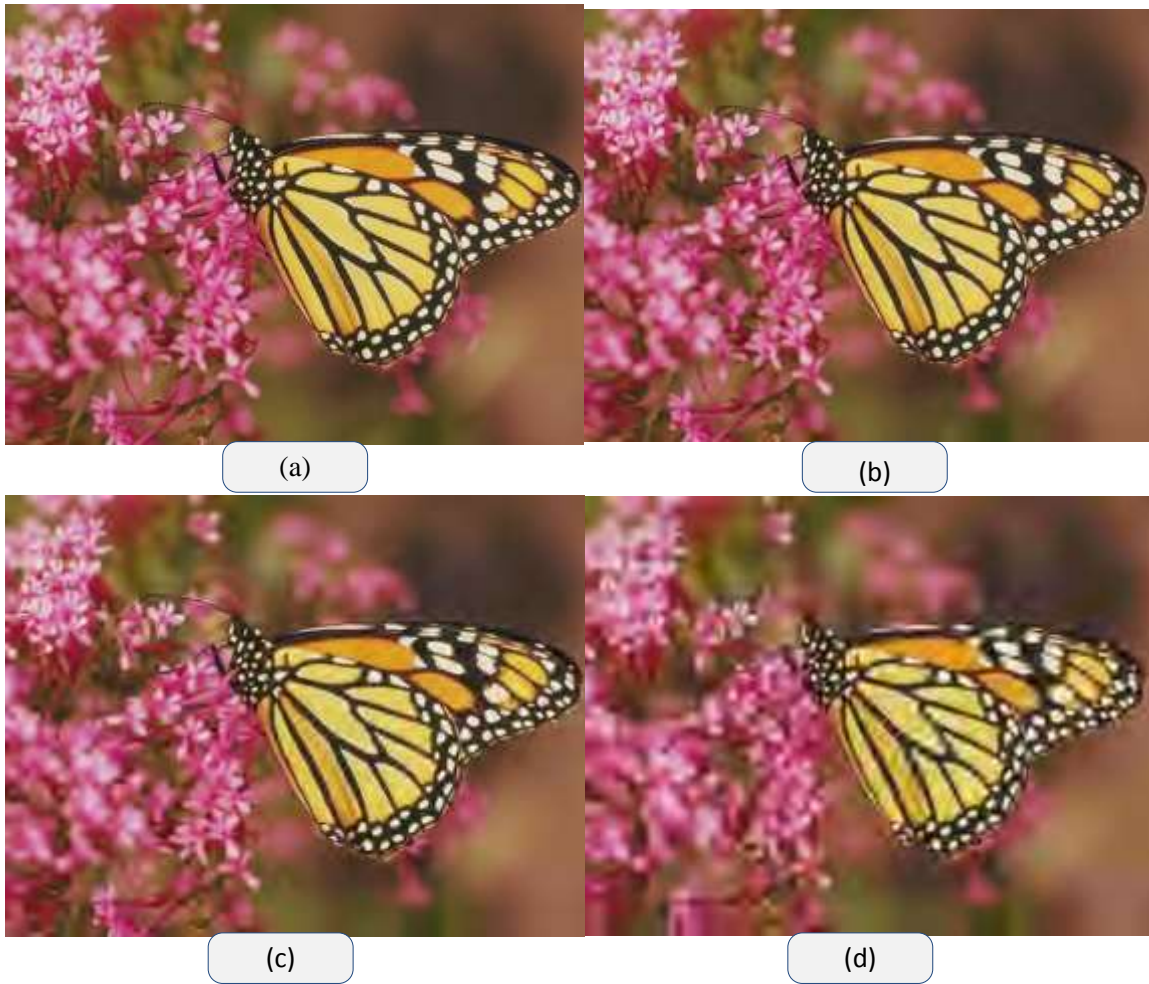
- **JPEG compression:** The images are exposed to JPEG compression technique at bit rates ranging from 0.15 bits per pixel (bpp) to 3.34 bpp as shown in Figure 3.3. This gives a total of 175 distorted images.



**Figure 3.3: JPEG compressed images**

(a) Reference image; (b), (c), (d) JPEG compressed images coded with varied bit rates

- **JPEG2000 compression:** The images are exposed to JPEG2000 compression technique at bit rates ranging from 0.028 bpp to 3.150 bpp as shown in Figure 3.4. This gives a total of 169 distorted images.



**Figure 3.4: JPEG2000 compressed images**

(c) Reference image; (b), (c), (d) JPEG2000 compressed images coded with varied bit rates

Further details about the LIVE image database can be obtained from [92].



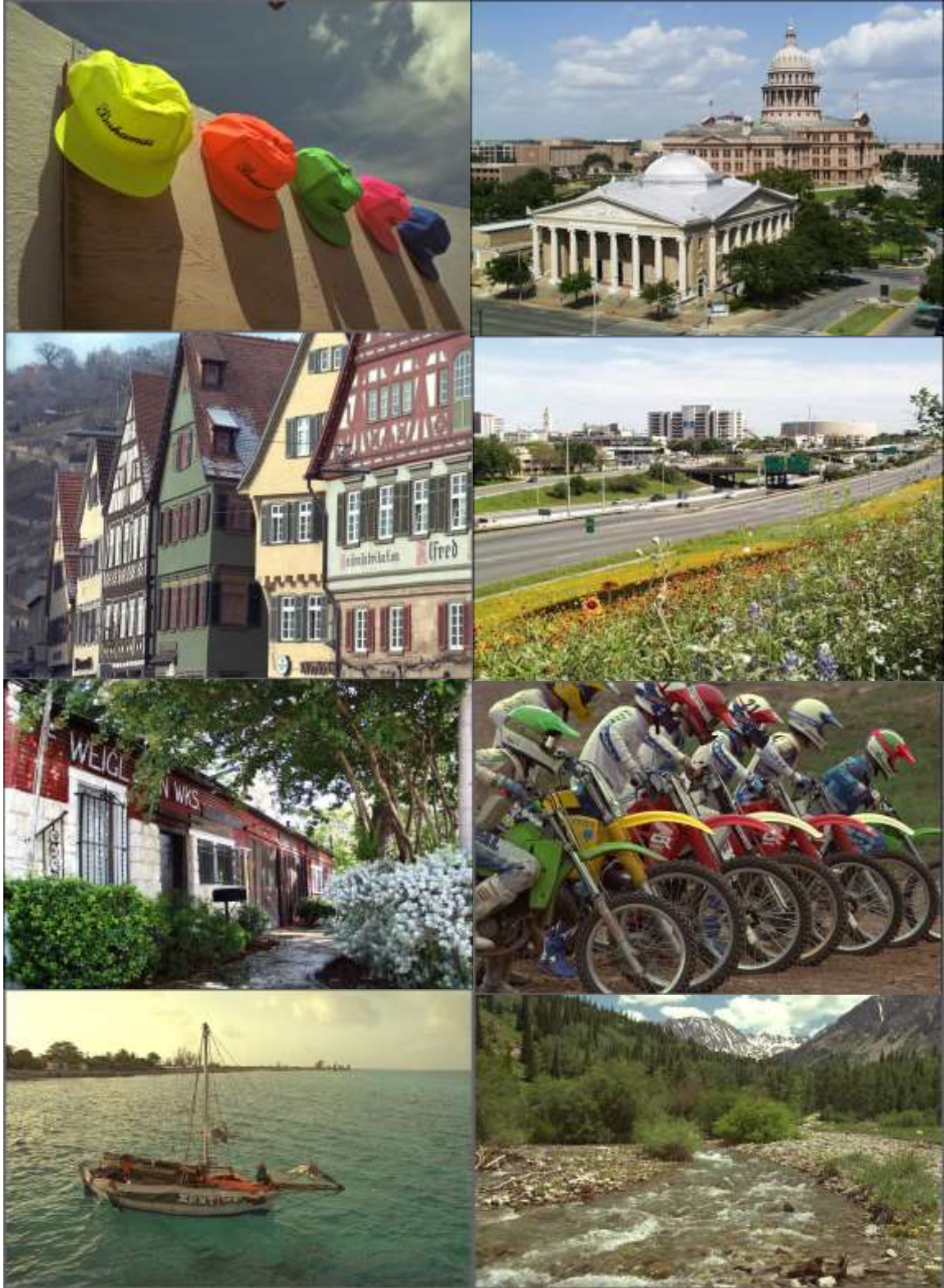


Figure 3.5: Reference images used in LIVE database



### 3.4.2 Performance Evaluation Criteria

According to the VQEG group formed in 1997, the performance of objective quality measures can be evaluated using the following evaluation criteria [6]:

- a. *Pearson's Correlation Coefficient (PCC)* is employed to evaluate the accuracy of the prediction score on a scale of -1 to 1:

$$PCC = \frac{\sum_{i=1}^N (DMOS(i) - \overline{DMOS})(DMOS_p(i) - \overline{DMOS_p})}{\sqrt{\sum_{i=1}^N (DMOS(i) - \overline{DMOS})^2} \sqrt{\sum_{i=1}^N (DMOS_p(i) - \overline{DMOS_p})^2}} \quad (3.48)$$

In the above equation,  $i$  denotes the index of image/video samples  $DMOS(i)$  is the subjective quality score,  $DMOS_p(i)$  is the predicted subjective quality score.  $\overline{DMOS}$  and  $\overline{DMOS_p}$  are the average of  $DMOS(i)$  and  $DMOS_p(i)$  respectively, and  $N$  denotes the total number of images/video samples.

- b. *Spearman's Rank-Order Correlation Coefficient (SROCC)* is used as a measure for the prediction monotonicity criterion. To compute SROCC,  $DMOS$  and  $DMOS_p$  are represented into the ranks, and the SROCC is calculated with the data of ranks using the equation (3.48). The computation of SROCC can be independent of the compensation by a mapping function, as it is calculated based on the ranks of quality data.
- c. *Outlier Ratio (OR)* is a measure to evaluate the consistency of the prediction, and is defined as the ratio of outlier points to the total points given by:

$$OR = \frac{\text{total number of outliers}}{N} \quad (3.49)$$

The point is said to be an outlier, if the prediction error in equation (3.50) crosses the threshold. Normally, the value of the threshold is twice the standard deviation of individual subjective quality scores assigned to an image/video sample. The prediction error and the condition for the outlier are given by:

$$P_{error}(i) = DMOS(i) - DMOS_p(i) \quad (3.50)$$

$$|P_{error}(i)| > 2\sigma_{DMOS}(i) \quad (3.51)$$

- d. *Root Mean Square Error* (RMSE) also provides an informative performance measure, to judge the relative performance between the quality measures under comparison. Based on the prediction error in the equation (3.50), the RMSE is measured using:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{error}(i))^2} \quad (3.52)$$

Here, N is the total number of image/video samples.

- e. *Mean Absolute Error* (MAE) is a measure, which specifies how close the predictions are to the ultimate outcomes. It is defined as the average of the absolute errors given by:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_{error}(i)| \quad (3.53)$$

- f. *Coefficient of determination* ( $R^2$ ) is a measure of accuracy of the regression models. It indicates how well the regression line fits the data set. Its value lies between 0 and 1. If the regression line fits the data well, the  $R^2$  is close to 1; otherwise, it is close to 0 for poor fitting. The value of  $R^2$  is given by:

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} = \frac{\sum_i (DMOS(i) - DMOS_p(i))^2}{\sum_i (DMOS(i) - \overline{DMOS})^2} \quad (3.54)$$

Here,  $SS_{err}$  is the sum of square of residual and  $SS_{tot}$  is the total sum of squares.

### 3.5 Summary

In this chapter, several image quality measures developed for quality assessment, under three scenarios were discussed. The vast majority of the image quality algorithms are based on FR quality assessment. The conventional FR image quality measures like MSE and PSNR, gained popularity because of their simple mathematical formulation. The traditional FR image quality algorithms are based on a variety of approaches, such as quality estimation based on HVS, based on image structure, etc. The RR image quality measures were also discussed; these are preferred in cases where the reference image is not fully accessible. Several NR-IQMs were also discussed; these are preferred over FR and RR quality measures in real-time applications. We have shown that most NR algorithms are developed for specific types of distortions like blocking, blurring, noise,

etc. for quality prediction. Some of these models are specifically designed for compression artifacts. The performance of most image quality measures is evaluated using images from LIVE database, which contained various distortions each with different levels. Finally, the five evaluation criteria recommended by the VQEG for validation of the objective measures are discussed at the end.

## **CHAPTER 4**

### **THE PROPOSED NO-REFERENCE IMAGE QUALITY**

### **ASSESSMENT ALGORITHM USING ARTIFICIAL**

### **NEURAL NETWORKS**

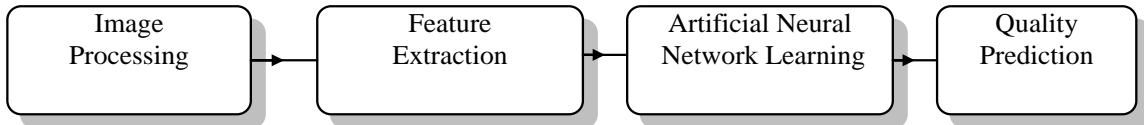
#### **4.1 Introduction**

In this chapter, the neural network-based no-reference image quality assessment algorithm is proposed, which attempts to estimate the quality through machine learning. In literature survey, it is found that most of the models developed for NR-IQA are distortion specific. For the known distortion, some distinct parameters from the distorted medium are extracted, and then modeled using curve fitting technique for the quality prediction. In this research, we aimed to develop a general-purpose NR-IQA learning algorithm that is not limited to some specific distortion. Thus, the framework of neural network-based pattern classification is utilized to develop such learning model. The ANN is used as a regressor in this approach for quality estimation. The idea is to extract the significant features from the distorted image, and train the neural networks for quality prediction.

In brief, the proposed algorithm involves two steps:

- Feature Extraction.
- ANN Learning.

The block diagram of the proposed model is given in Figure 4.1



**Figure 4.1: Block diagram of the proposed algorithm**

This chapter is organized as follows: few important concepts of the ANNs related to the proposed algorithm is discussed in section 4.2, and the proposed framework is discussed in detail in section 4.3.

## **4.2 Artificial Neural Networks**

In present era, computers play an important role in handling complex computational tasks. They are highly fast and execute sequence of instructions, designed for a specific task [93], [94]. On the other hand, humans can perform more complex tasks like object and speech recognition, more precisely and efficiently, better than any other manmade model. Various efforts have been made by the scientists and the engineers, to understand the functioning of the human brain, and mathematically model it. These models attempt to simulate certain properties of the HVS, and thus, known as Artificial Neural Networks (ANN). They often handle complex tasks involving large amount of experimental data

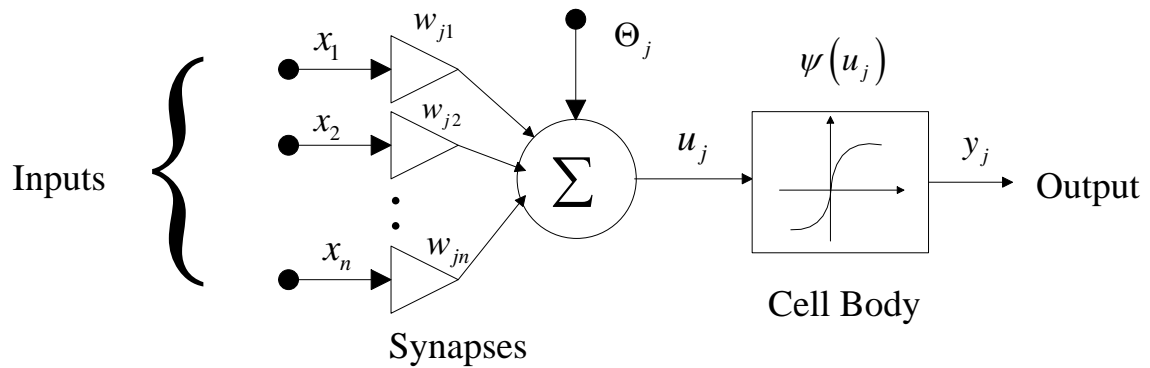
and observations. The ANN is an adaptive system, which is first trained to learn the task at hand, and then trained system is used to handle similar tasks. Such trained systems are able to execute similar tasks, not necessary the same. The ANN is widely used in various applications like, pattern and speech recognition, classification, image processing, etc. In the proposed approach, the artificial neural network is employed as a class of mathematical algorithms for the quality estimation.

The mathematical model of a biological neuron proposed by McCulloch and Pitts [95] became predominant. It inspired the research on parallel distributed processing systems, commonly known as artificial neural systems [96–98]. The researchers attempted to model the human abilities like speech recognition, classification, etc., using this artificial neural system. Various practical problems like pattern recognition [99], modeling [100], and prediction [101] can be solved using ANN. The ANN is an interconnection of the processing units between the input and the output layers. These processing units are known as “cells or neurons”.

#### **4.2.1 Artificial Neuron**

In ANN, the neuron is a processing unit, which processes the data or information it receives. The processed output from the neuron in the previous layer is given as in input to the next layer or represents the network’s output. The neural network comprises of three layers: input layer, hidden layer and output layer. The neurons of the input layer accept data provided by users and forward it to the hidden layer. The neurons in the hidden layer communicate only with neurons within the system, and propagate the

processed data to the output layer. Finally, the output layer neurons provide the final output of the system. The design structure of the network allows the neurons to perform many computational tasks concurrently. The basic model of the artificial neuron is given below:



**Figure 4.2: Artificial Neuron Model**

The neuron model comprises of four elementary modules: input vector ( $x_1$  to  $x_n$ ), summing junction ( $\Sigma$ ), threshold ( $\Theta$ ), and activation function (cell body), as shown in

Figure 4.2.

- Input vector is the experimental data provided to the network by the outside source. It is a vector  $X$  of size  $n \times 1$ , where  $X$  belongs to  $\mathfrak{R}$ . Every element of the vector  $x_i$ ,  $i = 1 \dots n$  is connected to the  $j^{\text{th}}$  neuron through the synaptic weights  $w_{ji}$



- The summing junction ( $\Sigma$ ) takes the weighted input product, which is the product of input signals and the synaptic weights, and summed to give a net input  $net_j$ .
- The activation function  $f(.)$  processes this net input and produces the final output of the neuron  $y_j$ . Generally, activation function can be linear or non-linear function depending on the characteristics of observed data.
- The threshold/ bias ( $\Theta$ ) term is added to control the cumulative input to the activation function. It takes a value either '-1' or '+1' given to the summing unit.

The mathematical representation of a simple neuron is given by:

$$net_j(X) = \sum_{i=1}^n w_{ji}x_i + \theta_j \quad (4.1)$$

Here,  $\theta$  is the bias and  $w_{ji}$  is the synaptic weight.

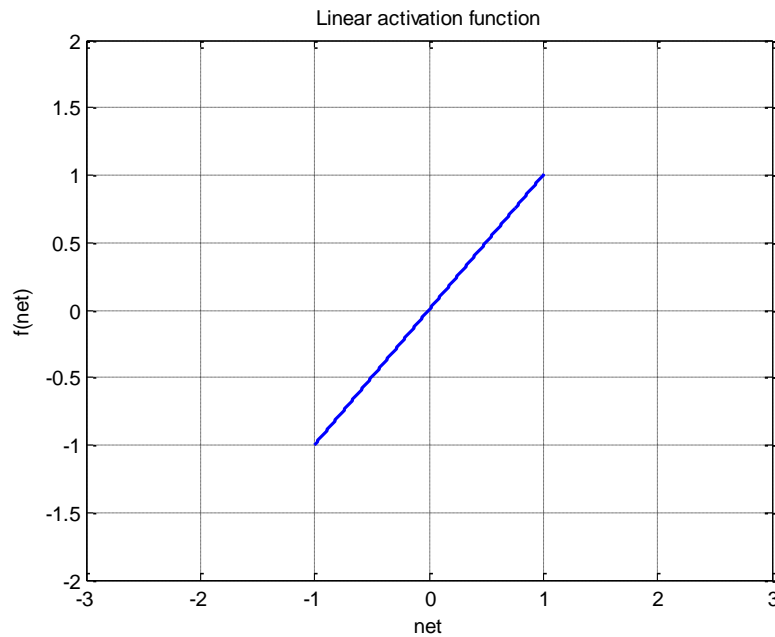
#### 4.2.2 Activation Function

There are various types of activation functions available and the selection of a proper function depends on the problem. A network with proper activation function produces efficient results. The activation functions used in our approach are discussed below:

- *Linear function*

It is a continuous valued function shown in Figure 4.3. It is like an identity function used in regression problems. The output of neuron  $j$  is given by:

$$y_j = f(\text{net}_j) = \text{net}_j \quad (4.2)$$



**Figure 4.3: Linear activation function**

- *Sigmoid function*

It is a non-linear activation function, used in a wide range of applications. The various characteristics of sigmoid function like non-linearity, monotonicity, differentiability, etc. makes it a suitable activation function for many non-linear problems. It has saturating limits at binary range (0 or 1).

The mathematical representation of binary sigmoid function for neuron  $j$  is given by:

$$y_j = f(\text{net}_j) = \frac{1}{1 + e^{-\alpha \text{net}_j}} \quad (4.3)$$

Here,  $\alpha$  is a parameter that controls the slope of the signal.

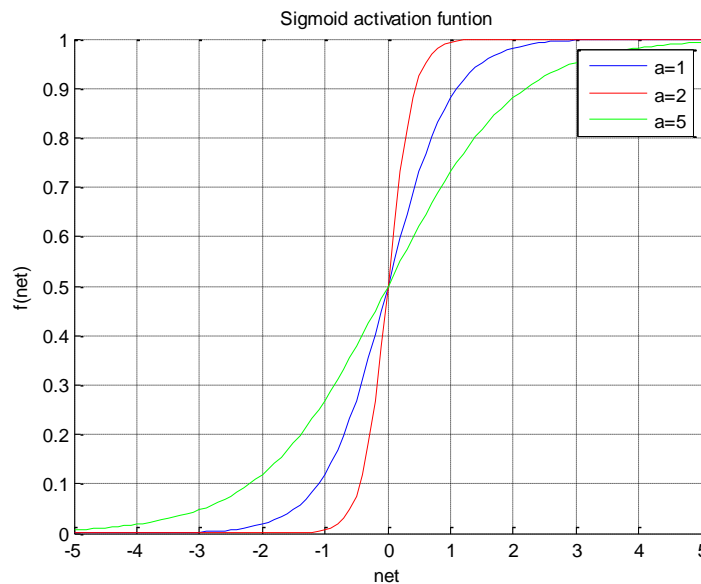
The graphical representation explaining the variation in shape of the sigmoidal signal with different values of  $\alpha$  is shown in Figure 4.4

- *Hyperbolic tangent sigmoid function*

This function exhibits similar properties like sigmoidal function, except with saturating limits at bipolar range (1 or -1). The mathematical representation of the output of  $j^{\text{th}}$  neuron is given by:

$$y_j = f(\text{net}_j) = \frac{1 - e^{-2\alpha \text{net}_j}}{1 + e^{-2\alpha \text{net}_j}} \quad (4.4)$$

Here,  $\alpha$  is a parameter controlling the slope of the signal.



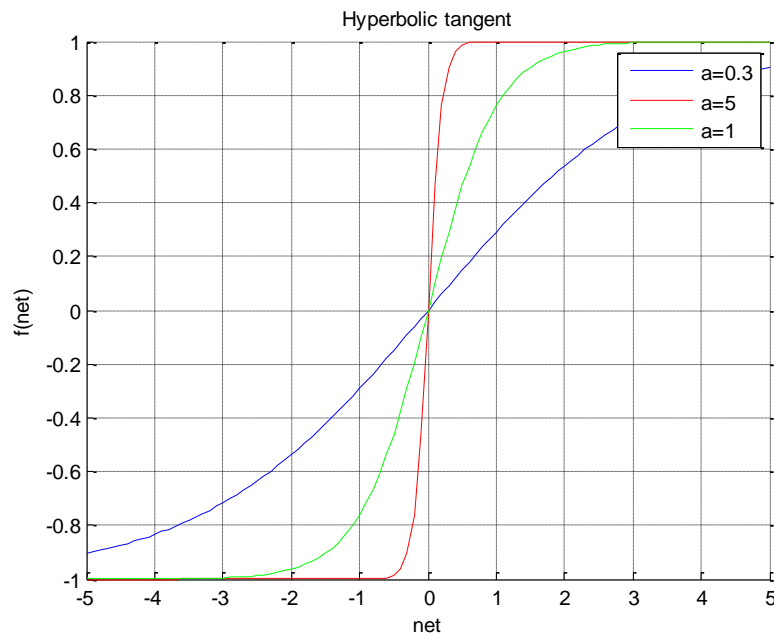
**Figure 4.4: Sigmoid Transfer Function**

Both functions, sigmoid and hyperbolic tangent functions are equivalent, and it is possible to transfer one function to the other with some linear transformation. The equations (4.5) and (4.6) represent the derivatives of both the functions with respect to net input. Differentiability is one of the important characteristics of both sigmoidal and hyperbolic tangent functions, used in the training algorithms.

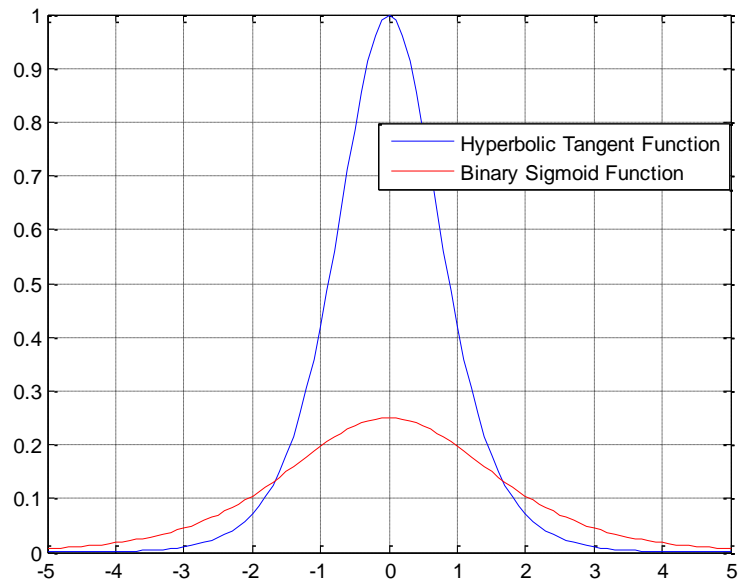
$$g_j(\text{net}_j) = \alpha f(\text{net}_j) [1 - f(\text{net}_j)] \quad (4.5)$$

$$g_j(\text{net}_j) = \alpha [1 + f(\text{net}_j)] [1 - f(\text{net}_j)] \quad (4.6)$$

The graphical representation of the hyperbolic tangent sigmoidal function is shown in Figure 4.5



**Figure 4.5: Hyperbolic Tangent Sigmoid Function**

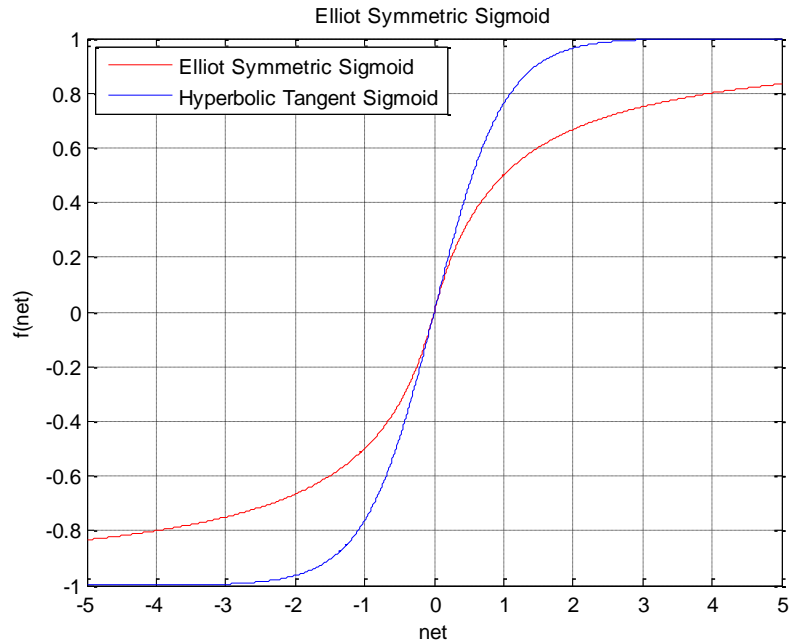


**Figure 4.6: Derivative of binary sigmoid and hyperbolic tangent function**

- *Elliot symmetric sigmoid function*

This function is similar to hyperbolic tangent sigmoid, except with no exponential term in the expression. It is faster than other sigmoid functions, as it doesn't include any exponential function. The mathematical representation of the output of  $j^{\text{th}}$  neuron is given by:

$$y_j = f(\text{net}_j) = \frac{\text{net}_j}{1 + |\text{net}_j|} \quad (4.7)$$



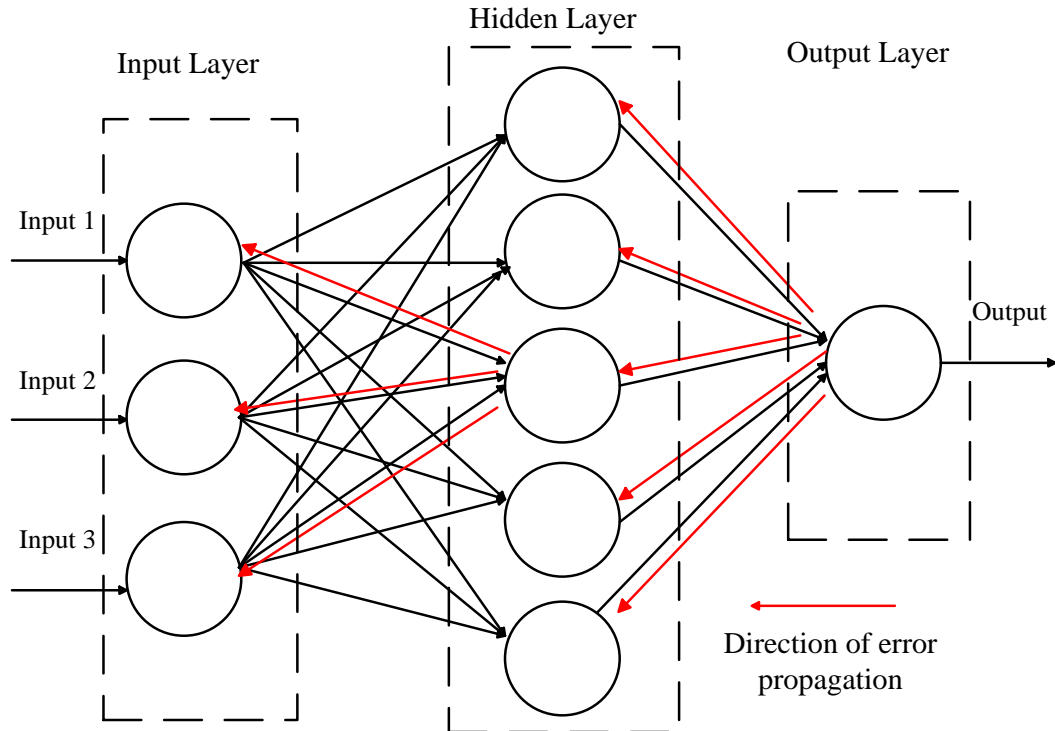
**Figure 4.7: Elliot Symmetric Sigmoid Function**

The graphical representation of the hyperbolic tangent sigmoidal function is shown in Figure 4.7

### 4.2.3 Back-Propagation Algorithm/ Delta Rule

The network's learning refers to the configuration of the synaptic weights between the consecutive nodes, to minimize the overall error of the network. This can be achieved by the backward propagation of the error (difference between target and obtained response) from the output to the previous layers as shown in Figure 4.8. Hence, this algorithm is termed as Back-Propagation (BP) algorithm/ Delta rule [98], [102], [103]. The aim of the learning algorithm is to tune the weights, and fit the hyper-plane to the input distribution

with respect to the target. The BP algorithm uses the error between the network's output  $o^p$  and the target  $t^p$  for the weights' tuning.



**Figure 4.8: Back-Propagation of network's error**

The overall error of the network is based on the summed squared error given by:

$$E = \sum_p E^p = \frac{1}{2} \sum_p (t^p - y^p)^2 \quad (4.8)$$

Here, index  $p$  represents the input patterns and  $E^p$  is the error for pattern  $p$ . The delta rule incorporates the concept of gradient descent, and search for the weights that minimizes the error function. According to gradient descent rule, “the weight change is proportional to the negative derivative of the error, measured on current training pattern, with respect to each weight” given by:

$$\Delta_p w_j = -\eta \frac{\partial E^p}{\partial w_j} \quad (4.9)$$

Here,  $\eta$  is a positive constant known as the *learning rate*.

The error derivative can be represented as:

$$\frac{\partial E^p}{\partial w_j} = \frac{\partial E^p}{\partial y^p} \frac{\partial y^p}{\partial w_j} \quad (4.10)$$

The output of a neuron is represented by a linear equation given by:

$$y_p = net_j(X) = \sum_{i=1}^n w_{ji} x_i + \theta_j \quad (4.11)$$

The derivative of the linear equation (4.11) given by:

$$\frac{\partial y^p}{\partial w_j} = x_j \quad (4.12)$$

And

$$\frac{\partial E^p}{\partial y^p} = -(t^p - y^p) \quad (4.13)$$

By substituting equations (4.12) and (4.13) in equation (4.9), we have

$$\Delta_p w_j = \eta \delta^p x_j \quad (4.14)$$

Here,  $\delta^p = t^p - y^p$  is the error between the neuron's output and the target for the pattern  $p$ .



This rule can be generalized for the neurons with non-linear activation function. Let's consider a function given by:

$$y_k^p = f(s_k^p) \quad (4.15)$$

Here,  $s_k^p$  is evaluated using (4.11). The delta rule is generalized using following steps:

$$\Delta_p w_{jk} = -\eta \frac{\partial E^p}{\partial w_{jk}} \quad (4.16)$$

The error measured  $E^p$  is given by:

$$E = \sum_p E^p = \frac{1}{2} \sum_{i=1}^K (t_i^p - y_i^p)^2 \quad (4.17)$$

The derivative of equation (4.17) is given by:

$$\frac{\partial E^p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}} \quad (4.18)$$

From equation (4.11), the second term in equation (4.18) is given by:

$$\frac{\partial s_k^p}{\partial w_{jk}} = y_j^p \quad (4.19)$$

Consider an equation

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} \quad (4.20)$$

We get an update rule, equivalent to the delta rule as defined in equation (4.14) . On the error surface, update rule performs the gradient descent by altering the weights using the rule given by:

$$\Delta_p w_{jk} = \eta \delta_k^p y_j^p \quad (4.21)$$

The  $\delta_k^p$  in equation (4.20) can be expressed as:

$$\delta_k^p = -\frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial s_k^p} \quad (4.22)$$

On observing the equation (4.15), we notice that

$$\frac{\partial y_k^p}{\partial s_k^p} = f'(s_k^p) \quad (4.23)$$

Let's assume that the neuron  $k$  is an output neuron, and the error between target and the measured output is given by:

$$\frac{\partial E^p}{\partial y_k^p} = -(t_k^p - y_k^p) \quad (4.24)$$

This is similar to the result obtained by standard delta rule. Substitute equations (4.23) and (4.24) in (4.22) to obtain the neuron's output given by:

$$\delta_k^p = (t_k^p - y_k^p) f'(s_k^p) \quad (4.25)$$

This representation works well for the output neurons.

Now, let's see the weight change of hidden neurons. The error in the output layer is represented as a function of the net inputs  $s_j$  from hidden to output layer,  $E^p = E^p(s_1^p, s_2^p, \dots, s_j^p, \dots)$ , and this chain rule can be used to represent:

$$\begin{aligned}
 \frac{\partial E^p}{\partial y_h^p} &= \sum_{o=1}^K \frac{\partial E^p}{\partial s_o^p} \frac{\partial s_o^p}{\partial y_h^p} \\
 &= \sum_{o=1}^K \frac{\partial E^p}{\partial s_o^p} \frac{\partial}{\partial y_h^p} \sum_{j=1}^N w_{ho} y_j^p \\
 &= \sum_{o=1}^K \frac{\partial E^p}{\partial s_o^p} w_{ho} \\
 &= \sum_{o=1}^K \delta_o^p w_{ho}
 \end{aligned} \tag{4.26}$$

Substitute equation (4.26) in equation (4.22), we get

$$\delta_h^p = f'(s_h^p) \sum_{o=1}^K \delta_o^p w_{ho} \tag{4.27}$$

All  $\delta$ 's of the network's neurons can be computed using a recursive method provided by equations (4.25) and (4.27). This can further be used to calculate the change in weight according to equation (4.21). Thus, this approach allows in developing the generalized delta rule for the non-linear networks, by error propagation from the output layer to the neurons in the hidden layer.

The speed of the network convergence depends on the optimum value of learning rate. The change in the gradient is directly proportional to the learning rate. The weight is updated by combing the current error gradient, with the one in the preceding training step.

The updated weight is given by:

$$\Delta_p w_{jk} = -\eta \frac{\partial E^p}{\partial w_{jk}} + \alpha \Delta_{old} w_{jk} \quad (4.28)$$

In the above equation,  $\Delta_{old} w_{jk}$  represent the latest weight change.

### 4.3 The Proposed Methodology

In this research, we focused on developing a learning-based NR-IQA algorithm, which is not restricted to only specific type of distortion. Thus, the ANN is employed as a learning-machine in our approach for quality estimation. The idea is to extract certain significant features from the distorted image, and use them to train neural networks for quality prediction. Hence, the proposed algorithm is a combination of two steps: feature extraction and ANN training.

#### 4.3.1 Feature Extraction

The selection of the appropriate features depends on the artifacts observed in the distortions. The proposed algorithm is tested on for the distortions: JPEG compression, JPEG2000 compression, blur and noise. The most common artifacts observed in these distortions are *blocking*, *blurring*, and *noise*. Various features quantifying these artifacts, and also their extraction details are given in this section.

The following are the features extracted for the quality assessment:

1. *Blocking Factor* [51]:

The blocking factor quantifies the blockiness observed in the compressed images. The blocking artifact refers to the noticeable artificial discontinuities, observed at regular interval in the DCT-based JPEG compressed images. Since the image is represented into 8x8 pixel blocks, the independent quantization of each coding block results in the appearance of the horizontal and vertical discontinuities at the block boundaries as shown in Figure 4.9. The overall blockiness is measured as the average differences across block boundaries given by:

$$B_h = \frac{1}{M \times \left( \left\lfloor \frac{N}{8} \right\rfloor - 1 \right)} \left( \sum_{i=1}^M \sum_{j=1}^{\left\lfloor \frac{N}{8} \right\rfloor - 1} |d_h(i, 8 \times j)| \right) \quad (4.29)$$

Here,  $B_h$  is the blocking factor around horizontal direction, and  $d_h(i, 8 \times j)$  is the horizontal gradient across the block boundary of an image  $x(i, j)$  given by:

$$d_h(i, 8 \times j) = x(i, 8j + 1) - x(i, 8j), j \in [1, N - 1] \quad (4.30)$$

The activity of an image is the measure of blurriness. It is defined as the “measure of existence of fine surfaces in the picture” [104]. The image with high spectral activity exhibits many fine structures, and hence, considered to be rich in detail. On the other hand, the unstructured monochrome image exhibits zero spatial activity. The image with poor spatial activity experiences blurring artifact as shown in Figure 4.10. The activity of

an image is measured using two factors: the average absolute difference and the zero-crossing rate.



**Figure 4.9: Blocking artifact**

2. *Average absolute difference* [51]:

Unlike blocking factor, the gradient is measured across the entire image to obtain horizontal and vertical details of the image. The average absolute difference between in-block image samples is given by:

$$A_h = \frac{1}{7} \left[ \frac{8}{M \times (N-1)} \left( \sum_{i=1}^M \sum_{j=1}^{N-1} |d_h(i, j)| - B_h \right) \right] \quad (4.31)$$

Here,  $A_h$  is the average absolute difference along horizontal direction, and  $d_h(i, j)$  is the horizontal gradient given by:

$$d_h(i, j) = x(i, j+1) - x(i, j), j \in [1, N-1] \quad (4.32)$$

3. *Zero-Crossing rate* [51]:

The Zero-Crossing (ZC) rate is defined as “the rate at which the signal changes from positive to negative or back” [105]. The ZC is said to occur, if the consecutive pixels have different signs. The image with high spatial activity exhibits high ZC rate, given by:

$$z_h(i, j) = \begin{cases} 1 & \text{horizontal ZC at } d_h(i, j) \\ 0 & \text{otherwise} \end{cases} \quad (4.33)$$

$$Z_h = \frac{1}{M \times (N-2)} \left( \sum_{i=1}^M \sum_{j=1}^{N-2} z_h(i, j) \right) \quad (4.34)$$

Here,  $z_h$  is the map of the horizontal ZC, and  $Z_h$  is the overall ZC rate of the image along horizontal direction.



Figure 4.10: Blurring artifact due to poor spatial activity



4. *Standard deviation of difference image:*

The standard deviation is a measure of variation or deviation from the expected value.

The smaller the standard deviation, the closer the data point to its mean.

The standard deviation of the difference image  $d_h$  along the horizontal direction is given by:

$$\sigma_{d-h} = \sqrt{\frac{1}{M \times (N-1)} \left( \sum_{i=1}^M \sum_{j=1}^{N-1} |d_h(i, j) - \bar{d}_h| \right)^2} \quad (4.35)$$

Here,  $\bar{d}_h$  is the mean of the difference image.

5. *Standard deviation of zero-crossing rate:*

The standard deviation of the zero-crossing rate along horizontal direction is given by:

$$\sigma_{z-h} = \sqrt{\frac{1}{M \times (N-2)} \left( \sum_{i=1}^M \sum_{j=1}^{N-2} z_h(i, j) - Z_h \right)^2} \quad (4.36)$$

Similarly, these measures along vertical direction are given by  $B_v, A_v, Z_v, \sigma_{z-v}, \sigma_{d-v}$

The mean of horizontal and vertical measures gives the overall value of the features, given by:

$$B = \frac{B_h + B_v}{2} \quad (4.37)$$

$$A = \frac{A_h + A_v}{2} \quad (4.38)$$

$$Z = \frac{Z_h + Z_v}{2} \quad (4.39)$$

$$\sigma_d = \frac{\sigma_{d-v} + \sigma_{d-h}}{2} \quad (4.40)$$

$$\sigma_z = \frac{\sigma_{z-v} + \sigma_{z-h}}{2} \quad (4.41)$$

#### 6. Noise Mean [80]:

Image noise is defined as the “random variation in the brightness or color information of the image” [106]. Edge detection becomes difficult by the presence of noise. Thus, it is important to filter out the noise before edge detection. This is done using the average filter, and the filtered image is given by:

$$g(x, y) = \frac{1}{3 \times 3} \left[ \sum_{i=-1}^1 \sum_{j=-1}^1 f(x+i, y+j) \right] \quad (4.42)$$

The candidates for the noisy pixels from the filtered image in equation (4.42) are given by:

$$D_h(x, y) = |g(x, y+1) - g(x, y-1)| \quad (4.43)$$

$$D_{h-mean} = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N D_h(x, y) \quad (4.44)$$

Here,  $D_{h-mean}$  is the mean of noise candidates along horizontal direction  $D_h(x, y)$ .

Similarly,  $D_v(x, y)$  is the noise candidates along vertical direction and  $D_{v-mean}$  is its mean.

Then,

$$N_{cand}(x, y) = \begin{cases} \max(D_h(x, y), D_v(x, y)) & \text{if } D_h(x, y) \leq D_{h-mean} \text{ and } D_v(x, y) \leq D_{v-mean} \\ 0 & \text{otherwise} \end{cases} \quad (4.45)$$

$$N_{cand-mean}(x, y) = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N N_{cand}(x, y) \quad (4.46)$$

In the above equation,  $N_{cand}(x, y)$  is the noise candidate, which is zero along the edges,

and  $N_{cand-mean}(x, y)$  is its mean.

Finally, the noisy pixels detected using:

$$N(x, y) = \begin{cases} N_{cand}(x, y) & \text{if } N_{cand}(x, y) > N_{cand-mean} \\ 0 & \text{otherwise} \end{cases} \quad (4.47)$$

The noise mean is given by:

$$Noise_{mean} = \frac{Sum_{noise}}{Noise_{cnt}} \quad (4.48)$$

Here,  $Sum_{noise}$  is the sum of the pixels  $N(x, y)$  and  $Noise_{cnt}$  is the total number of noisy

pixels.

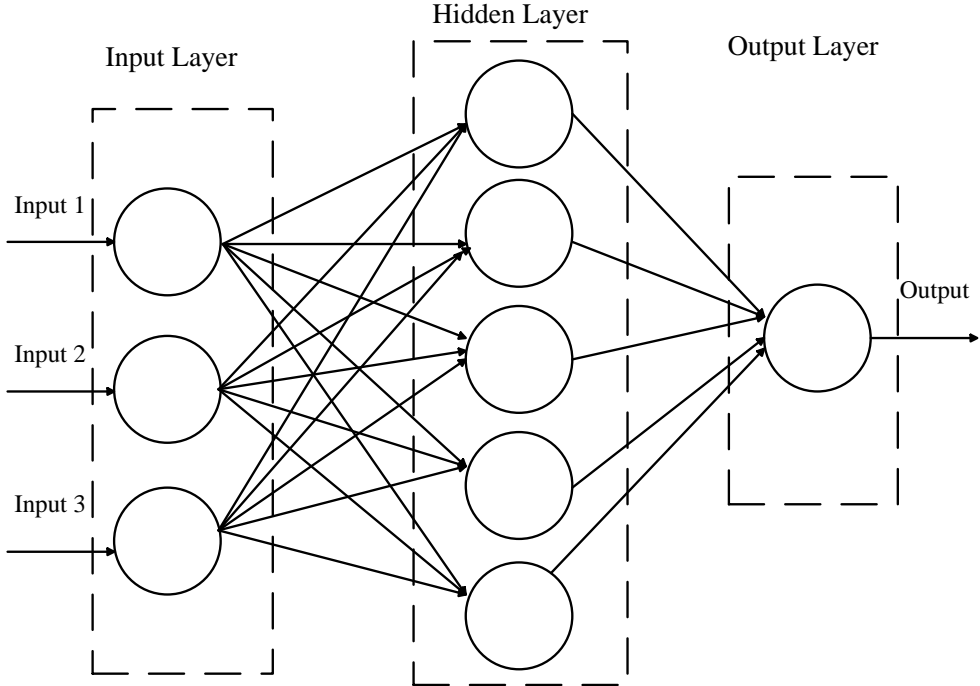


**Figure 4.11: Image contaminated with Gaussian noise**

### **4.3.2 Neural Network Training**

This section illustrates the training of the neural networks, using the statistical features extracted from the distorted images. The idea is to train the network with features from the distorted medium, against their subjective scores (MOS/DMOS), for quality

estimation. In our framework, the neural network with feedforward architecture is incorporated. This network structure allows propagation of data from input to output layer in forward direction, and doesn't facilitate the data flow within the same layer or to the preceding layer as shown in Figure 4.12. Three activation functions; namely, *sigmoid function*, *hyperbolic tangent sigmoid function*, and *Elliot symmetric sigmoid function* are used in hidden layer, and network's performance using each activation function is evaluated and compared. While, the linear activation function is employed in the output layer. The backpropagation algorithm/delta rule is adopted for network's training, discussed in the section 4.2.3.



**Figure 4.12: Feedforward Network**

The aim of this research is to develop the NR-IQA algorithm for the following five distortion cases:

1. Blur.
2. Noise.
3. JPEG compression.
4. JPEG2000 compression.
5. Across all distortions.

There are certain issues to be looked at while training the network, as follows:

*a. Optimum feature set:*

The selection of the proper feature set plays a vital role in neural network's learning. The statistical features must correlate well the perceived image quality, for better training of the network. To select AN optimum feature set for each of the distortion case, forward-selection rule is employed. "Forward-selection involves starting with no variables in the model, testing the addition of each variable, adding the variable that improves the model most, and repeating this process until none improves the model" [107]. Thus, this rule provides the optimum features for each distortion case.

*b. Optimum network size:*

The network size represents the number of neurons in the hidden layer, and the complexity of the network depends on the number of hidden layer nodes. Besides complexity, accuracy and generalization capabilities of the network depends upon its size too. Generally, the number of hidden layer nodes must be large enough to represent the problem correctly, but at the same time small enough to maintain the

network's generality. The effect of underfitting is observed in very small networks, wherein the network fails to identify the internal structure of the data. On the other hand, very large networks exhibit the overfitting effect, in which they are more likely to become over specific to the training data. Such over specific networks performs well on training set, but fails for newer data outside the training set. The under-specificity and the over-specificity of the networks lead to poor accuracy in quality prediction. Thus, identifying the optimum number of hidden neurons is an important issue in network design. Several researchers like Lippmann, Cybenko [108], [109] has reported that single hidden layer is sufficient to model most problems. In this approach, a suitable network size is obtained by training the network with different hidden neurons, and evaluating their performance based on the Mean Square Error (MSE). The network with minimum prediction error provides the optimum hidden neurons.

The learning of artificial neural network involves the following steps:

1. Three network topologies, each with different activation functions are selected. The activation functions include the sigmoid function, hyperbolic tangent and Elliot symmetric sigmoid function.
2. To speed up the training time, extracted features are first normalized to same range of values. This minimizes the bias within the neural network for one feature over another.
3. For all considered distortion cases, the optimum statistical features are selected using the forward selection rule.

4. Networks with varied number of hidden nodes are trained to find the optimum number of hidden neurons for each distortion case.
5. The network with optimum feature set and the optimum hidden neurons is trained for quality prediction. For network's training, it is provided with the selected feature set and the desired subjective scores (MOS/DMOS) from the training samples and is trained using BP algorithm, discussed in section 4.2.3.
6. Performance of the trained network is evaluated using the features extracted from the test samples. When the trained network is provided with the test features, it outputs the estimate of the quality score for the given feature set.
7. The predicted quality scores are compared with the actual subjective scores of the test samples, and the performance is evaluated based on the criteria, discussed in section 3.4.2.
8. The steps 3-7 are repeated for all three network topologies with different activation functions.

#### **4.4 Summary**

The framework of the proposed algorithm for the neural network-based NR-IQA is explained in this chapter. Certain important concepts of the ANN related to the proposed framework are discussed, which are incorporated in the neural network design. Both steps of the algorithm: feature extraction and ANN learning, are discussed in detail. The extraction of the six statistical features quantifying diverse artifacts is discussed under feature extraction, along with their mathematical representations. For ANN training, few



important issues that influence the performance the neural network are also reported. Finally, all steps involved in neural network training for quality predictions are enumerated.

## CHAPTER 5

### PERFORMANCE EVALUATION AND COMPARISON

#### 5.1 Introduction

In this chapter, the performance of the proposed algorithm is evaluated and compared with the traditional models. The database employed for this purpose is the standard LIVE image database [91][92], which has been widely used in performance evaluation of many image quality measures. It comprises of the following datasets: white noise contaminated images, Gaussian blurred images, JPEG and JPEG2000 compressed images, and also their corresponding subjective scores (DMOS). The information about the number of images in each dataset and their distortion parameters is given in Table 5.1. The DMOS value varies on a scale ranging from 0 to 100, where 0 represents an excellent and 100 represents a bad visual quality. More details of the LIVE image database are given in Section 3.4.1.

**Table 5.1: LIVE database**

Dataset	Number of images	Distortion Parameter	Parameter Range	Subjective Scores
White noise	145	Standard deviation of white Gaussian noise	0.012-2.00	Realigned DMOS (exclude the reference images)
Gaussian blur	145	Standard deviation of Gaussian filter	0.42-15.00	
JPEG	175	Bit rate (bits / pixel)	0.15-3.34	
JPEG2000	169	Bit rate (bits / pixel)	0.028-3.15	

## **5.2 Performance evaluation**

To evaluate the performance of the proposed approach, the following distortion sets are considered from the LIVE database: blur, noise, JPEG and JPEG2000 compression. Each distortion set is divided into two sets: a training set of size 60% and a testing set of size 40%. The objective is to develop the model for five distortion cases; namely, blur, noise, JPEG compression, JPEG2000 compression, and across all distortions. Thus, the six statistical features are extracted from the training samples using the procedure discussed in section 4.3.1, and are normalized to same range of values. These features form the basis for the ANN training.

It is to be noted that three network topologies, each with different activation function are selected. The activation functions employed are sigmoid function, hyperbolic tangent sigmoid function, and Elliot sigmoid function. For a better learning of the network, it should be provided with the optimum feature set, and also should have the optimum number of hidden neurons. The extraction of both parameters is discussed in the following sections.

### **5.2.1 Optimum feature set**

The optimum feature set for all distortion cases are obtained through forward selection rule. The networks with different combinations of feature sets are trained against the desired subjective scores, and the performance is then evaluated based on the MSE. The details of the optimum feature sets for the different distortions cases are shown in the

tables, Table 5.3 to Table 5.7. The results are shown for the considered five distortions cases.

**Table 5.2: Symbolic representation of the features**

Feature Number	Feature	Symbol
1	Blocking factor	$B$
2	Average Absolute Difference	$A$
3	Zero-Crossing rate	$ZC$
4	Standard deviation of diff. image	$\sigma_d$
5	Standard deviation of ZC	$\sigma_{zc}$
6	Noise Mean	$N$

In Table 5.2, the symbolic representation used for different statistical extracted features are shown.

**Table 5.3: Optimum features selection for JPEG compression**

JPEG Compression					
Sigmoid	MSE	Hyperbolic Tangent Sigmoid	MSE	Elliot Sigmoid	MSE
$\sigma_{zc}$	97.38	$\sigma_{zc}$	103	$\sigma_{zc}$	107
$\sigma_{zc} - B$	47.54	$\sigma_{zc} - B$	50.84	$\sigma_{zc} - B$	33.9
$\sigma_{zc} - B - A$	41.78	$\sigma_{zc} - B - A$	44.46	$\sigma_{zc} - B - A$	29.24
$\sigma_{zc} - B - A - ZC$	<b>37.13</b>	$\sigma_{zc} - B - A - ZC$	<b>39.90</b>	$\sigma_{zc} - B - A - ZC$	<b>25.17</b>
$\sigma_{zc} - B - A - ZC - \sigma_d$	40.08	$\sigma_{zc} - B - A - ZC - \sigma_d$	40.01	$\sigma_{zc} - B - A - ZC - \sigma_d$	26.12
$\sigma_{zc} - B - A - ZC - \sigma_d - N$	42.1	$\sigma_{zc} - B - A - ZC - \sigma_d - N$	44.85	$\sigma_{zc} - B - A - ZC - \sigma_d - N$	27.69

In Table 5.3, the performance of the networks in terms on the MSE for JPEG compression is shown. The results are shown for three networks each with the following activation function: sigmoid, hyperbolic tangent sigmoid and Elliot symmetric sigmoid. All three networks are trained with different combinations of feature set as per the forward selection rule. The results show that for all three activation functions, the following four features; namely, *standard deviation of ZC rate*, *blocking factor*, *average absolute difference*, and *ZC rate* improved the network's performance to its maximum. There is no further improvement with addition of other features. Thus, these four features are considered to be the optimum feature set for JPEG compressed distortion.

**Table 5.4: Optimum features selection for JPEG2000 compression**

JPEG2000 Compression					
Sigmoid	MSE	Hyperbolic Tangent Sigmoid	MSE	Elliot Sigmoid	MSE
$\sigma_{zc}$	106.2	$\sigma_{zc}$	106.9	$\sigma_{zc}$	110
$\sigma_{zc} - A$	65.36	$\sigma_{zc} - A$	68.01	$\sigma_{zc} - A$	50.04
$\sigma_{zc} - A - B$	44.87	$\sigma_{zc} - A - B$	43.9	$\sigma_{zc} - A - B$	42.9
$\sigma_{zc} - A - B - ZC$	<b>39.9</b>	$\sigma_{zc} - A - B - ZC$	<b>42.33</b>	$\sigma_{zc} - A - B - ZC$	<b>39.01</b>
$\sigma_{zc} - A - B - ZC - \sigma_d$	40.59	$\sigma_{zc} - A - B - ZC - N$	48.59	$\sigma_{zc} - A - B - ZC - N$	39.9
$\sigma_{zc} - A - B - ZC - \sigma_d - N$	41.8	$\sigma_{zc} - A - B - ZC - N - \sigma_d$	47.02	$\sigma_{zc} - A - B - ZC - N - \sigma_d$	40.21

The networks' performance for JPEG2000 compression is shown in Table 5.4. An optimum feature set for each network with different activation function is selected as per the forward selection rule. The following four features; namely, *standard deviation of ZC*

rate, average absolute difference, blocking factor, and ZC rate improved the network's performance for all three activation functions. There is no further improvement with addition of other features. Hence, these four features are considered to be an optimum feature set for JPEG2000 compressed distortion.

**Table 5.5: Optimum features selection for blur distortion**

Blur					
Sigmoid	MSE	Hyperbolic Tangent Sigmoid	MSE	Elliot Sigmoid	MSE
$\sigma_{zc}$	52.82	ZC	51.60	ZC	53.95
$\sigma_{zc} - A$	42.55	ZC - A	43.1	ZC - A	48.02
$\sigma_{zc} - A - \sigma_d$	30.51	ZC - A - $\sigma_d$	38.32	ZC - A - $\sigma_d$	39.79
$\sigma_{zc} - A - \sigma_d - B$	<b>27.63</b>	ZC - A - $\sigma_d - B$	<b>31.53</b>	ZC - A - $\sigma_d - B$	<b>36.80</b>
$\sigma_{zc} - A - \sigma_d - B - ZC$	35.52	ZC - A - $\sigma_d - B - ZC$	37.22	ZC - A - $\sigma_d - B - ZC$	36.12
$\sigma_{zc} - A - \sigma_d - B - ZC - N$	38.79	ZC - A - $\sigma_d - B - ZC - N$	35.74	ZC - A - $\sigma_d - B - ZC - N$	37.28

The results in Table 5.5 show the selection of suitable features for blur distortion. The networks with following activation functions: sigmoid, hyperbolic tangent sigmoid and Elliot symmetric sigmoid, are evaluated based on MSE. The selection of an optimum feature set is done using forward-selection rule. Simulation results show that following four features: *standard deviation of ZC rate, average absolute difference, standard deviation of difference image, and blocking factor*, improved the network's performance for sigmoid function. Whereas, *ZC rate, average absolute difference, standard deviation of difference image, and blocking factor* improved the network's performance for the

hyperbolic tangent sigmoid and Elliot symmetric sigmoid functions, to its maximum. The addition of any other feature didn't improve the networks' performance.

**Table 5.6: Optimum features selection for noise distortion**

Noise					
Sigmoid	MSE	Hyperbolic Tangent Sigmoid	MSE	Elliot Sigmoid	MSE
$B$	17.1	$B$	18.7	$B$	17.7
$B - ZC$	5.96	$B - ZC$	5.96	$B - ZC$	6.96
$B - ZC - N$	<b>5.02</b>	$B - ZC - N$	<b>5.17</b>	$B - ZC - N$	<b>5.64</b>
$B - ZC - N - \sigma_{zc}$	5.16	$B - ZC - N - \sigma_{zc}$	6.56	$B - ZC - N - \sigma_{zc}$	5.82
$B - ZC - N - \sigma_{zc} - \sigma_d$	5.4	$B - ZC - N - \sigma_{zc} - \sigma_d$	6.18	$B - ZC - N - \sigma_{zc} - \sigma_d$	5.91
$B - ZC - N - \sigma_{zc} - \sigma_d$ - $A$	5.23	$B - ZC - N - \sigma_{zc} - \sigma_d - A$	6.96	$B - ZC - N - \sigma_{zc} - \sigma_d$ - $A$	5.76

The performance of the networks in terms on the MSE for noise distortion is shown in Table 5.6. The following three features; namely, *blocking factor*, *ZC rate*, and *noise mean* improved the network's performance for all three activation functions. The forward-selection rule is employed for feature selection. The addition of any other feature resulted in no improvement in network's performance. Thus, these three features are selected for network's training for noise distortion.

**Table 5.7: Optimum features selection across all distortions**

Across All Distortions					
Sigmoid	MSE	Hyperbolic Tangent Sigmoid	MSE	Elliot Sigmoid	MSE
$ZC$	165.9	$\sigma_{zc}$	167.7	$\sigma_{zc}$	167.9
$ZC - \sigma_{zc}$	99.93	$\sigma_{zc} - ZC$	111.3	$\sigma_{zc} - ZC$	95.5
$ZC - \sigma_{zc} - B$	67.96	$\sigma_{zc} - ZC - B$	88.61	$\sigma_{zc} - ZC - B$	57.04
$ZC - \sigma_{zc} - B - N$	62.27	$\sigma_{zc} - ZC - B - \sigma_d$	83.15	$\sigma_{zc} - ZC - B - N$	52.07
$ZC - \sigma_{zc} - B - N - \sigma_d$	<b>58.18</b>	$\sigma_{zc} - ZC - B - \sigma_d - N$	71.02	$\sigma_{zc} - ZC - B - N - \sigma_d$	<b>51.5</b>
$ZC - \sigma_{zc} - B - N - \sigma_d - A$	59.85	$\sigma_{zc} - ZC - B - \sigma_d - N - A$	<b>67.58</b>	$\sigma_{zc} - ZC - B - N - \sigma_d - A$	53.16

The networks' performance across all distortions is shown in Table 5.7. The performance is evaluated based on MSE between predicted and desired scores. The simulation results show that for sigmoid and Elliot symmetric sigmoid functions, following five features: *ZC rate, blocking factor, noise mean, standard deviation of ZC rate, and standard deviation of difference image* improved the network's performance. While, all six features are considered in network's training for the hyperbolic tangent function. The order of the selected features is decided using forward-selection rule.

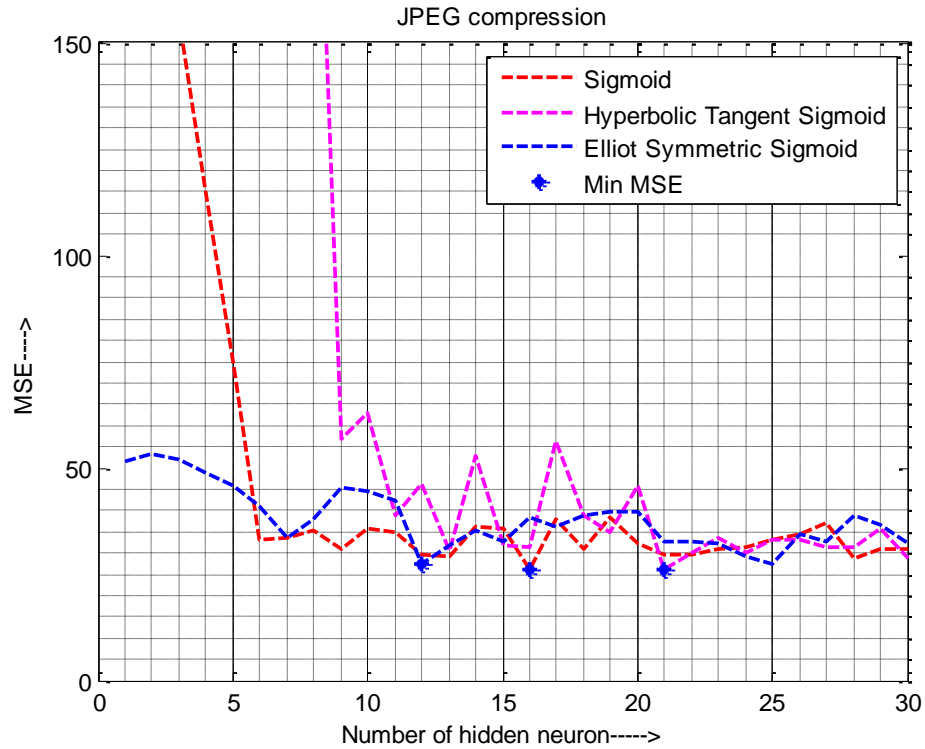


## 5.2.2 Optimum number of hidden neurons

After the selection of the optimum feature set, the next task is to find the optimum number of neurons in the network's hidden layer. This is obtained by the neural networks with varied number of hidden neurons, and evaluating their performance using the MSE. The network with the lowest MSE gives the minimum number of hidden neurons required for successful learning of the neural network besides maintaining its generalization capability. The optimum number of hidden nodes obtained for the different distortion cases are shown in Table 5.8.

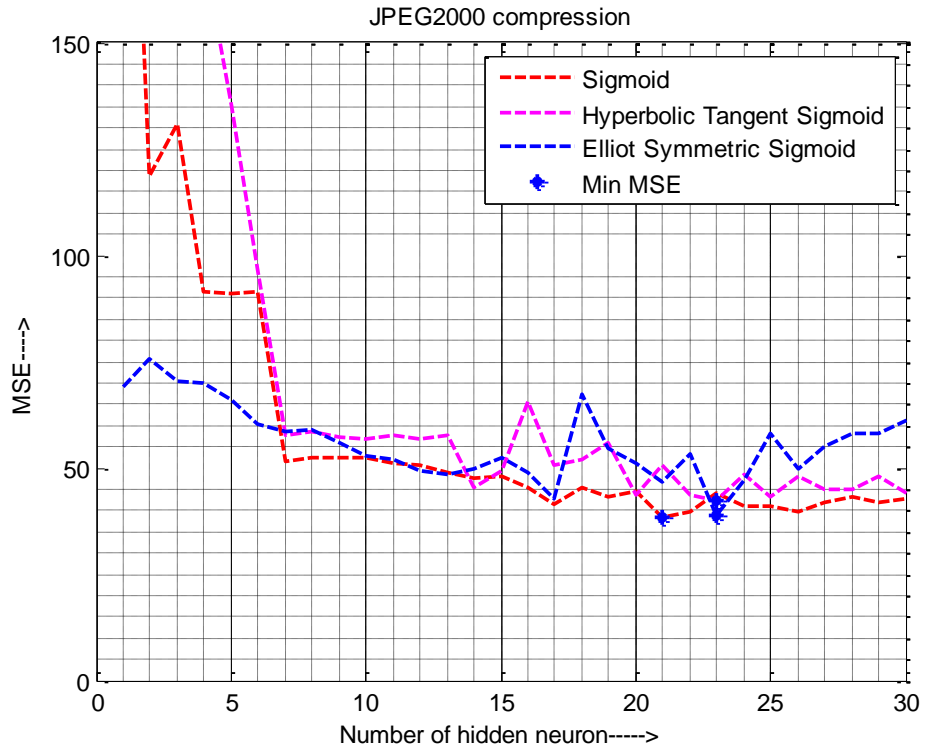
**Table 5.8: Selection of optimum number of hidden neurons**

Distortion Type	Sigmoid		Hyperbolic Tangent Sigmoid		Elliot Sigmoid	
	Number of neurons	MSE	Number of neurons	MSE	Number of neurons	MSE
JPEG	16	25.88	21	25.92	12	27.25
JPEG2000	21	38.59	23	42.46	23	39.02
Blur	11	28.23	12	35.72	16	38.64
Noise	14	4.92	12	5.2	10	6.16
Across all	34	49.45	50	67.59	37	52.07



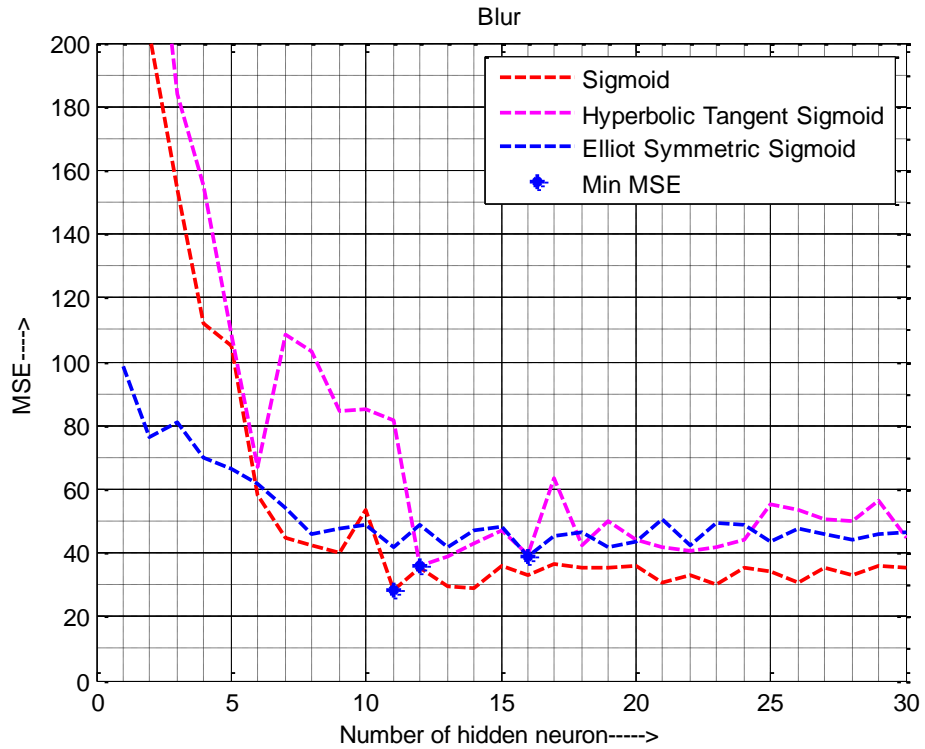
**Figure 5.1: MSE vs. varied number of hidden neurons for JPEG compression**

In Figure 5.1, the error rates of the networks with sigmoid, hyperbolic tangent, and Elliot sigmoid functions for JPEG compression are shown. The performance graph shows that the lowest MSE is obtained at **HN=16** for the neural network with sigmoid function, **HN=21** for hyperbolic tangent, and **HN=12** for Elliot sigmoid function, where HN indicated the number of neurons in hidden layer. These HN values are considered to be the optimum number of hidden neurons required for the effective training of the neural network for the JPEG compression distortion.



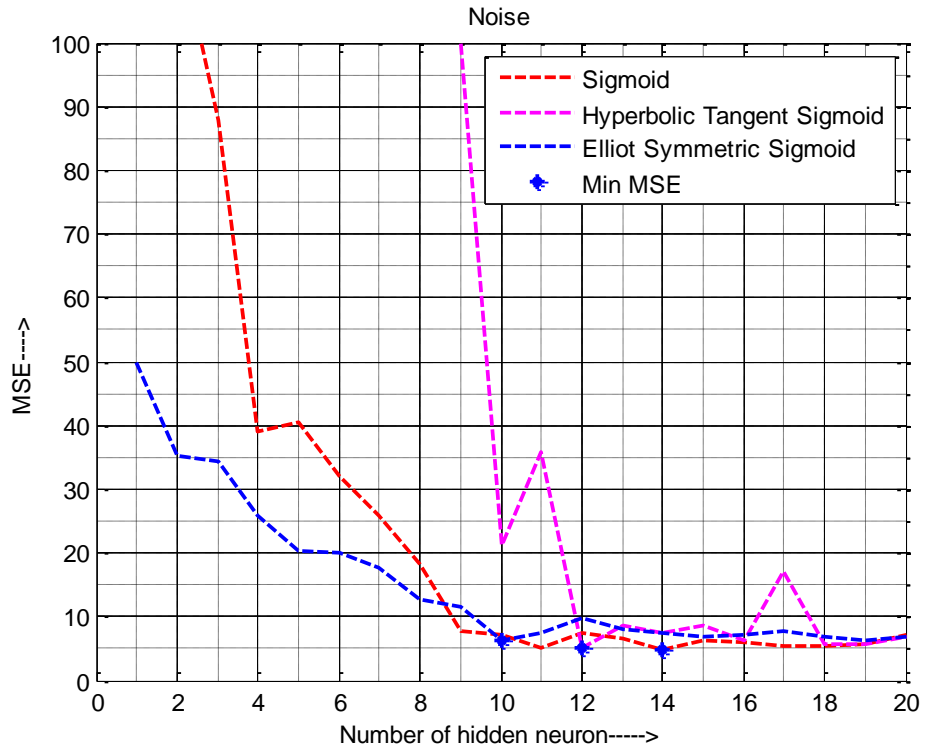
**Figure 5.2: MSE vs. varied number of hidden neurons for JPEG2000 compression**

The results in Figure 5.2 show the error rates of the networks trained with varied number of hidden neurons (HN) for JPEG2000 compression. For neural network with sigmoid function, the lowest MSE is obtained at **HN=21**, whereas, **HN=23** gave the lowest value of MSE for hyperbolic tangent and Elliot sigmoid functions. These HN values are considered to be the optimum number of hidden neurons for effective network's training for JPEG2000 compression distortion.



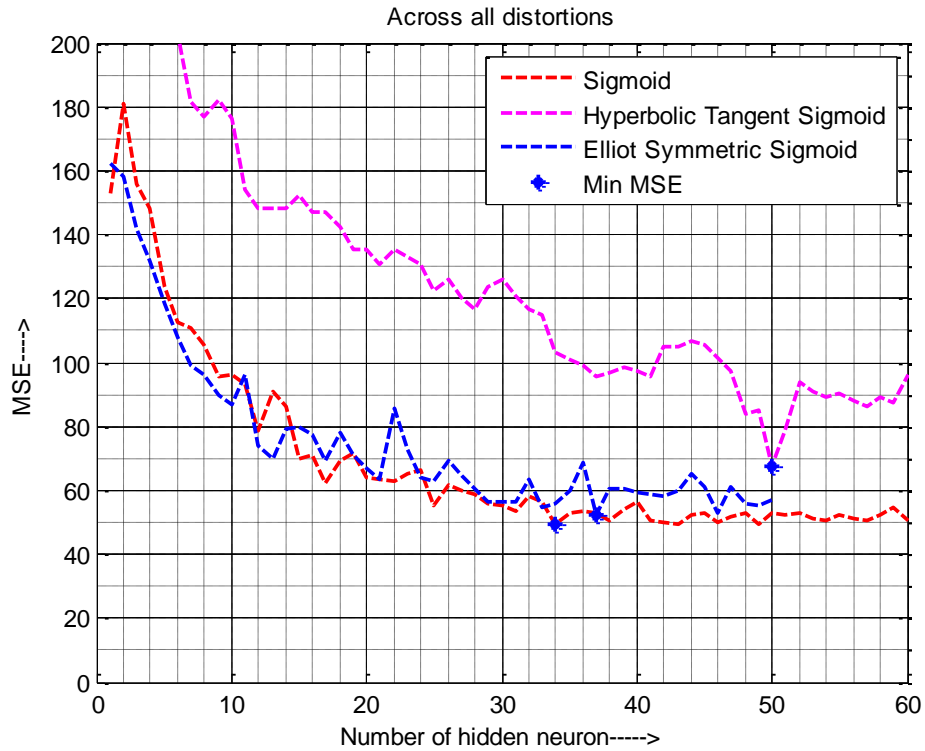
**Figure 5.3: MSE vs. varied number of hidden neurons for blur distortion**

The error rates of the networks for blur distortion are shown in Figure 5.3. The network with sigmoid function showed lowest MSE at hidden neurons **HN=11**. While, hyperbolic tangent and Elliot sigmoid functions showed minimum MSE at **HN=12** and **HN=16** respectively. These HN values are chosen for effective training of the neural network for the blur distortion.



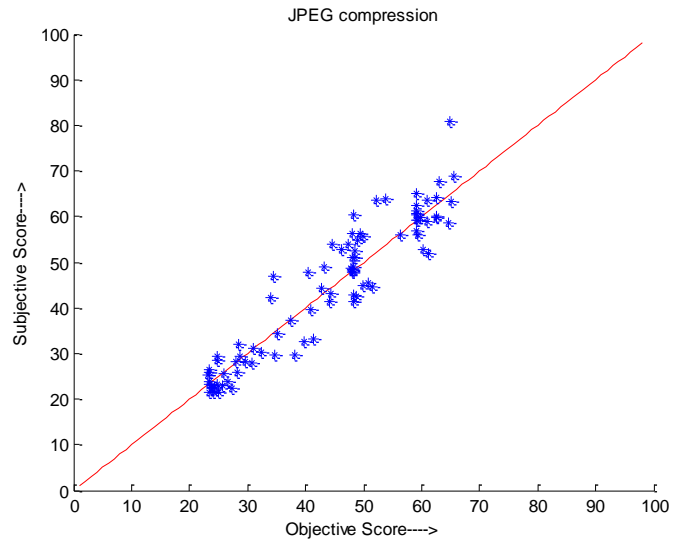
**Figure 5.4: MSE vs. varied number of hidden neurons for noise distortion**

The error rates of the networks for noise distortion are shown in Figure 5.4. For hidden neurons **HN=14**, the network showed minimum MSE for sigmoid function, whereas, **HN=12** and **HN=10** provided minimum MSE for hyperbolic tangent and Elliot sigmoid functions respectively. The networks with these HN values are designed for quality prediction of noise distortion.

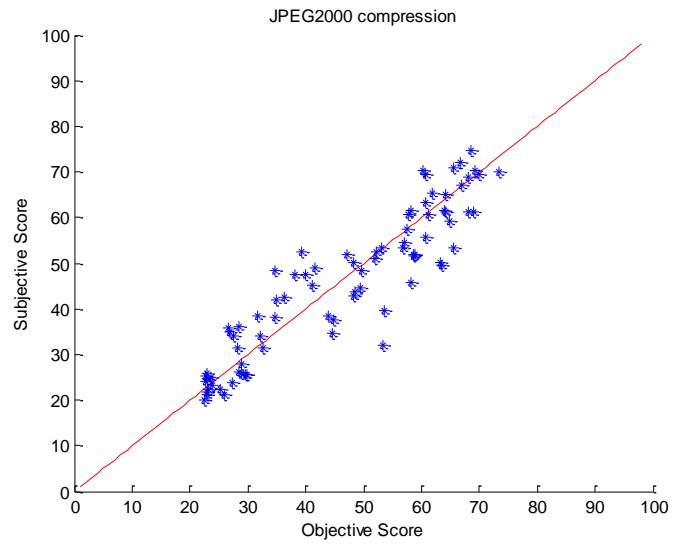


**Figure 5.5: MSE vs. varied number of hidden neurons across all distortions**

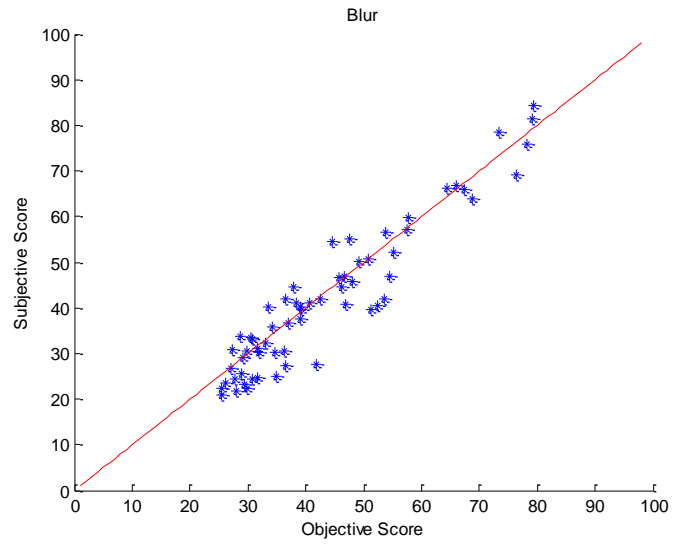
In Figure 5.5, the error rates of the networks with sigmoid, hyperbolic tangent and Elliot sigmoid functions across all distortions are shown. The lowest MSE is observed at hidden neurons **HN=34**, **HN=50**, and **HN=37** for the networks with sigmoid, hyperbolic tangent, and Elliot sigmoid functions respectively. For the better training of the neural network, these HN values are considered across all distortions.



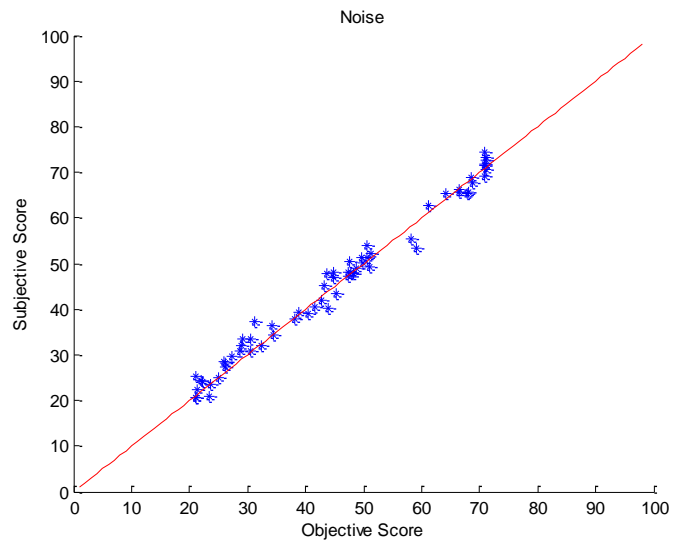
**Figure 5.6: Scatter plot of Subjective scores vs. Objective scores for JPEG compression**



**Figure 5.7: Scatter plot of Subjective scores vs. Objective scores for JPEG2000 compression**

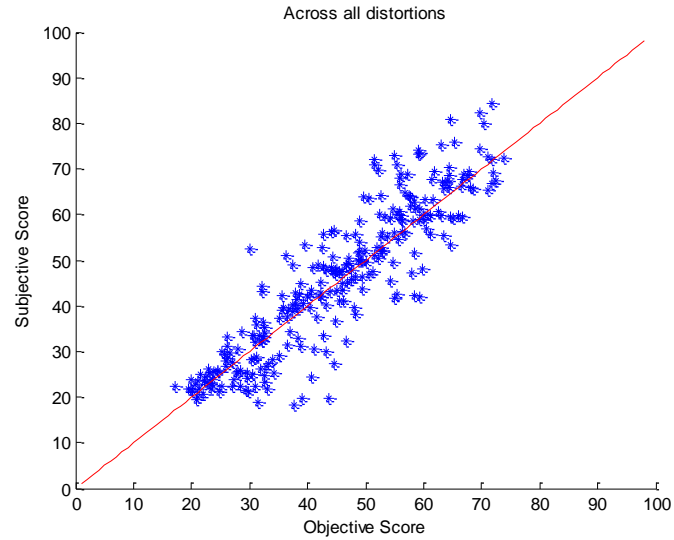


**Figure 5.8: Scatter plot of Subjective scores vs. Objective scores for blur distortion**



**Figure 5.9: Scatter plot of Subjective scores vs. Objective scores for noise distortion**





**Figure 5.10: Scatter plot of Subjective scores vs. Objective scores across all distortions**

The scatter plots in Figures 5.6 – 5.10 show the correlation between the subjective scores from the LIVE database and the scores predicted by the model. The linearity of the scatter points depends on the prediction accuracy. The results show that the network performed significantly well, and predicts the quality score that correlates well with the human observers' evaluations.

### **5.2.3 Performance comparison with traditional algorithms**

The neural network with optimum number of hidden neurons is trained with the optimum feature set, to estimate the subjective quality score. The performance of the proposed approach for all the distortion cases is evaluated, and compared with existing models. The evaluate criteria considered are, the Pearson's Correlation Coefficients (PCC), Spearman's Rank-Order Correlation Coefficient (SROCC), Coefficient of Determination

**Table 5.9: Performance evaluation and comparison of the proposed model**

	Performance Evaluation		PCC	SROCC	R <sup>2</sup>	RMSE	MAE	OR
JPEG	Proposed	Sigmoid	0.94	0.92	0.86	5.09	3.86	0.21
		Hyp. Tan.	0.93	0.91	0.86	5.36	4.21	0.25
		Elliot Sig.	0.94	0.92	0.87	5.22	3.97	0.23
	Bovik's		0.91	0.90	0.81	7.02	5.35	0.40
JPEG2000	Proposed	Sigmoid	0.92	0.92	0.84	6.45	5	0.46
		Hyp. Tan.	0.92	0.92	0.84	6.52	5.18	0.51
		Elliot Sig.	0.93	0.93	0.86	6.25	4.99	0.42
	Bovik's		0.88	0.88	0.79	7.67	6.21	0.56
BLUR	Proposed	Sigmoid	0.95	0.91	0.87	5.31	4.05	0.47
		Hyp. Tan.	0.94	0.91	0.85	5.62	4.52	0.5
		Elliot Sig.	0.93	0.9	0.86	6.22	4.75	0.4
	Bovik's		0.91	0.89	0.83	6.81	5.53	0.59
	CPBD		0.89	0.90	0.73	7.33	5.17	0.55
NOISE	Proposed	Sigmoid	0.99	0.99	0.98	2.22	1.73	0.03
		Hyp. Tan.	0.99	0.99	0.98	2.28	1.74	0.06
		Elliot Sig.	0.99	0.99	0.98	2.48	1.96	0.05
	Bovik's		0.97	0.96	0.96	3.34	2.64	0.15
ALL	Proposed	Sigmoid	0.90	0.90	0.74	7.23	5.34	0.41
		Hyp. Tan.	0.89	0.90	0.7	7.72	5.58	0.42
		Elliot Sig.	0.90	0.90	0.75	7.19	5.33	0.42
	Bovik's		0.80	0.84	0.64	9.60	7.66	0.62

( $R^2$ ), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Outlier Ratio (OR). The details of these evaluation criteria are given in section 3.4.2.

Table 5.9, the performance evaluation of the proposed ANN based NR image quality assessment model is shown. The performance of the neural networks using sigmoid, hyperbolic tangent and Elliot sigmoid activation functions is evaluated with the six evaluation criteria (PCC, SROCC,  $R^2$ , RMSE, MAE, and OR) discussed in section 3.4.2 . The results show that all three neural networks demonstrate similar performance for all the considered distortion cases. Furthermore, the proposed image quality measure is compared to the existing models in [51] and [82]. In [51], Bovik proposed a non-linear fitting technique to predict quality of JPEG compressed images. His framework is extended for other 4 distortion cases (JPEG2000 compression, blur, noise, and across all distortions), and used as a benchmark to study the performance of the proposed NR IQM. The model based on Cumulative Probability of Blur Distortion (CPBD) for blurred image is also employed for comparison [82]. From the results in Table 5.9, it is observed that there is a substantial improvement in all the evaluation criteria for the five distortion cases. There is a significant improvement in prediction accuracy (PCC) by 3% for JPEG, 4% for JPEG2000, 4% blur, and 10% across all distortions. The results show that the proposed NR IQM outperforms the traditional models in [51] and [82] by a very good margin.

### **5.3 Summary**

A machine learning-based NR-IQA algorithm using neural network is developed. The proposed approach is modeled for five distortion cases; namely, blur, noise, JPEG compression, JPEG2000 compression, and across all distortions. The performance of the proposed algorithm is evaluated using six evaluation criteria recommended by VQEG. The performance of the three network topologies with different activation functions is evaluated for all distortion cases, and even compared with the traditional algorithm. Simulation results show that there exist an excellent correlation between the subjective and the predicted quality scores. The proposed algorithm showed excellent results in terms of prediction consistency (low OR), prediction accuracy (high PCC), prediction monotonicity (high SROCC), with low prediction error (RMSE and MAE). The simulation results also showed that neural network-based NR-IQA outperforms the traditional algorithm.

## CHAPTER 6

### CONCLUSION AND FUTURE WORKS

#### 6.1 Conclusion

This research focuses on the development of a NR image quality measure. The field of NR image quality assessment is still in its beginning. The intrinsic complexity and limited knowledge of the human visual perception, makes the task of NR-IQM very challenging. In this thesis, we addressed the problem of visual quality assessment in the absence of reference image. We proposed an algorithm for the NR image quality assessment based on Artificial Neural Networks. The proposed framework involves the combination of two steps: feature extraction and neural network training. The ANN is used as a regressor to formulate the IQA problem, by training with the selected statistical features. The proposed algorithm is designed for five distortion cases: blur, noise, JPEG compression, JPEG2000 compression, and across all distortions. Depending on the distortions, six statistical features are extracted that correlated well with the perceived quality. These features include: blocking factor, average absolute difference, zero-crossing rate, standard deviations of the difference image, standard deviation of the zero-crossing rate and noise mean.

An intense training of the neural network is performed using different combination of extracted features and varied number of hidden neurons, to obtain the optimum network. The performance of the propose algorithm is evaluated based on the standard evaluation

criteria, recommend by the VQEG. The proposed algorithm is also compared with existing quality measures presented by Bovik [51] and Karam [82]. The simulation results showed that the proposed algorithm outperforms existing models, with excellent correlation with human observer evaluation. The performance of the algorithm affirms that the machine-learning approach is a powerful technique, and can be implemented for any type of distortion.

## **6.2 Future Works**

To further enhance the neural network-based approach for no-reference image quality assessment, we list a number of ideas below:

- Other architectures of ANNs can be designed and compared with the network topologies incorporated in this research.
- Other statistical features can be investigated for machine learning, and compared to those employed in this research.
- The scope of neural network-based NR-IQM can be extended to other single-distortion types and to multiple-distortions as well.

## REFERENCES

- [1] U. Engelke, “Perceptual-based Quality Metrics for Image and Video Services : A Survey,” pp. 190–197, 2007.
- [2] K. Thung, “A Survey of Image Quality Measures.”
- [3] S. E. Palmer, *Vision science – Photons to phenomenology*, 3rd ed. The MIT Press, Cambridge, 2002.
- [4] M. D. Fairchild, *Color Appearance Models*. Addison-Wesley, 1998.
- [5] M. Livingstone and D. Hubel, “Segregation of Depth : Anatomy , Form ,” vol. 240, no. 4853, pp. 740–749, 2009.
- [6] “VQEG: The Video Quality Experts Group.” [Online]. Available: <http://www.vqeg.org>.
- [7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley, New York, 1992.
- [8] Z. Wang, A. Bovik, and Cl. Lu, “WHY IS IMAGE QUALITY ASSESSMENT SO DIFFICULT?,” pp. 3313–3316, 2002.
- [9] A. M. Eskicioglu and P. S. Fisher, “Image Quality Measures and Their Performance,” vol. 43, no. 12, pp. 2959–2965, 1995.
- [10] Z. Wang and A. C. Bovik, “Mean squared error: love it or leave it,” no. January, pp. 98–117, 2009.
- [11] S. Winkler, “A Perceptual Distortion Metric for Digital Color Video,” *Proc. SPIE*, vol. 3644, pp. 175–184, 1999.
- [12] T. N. Pappas and R. J. Safranek, “Perceptual criteria for image quality evaluation,” *in Handbook of Image and Video Processing*, no. New York: Academic Press, 2000.
- [13] S. Winkler, “Issues in vision modeling for perceptual video quality assessment,” *Signal Processing*, vol. 78, no. 2, pp. 231–252, Oct. 1999.
- [14] J. Lubin, “The use of psychophysical data and models in the analysis of display system performance,” *in Digital Images and Human Vision*, Cambridge, MA: The MIT Press, 1993, pp. 163–178.

- [15] JNDmetrix Technology Sarnoff Corp., “Sarnoff JND Metrix, Evaluation Version,” 2003. [Online]. Available: <http://www.sarnoff.com/products->.
- [16] P. J. Burt and E. H. Adelson, “The Laplacian Pyramid as a Compact Image Code,” vol. C, no. 4, pp. 532–540, 1983.
- [17] E. Peli, “Contrast in complex images.,” *Journal of the Optical Society of America. A, Optics and image science*, vol. 7, no. 10, pp. 2032–40, Oct. 1990.
- [18] W. T. Freeman and E. H. Adelson, “The design and use of steerable filters,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
- [19] P. C. Teo and D. J. Heeger, “Perceptual image distortion,” *Proceedings of 1st International Conference on Image Processing*, vol. 2, pp. 982–986.
- [20] P. Teo and D. J. Heeger, “Perceptual image distortion,” *Proc. SPIE*, vol. 2179, pp. 127–141, 1994.
- [21] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.
- [22] A. B. Watson, “QUANTIZATION MATRICES FOR INDIVIDUAL IMAGES .,” *Society for Information Display Digest of Technical Papers*, vol. xxiv, pp. 946–949, 1993.
- [23] S. A. Karunasekera and N. G. Kingsbury, “A distortion measure for blocking artifacts in images based on human visual sensitivity.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 4, no. 6, pp. 713–24, Jan. 1995.
- [24] M. Miyahara, S. Member, K. Kotani, V. R. Algazi, and L. S. Member, “Objective Picture Quality Scale ( PQS ) for Image Coding,” *IEEE Trans. Communications*, vol. 46, no. 9, pp. 1215–1226, 1998.
- [25] S. Wankler, “A perceptual distortion metric for digital color images,” *IEEE Int. Conf. Image Processing*, vol. 3, pp. 399–403, 1998.
- [26] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and a C. Bovik, “Image quality assessment based on a degradation model.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 9, no. 4, pp. 636–650, Jan. 2000.
- [27] Z. Wang, A. C. Bovik, and L. Lu, “Foveated Wavelet Image Quality Index,” *Proc. SPIE Int. Society for Optical Engineering*, vol. 4472, pp. 42–52, 2001.



- [28] W. Lin, L. Dong, and P. Xue, “Visual distortion gauge based on discrimination of noticeable contrast changes,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 900–909, 2005.
- [29] D. M. Chandler and S. S. Hemami, “VSNR: a wavelet-based visual signal-to-noise ratio for natural images.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
- [30] D. M. Chandler and S. S. Hemami, “Visual Signal-to-Noise Ratio.” [Online]. Available: <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
- [31] H. R. Sheikh, “An Information Fidelity Criterion for Image Quality Assessment.” [Online]. Available: [http://live.ece.utexas.edu/research/quality/ifcvec\\_release.zip](http://live.ece.utexas.edu/research/quality/ifcvec_release.zip).
- [32] H. R. Sheikh, “Visual Information Fidelity Measure for Image Quality Assessment.” [Online]. Available: [http://live.ece.utexas.edu/research/quality/vifvec\\_release.zip](http://live.ece.utexas.edu/research/quality/vifvec_release.zip).
- [33] H. R. Sheikh and A. C. Bovik, “Image information and visual quality.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [34] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
- [35] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image Denoising Using Scale Mixtures of Gaussians in the Wavelet Domain,” *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [36] D. Van der Weken, M. Nachtgael, and E. E. Kerre, “Using similarity measures and homogeneity for the comparison of images,” *Image and Vision Computing*, vol. 22, no. 9, pp. 695–702, Aug. 2004.
- [37] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, “An SVD-based grayscale image quality measure for local and global assessment.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 2, pp. 422–9, Feb. 2006.
- [38] H. Han, D. Kim, S. Member, R. Park, and S. Member, “Structural Information-Based Image Quality Assessment using LU factorization,” *IEEE Trans. Consumer Electronics*, vol. 55, no. 1, pp. 165–171, 2009.

- [39] A. Bouzerdoum, A. Havstad, and A. Beghdadi, "Image Quality Assessment using a Neural Network Approach," *4th IEEE International Symposium on Signal Processing and Information Technology*, pp. 330–333, 2004.
- [40] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Human Vision and Electronic Imaging X, Proc. SPIE*, vol. 5666, pp. 149–159, Mar. 2005.
- [41] T. M. Cover, J. A. Thomas, J. Bellamy, R. L. Freeman, and J. Liebowitz, *Elements of Information Theory WILEY SERIES IN Expert System Applications to Telecommunications*. 1991.
- [42] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 2091–2110, 1989.
- [43] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 6, pp. 1680–9, Jun. 2006.
- [44] A. Chetouani, A. Beghdadi, M. Deriche, and A. Bouzerdoum, "A Reduce Reference Image Quality Metric based on Feature Fusion and Neural Networks," *19th European Signal Processing Conference*, pp. 589–593, 2011.
- [45] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 18, no. 7, pp. 1409–23, Jul. 2009.
- [46] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electronics*, vol. 38, no. 1, 1992.
- [47] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–278, Nov. 1998.
- [48] K. S. Randhawa and P. Kumar, "A Novel Approach for Blocking Artifact Reduction in JPEG Compressed Images," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 2, 2012.
- [49] L. Meesters and J.-B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Processing*, vol. 82, no. 3, pp. 369–387, Mar. 2002.
- [50] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu, and X. Yang, "A locally adaptive algorithm for measuring blocking artifacts in images and videos," *Signal Processing: Image Communication*, vol. 19, no. 6, pp. 499–506, Jul. 2004.

- [51] Z. Wang, H. R. Sheikh, and A. C. Bovik, "NO-REFERENCE PERCEPTUAL QUALITY ASSESSMENT OF JPEG COMPRESSED IMAGES," *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 477–480., 2002.
- [52] X. Li, "Blind image quality assessment," *IEEE Int. Conf. Image Processing*, vol. 1, pp. 449–452, 2002.
- [53] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 4, no. 11, pp. 317–320, Nov. 1997.
- [54] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," *Proc. IEEE Int. Conf. Image Processing*, vol. 3, pp. 981–984, 2000.
- [55] S. Liu and A. C. Bovik, "Efficient DCT-Domain Blind Measurement and Reduction of Blocking Artifacts," vol. 12, no. 12, pp. 1139–1149, 2002.
- [56] T. Brandão and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics," *Signal Processing*, vol. 88, no. 4, pp. 822–833, Apr. 2008.
- [57] R. Venkatesh Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Processing*, vol. 87, no. 6, pp. 1493–1503, Jun. 2007.
- [58] J. Martens, "The Hermite Transform--Theory," *IEEE Trans. Acoustics Speech Signal Processing*, vol. 38, no. September, pp. 1595–1618, 1990.
- [59] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 Still Image," *IEEE Signal Processing Magazine*, vol. 18, no. September, pp. 36–58, 2001.
- [60] S. H. Oguz, Y. H. Hu, and T. Q. Nguyen, "Image coding ringing artifact reduction using morphological post-filtering," *Proc. IEEE 2nd Workshop on Multimedia Signal Processing*, pp. 628–633, 1998.
- [61] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [62] H. Tong, M. Li, H. J. Zhang, and C. Zhang, "No-reference quality assessment for jpeg2000 compressed images," *2004 International Conference on Image Processing, 2004. ICIP '04.*, vol. 5, pp. 3539–3542.
- [63] R. Barland and A. Saadane, "Reference free quality metric for JPEG2000 compressed images," *Proc. IEEE 8th Int. Symp. Signal Processing and Its Applications*, vol. 1, pp. 351–354, 2005.

- [64] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 14, no. 11, pp. 1918–27, Nov. 2005.
- [65] H. R. Sheikh, "No Reference Image Quality Assessment for JPEG2000." [Online]. Available: [http://live.ece.utexas.edu/research/quality/jp2knr\\_release.zip](http://live.ece.utexas.edu/research/quality/jp2knr_release.zip).
- [66] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 8, no. 12, pp. 1688–701, Jan. 1999.
- [67] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 257–268, Apr. 2008.
- [68] H. Liu, S. Member, N. Klomp, and I. Heynderickx, "A No-Reference Metric for Perceived Ringing Artifacts in Images," vol. 20, no. 4, pp. 529–539, 2010.
- [69] H. Liu, N. Klomp, and I. Heynderickx, "A perceptually relevant approach to ringing region detection.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 19, no. 6, pp. 1414–26, Jun. 2010.
- [70] C. Tomasi, "Bilateral Filtering for Gray and Color Images," *Proc. IEEE Int. Conf. Computer Vision*, pp. 836–846, 1998.
- [71] J. Canny, "A computational approach to edge detection.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 8, no. 6, pp. 679–98, Jun. 1986.
- [72] S. Wu, W. Lin, L. Jiang, W. Xiong, L. Chen, and S. H. Ong, "An Objective Out-of-Focus Blur Measurement," *2005 5th International Conference on Information Communications & Signal Processing*, pp. 334–338, 2005.
- [73] P. Marziliano, F. Dufaux, and S. Winkler, "Perceptual Blur and Ringing Metrics : Application to JPEG2000," *Signal Processing: Image Communication*, no. 2, pp. 163–172, 2004.
- [74] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiarrg, and F. Moscheni, "A no-reference quality metric for measuring image blur," *Proc. IEEE 7th Int. Symp. Signal Processing and Its Applications*, vol. 1, no. 4, pp. 169–472, 2001.
- [75] C.-Y. Wee and R. Paramesran, "Image sharpness measure using eigenvalues," *2008 9th International Conference on Signal Processing*, pp. 840–843, Oct. 2008.

- [76] V. Cuong and D. M. Chandler, "S3: A Spectral and Spatial Sharpness Measure," *First International Conference on Advances in Multimedia*, pp. 37–43, Jul. 2009.
- [77] M. J. Chen and A. C. Bovik, "NO-REFERENCE IMAGE BLUR ASSESSMENT USING MULTISCALE GRADIENT," 2000.
- [78] C. V Vu and D. M. Chandler, "A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation," vol. 19, no. 7, pp. 423–426, 2012.
- [79] R. Hassen, W. Zhou, and M. Salama, "NO-REFERENCE IMAGE SHARPNESS ASSESSMENT BASED ON LOCAL PHASE COHERENCE MEASUREMENT," *Conference, Ieee International Processing, Signal*, no. x, pp. 1–4, 2010.
- [80] M. G. Choi, J. H. Jung, and J. W. Jeon, "No-Reference Image Quality Assessment using Blur and Noise," pp. 318–322, 2009.
- [81] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," *International Workshop on Quality of Multimedia Experience*, pp. 87–91, 2009.
- [82] N. D. Narvekar and L. J. Karam, "Correspondence Cumulative Probability of Blur Detection ( CPBD )," vol. 20, no. 9, pp. 2678–2683, 2011.
- [83] A. Chetouani, A. Beghdadi, S. Chen, and G. Mostafaoui, "A Novel Free Reference Image Quality Metric Using Neural Network Approach," *International Workshop on Video Processing and Quality Metrics*, pp. 62–65, 2010.
- [84] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," *IEEE International Conference on Image Processing*, vol. 3, pp. 981–984, 2000.
- [85] A. Chetouani, A. Beghdadi, and M. Deriche, "A new free reference image quality index for blur estimation in the frequency domain," *IEEE ISSPIT*, 2009.
- [86] A. K. Moorthy and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, May 2010.
- [87] M. A. Saad and A. C. Bovik, "BLIINDS: BLind Image Integrity Notator using DCT Statistics," 2012. [Online]. Available: <http://live.ece.utexas.edu/research/quality/BLIINDS.zip>.
- [88] A. Srivastava, A. B. Lee, and E. P. Simoncelli, "On Advances in Statistical Modeling of Natural Images," pp. 17–33, 2003.

- [89] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer Verlag, 2000.
- [90] M. A. Saad, S. Member, A. C. Bovik, and C. Charrier, "A DCT Statistics-Based Blind Image Quality Index," vol. 17, no. 6, pp. 583–586, 2010.
- [91] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," 2005. [Online]. Available: <http://live.ece.utexas.edu/research/quality>.
- [92] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3441–3452, 2006.
- [93] W. C. Lefebvre, J. C. Principe, and N. R. Euliano, *Neural and adaptive system: fundamentals through simulation*. John Wiley & sons, INC, 2000.
- [94] J. M. Zurada, *Introduction to artificial neural systems*. PWS publishing company, 1992.
- [95] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," in *The Bulletin of Mathematical Biophysics*, 1943, pp. 115–133.
- [96] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon. 1995.
- [97] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice-Hall International, London, UK, 1999.
- [98] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning internal representation by error propagation, Parallel Distributed*. MIT Press, Cambridge, Mass, 1986, pp. 318–362.
- [99] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, UK, 2008.
- [100] M. M. Polycarpou and P. A. Ioannou, "Modelling, Identification and Stable Adaptive Control of Continuous-Time Nonlinear Dynamical Systems Using Neural Networks.," *American Control Conference*, pp. 36–40, 1992.
- [101] B. Ozkaya, A. Demir, and M. Bilgili, "Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors," in *Environmental Modelling & Software*, 2007, pp. 815–822.
- [102] D. B. Parker, *Learning-logic*. 1985.

- [103] P. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Harvard University, Cambridge, MA, 1974.
- [104] W. Fischer, *Digital Television: A Practical Guide for Engineers*. Springer-Verlag Berlin, New York, USA.
- [105] K. Elleithy, *Advance Techniques in Computing Sciences and Software Engineering*. Springer, New York, USA.
- [106] S. V Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, Fourth Edi. John Wiley & sons, The Atrium, Southern Gate, West Sussex, UK.
- [107] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press Inc., San Diego, California, 1995.
- [108] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303–314, 1989.
- [109] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, pp. 4–22.

## VITAE

Name : Amer Bin Ziyad

Nationality : Indian

Date of Birth : 12<sup>th</sup> February, 1986

Present Address : King Fahd University of Petroleum and Minerals, P.O.  
Box 8656, Dhahran 31261, Saudi Arabia

Email : amerbinziyad@hotmail.com

Permanent Address : 18-1-350/13, Fatima Nagar, Chandrayan Gutta,  
Hyderabad 500005, Andhra Pradesh, India

Contact Numbers : +966560809117, +919966292248

Academic Background : Bachelor of Engineering in Electronics and  
Communications Engineering from Osmania University,  
India in 2008.