DEVELOPMENT OF AN ROI AWARE FULL-REFERENCE OBJECTIVE PERCEPTUAL QUALITY METRIC ON IMAGES OVER FADING CHANNEL

Sri Lohith Gogineni

Faculty of Engineering
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden
This thesis is submitted to the Faculty of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering with emphasis on Radio Communications. The thesis is equivalent to 26 weeks of full-time studies.

Contact Information:
Author:
Sri Lohith Gogineni
E-mail: srilohithgogineni@gmail.com

University advisor:
Prof. Hans-Jürgen Zepernick
Department of Communication Systems
Abstract

In spite of technological advances in wireless systems, transmitted data suffers from impairments through both lossy source coding and transmission over error prone channels. Due to these errors, the quality of multimedia content is degraded. The major challenge for service providers in this scenario is to measure the perceptual impact of distortions to provide certain Quality of Experience (QoE) to the end user.

The general tendency of the Human Visual System (HVS) suggests that the artifacts in the Region-of-Interest (ROI) are perceived to be more annoying compared to the artifacts in Background (BG). With this assumption, the thesis aims to measure the quality of image over ROI and BG independently. Visual Information Fidelity ($VIF$), a full-reference image quality assessment is chosen for this purpose. Finally, the metric measured over ROI and BG are pooled to get a ROI aware metric. The ROI aware metric is used to predict the Mean Opinion Score (MOS) of an image.

In this study, an ROI aware quality metric is used to measure the quality of a set of distorted images generated using a wireless channel. Eventually, MOS of the distorted images is estimated. Lastly, the predicted MOS is validated with the MOS obtained from subjective tests.

Testing the proposed image quality assessment approach shows an improved prediction performance of ROI aware quality metric over traditional image quality metrics. It is also observed that the above approach provides a consistent improvement over a wide variety of distortions. After extensive research, the obtained results suggest that the impairments in the ROI are perceived to be more annoying than that of the BG.

Keywords: Visual Information Fidelity, Quality of Experience, Image, Full-Reference, Objective Perceptual Metrics, Region-of-Interest.
Acknowledgements

The journey of this thesis has been truly a great experience. I am grateful to have succeeded in doing my research in the field of wireless communications. It is a genuine pleasure to express my deep sense of gratitude to Prof. Dr. Hans-Jürgen Zepernick for providing me the opportunity to pursue my master thesis under his supervision. I sincerely thank Dr. Thi My Chinh Chu for her valuable suggestions throughout my thesis.

I am grateful to my parents, Mr. G. Ramesh and Ms. Manjula, my sister, Ramya Shuba for their constant support and encouragement in all my endeavors in life. In addition, I would like to thank my friends for their help and support, without which my thesis work could not have been successful.

Sri Lohith Gogineni
## Contents

Abstract i

Acknowledgements ii

Acronyms v

1 Introduction 1
  1.1 Research Questions .................................. 3
  1.2 Main Contributions .................................. 4

2 Background and Related Work 5
  2.1 Subjective Image Quality Assessment ................ 5
  2.1.1 Popular Image Quality Assessment Databases ...... 6
  2.2 Objective Image Quality Assessment .................. 7
  2.2.1 Metrics based on the HVS .......................... 7
  2.2.2 Metrics based on the Distortions in Visual Content ... 7
  2.2.3 Metrics based on the Reference Information ........ 7
  2.3 Full-Reference Quality Assessment .................... 8
  2.3.1 Literature Review .................................. 8
  2.3.2 Visual Information Fidelity ....................... 10
  2.3.3 Structural Similarity ............................. 10
  2.4 Visual Attention Framework ........................... 11
  2.4.1 Review of Visual Attention with Explicit Design Methods 11
  2.5 Motivation for Visual Attention based Quality Assessment ... 12

3 Methodology 13
  3.1 Overview of the Method Implemented .................. 13
  3.1.1 Framework for FR Quality Assessment in a Wireless Communication System .......................... 13
  3.1.2 Reference and Distorted Images .................... 14
  3.1.3 Objective Metric Model ............................ 14
  3.1.4 Evaluating ROI aware Objective Quality Metric .... 15
  3.2 Perceptual Metric Weight Optimisation ................ 16
  3.2.1 Mapping Function ................................. 18
Acronyms

BG  Background
FR  Full-Reference
MOS Mean Opinion Score
$P_{corr}$ Pearson Correlation
$PMOS$ Predicted Mean Opinion Score using $VIF$
$PMOS_{RA}$ Predicted Mean Opinion Score using $VIF_{RA}$
QoE Quality of Experience
RA ROI aware
RMSE Root Mean Squared Error
ROI Region-of-Interest
$S_{corr}$ Spearman Correlation
$SSIM$ Structural Similarity
VA Visual Attention
$VIF$ Visual Information Fidelity
$VIF_{ROI}$ Visual Information Fidelity of ROI
$VIF_{BG}$ Visual Information Fidelity of BG
$VIF_{RA}$ ROI aware Visual Information Fidelity
Chapter 1

Introduction

The evolution in wired and wireless communication namely Internet, third and fourth generation mobile networks support traditional voice services to mobile multimedia services. Currently, the widespread adoption of smart phones and rapid increase in the number of devices let users consume more mobile video services. Services like Skype, Facetime and other social network applications along with entertainment applications like streaming music, online games demonstrate the growing importance of these devices. Mobile education services have also become popular among mobile users. Edutainment (entertainment designed to educate as well as amuse) services are also emerging that let young children use mobile devices.

Global mobile data traffic grew 74 percent in 2015 due to the drastic increase in usage of mobiles and fourth-generation (4G) traffic [1]. For the first time in 2015, 4G traffic surpassed third-generation (3G) traffic. By the end of 2015, the mobile data traffic increased from 2.2 exabytes per month to 3.6 exabytes per month. 4G mobile connections represented 47 percent of mobile data traffic, while 33 percent of data traffic is represented by 3G in 2015. It is estimated that global mobile traffic will increase by eightfold between 2015 and 2020 [1]. Worldwide wireless devices with access to mobile data networks are the primary contributors to the growth in mobile traffic. Increase in usage of smarter user devices and growth in mobile to mobile connections clearly indicates the growth of Internet of Everything (IoE).

Due to the rapid growth of wireless communications, the demand for robust multimedia transmission with better quality, coverage, and more power is increasing. As the availability of bandwidth is limited, there is a need for a communication system that does not consume more bandwidth for achieving better image quality. Furthermore, real-time applications are important because it is widespread. Reliable image communication with the real-time requirement needs suitable bit-rate, low power, low delay and low complexity maintaining good image quality.

The limited bandwidths in the networks along with the compression techniques result in the data being highly prone to distortions through bit errors and packet loss during transmission. Visual data may go through many stages of
processing before being presented to a human observer. Each stage of processing may introduce distortions that could reduce the quality of final display. Hence it is of vital importance for service providers to objectively measure the impact of these distortions. It is necessary that service providers maintain a certain level of Quality of Experience (QoE) to the end user [2]. However, the limited bandwidth in the network in conjunction with the large amount of image and video data comprise a highly complex and intricate scenario.

Human Visual System (HVS) is one of the sense organs which is vital to gather information from the outside world [3]. Our visual system adapted to changes in the environment during human evolution with the development of new technologies like computers, television and mobile phones. The HVS is getting accustomed to look at these new technologies such as digital images and videos. Our visual system is habituated to look at the immaculate quality of real world environment, so we expect minimum quality of the digital image or video. But the quality is degraded due to several factors such as transmission, capture, compression and display of image or video. The sources of distortions could range from sensor inadequacy, compression, Gaussian noise and motion blurring. Therefore, it is necessary to measure the quality of the content before it reaches the end user. Hence, development of a reliable Image Quality Metric (IQM) is vital on the image and video signal processing systems. Service providers measure the quality of image or video using different metrics to guarantee certain level of quality.

In digital imaging systems, it is crucial to be able to control and predict the image quality so as to enhance and maintain the quality of an image before transmission or storage. Quality of Service (QoS) requirements of wireless image and video services is one of the main challenges for the network operators in the communication systems. However, the quality measurement in wireless communications is a complex task due to three reasons: the computational complexity should be low considering the limited processing power of the mobile devices; as the original image information is not available at the receiver the quality assessment needs to be done with just the received image; the distortions in the image are random and complex with their intensities and distributions.

In order to monitor the quality of wireless communication services, certain perceptual quality metrics are required. The metrics are chosen such that they can precisely quantify the end-to-end visual quality in the communication system. These metrics are utilized to achieve the stern requirements of QoS.

However, the requirement of metrics for visual quality assessment have increased in recent years. Traditionally, fidelity metrics such as the Mean Squared Error (MSE) and the Peak Signal-to-Noise Ratio (PSNR) are used to analyse the performance of the system. Unfortunately, the pixel based assessment used in MSE and PSNR exhibit poor correlation with perceived visual quality [4]. This resulted in development of new objective quality metrics alternative to the traditional metrics.

Historically, the quality of images is described in terms of the distortions in the
visibility of an image, such as blurriness, blockiness and colour shifts. The Just Noticeable Difference (JND) model [5] by Sarnoff assesses the distortions of an image by rating the image subjectively. Janssen has proposed a new philosophy [6] where images are regarded as carriers of visual information instead of two-dimensional signals and visual-cognitive processing is regarded as information processing rather than signal processing.

The Image Quality Assessment (IQA) techniques are basically divided into two types, the subjective method and the objective method. The subjective method involves human beings to evaluate the image quality. The subjective method is perhaps accurate and it is considered as a reference method as human beings are the ultimate users of the multimedia applications. Whereas in the objective methods, an objective metric is chosen to predict the perceived image quality as closely as possible to the subjective method. A Full-Reference (FR) objective metric is chosen for the thesis as it provides reference information while assessing the quality of an image.

Visual Attention (VA) is one of the important features of the HVS as it reduces the complexity of analysis. Although many research works proposed quality prediction methods for images, the impact of VA is not observed while predicting the quality of an image. Therefore, it is necessary to study the quality prediction performance by including the VA into an FR metric.

The content of this thesis is broadly divided into four chapters. Firstly, Chapter 2 presents the background of image quality metrics and its design along with the literature review of various metrics. The motivation of this thesis work is also explained in Chapter 2. Chapter 3 is concerned with the methodology of quality assessment and design of quality metric with VA. Chapter 4 demonstrates the impact of VA on quality assessment with the help of results and analysis. Chapter 5 provides a summary and findings of the thesis along with limitations and future work. Finally, Chapter 5 ends with conclusions of this thesis.

1.1 Research Questions

The research questions which are formulated based on aims and objectives of this thesis are as follows:

- Does inclusion of visual attention with full-reference image quality metrics (VIF, SSIM) improve the quality prediction?

- Does the method of pooling ROI and BG metrics to obtain an ROI aware metric provide a better prediction of human perception?

- Does the scale of distortion in an image affect the performance of an ROI aware metric?

- How does the location of artifact affect the image quality assessment?
1.2 Main Contributions

The main contributions of this thesis are:

- A new technique is implemented to improve the performance of an FR objective quality metric.

- A method of pooling ROI and BG metrics is done to further improve the performance of the image quality metric.

- An ROI aware metric which incorporates VA into quality assessment is proposed for predicting the quality of an image.

- A mapping function is derived for an ROI aware metric based on VIF to predict the mean opinion scores.

- An optimisation procedure to find the parameters of the mapping function is performed to increase the performance of an ROI aware metric based on VIF.

- A comprehensive analysis is performed to examine the impact of scale of distortions and locations of artifacts on the performance of an ROI aware metric based on VIF.
Chapter 2

Background and Related Work

The first essential step in the current research is to understand the process of quality assessment of images. In addition, it is necessary to study the literature and identify the research gaps in the field of quality prediction of images to perform this research. The following section provides a brief groundwork to acquire the knowledge of quality assessment along with the current state of research in the area of image quality metrics.

2.1 Subjective Image Quality Assessment

Since HVS is the ultimate receiver and judge for most of the images, video and graphics, perceptual quality metric development would be more logical and user-oriented. Therefore, the ‘base truth’ for quality prediction is the subjective assessment technique. Also, subjective assessment forms the highest end of a quality prediction performance scale.

The distortion can be calculated by subjective viewing tests with appropriate standard procedures. Subjective tests are performed based on two standards specified by the International Telecommunication Union (ITU). Single stimulus and double stimulus methods are specified by the Radio Communications Sector of the ITU (ITU-R) [7] for testing pictures.

In general, a subjective test can be double stimulus or single stimulus [4] based on the availability of original image. In the double-stimulus methodology, both source and test images are made available to the subject for quality assessment. Also, there are different double stimulus methods available which are based on presenting source image to the subject such as the Double-Stimulus Continuous Quality Scale (DSCQS) and Double-Stimulus Impairment Scale (DSIS). In the single-stimulus method, the quality of an image is evaluated on a linear quality scale without providing the source image.

The scores obtained from several subjects are then averaged to obtain a Mean Opinion Score (MOS) and Difference Mean Opinion Score (DMOS) for each test image. These subjective scores are used to analyse the objective metrics performance. However, these subjective tests are time consuming and expensive as the resultant MOS is obtained by multiple observers through repeated test sessions.
The subjective tests are not feasible for visual signal manipulations (such as encoding, transmission, reception etc.). Even in conditions where subjects are allowed for visual inspection, the assessment depends upon viewers’ physical conditions, personal experience, and emotional states. Therefore, computational models to predict the MOS are built to assess the quality of images. In other words, the objective methodology is employed to approximate the human perception results. The objective metrics are generally advantageous due to the repeatability nature of the measurements. The results of the subjective tests are therefore instrumental for the verification of perceptual IQM. These subjective tests also contribute helpful insight into human perception of images and videos in the presence and absence of distortions.

2.1.1 Popular Image Quality Assessment Databases

Many subjective tests have been performed in various institutions worldwide and the data is made available to public through databases. The most popular database which is publicly available is the LIVE database [8] [9] developed at University of Texas, Austin. It consists of 779 distorted images generated from 29 pristine images with five different distortions. The distortions are: JPEG compression, JPEG 2000 compression, Gaussian blur, fast fading channel distortion and white noise. The subjective scores of all images along with DMOS values are available in the LIVE database.

Another database available for the research community is Computational and Subjective Image Quality (CSIQ) [10] database developed at Oklahoma State University. The CSIQ database includes 866 images derived from 30 source images. Additive white Gaussian noise, Gaussian blurring, additive Gaussian pink noise, JPEG and JPEG2000 compression errors are the distortions used in the CSIQ database. The TID2008 database [11] developed as a joint effort between Finland, Italy and Ukraine includes 1700 impaired images created from 25 reference images. In 2013, the TID2008 database is updated to TID2013 [12] database as an extension to the 2008 database. It contains 3000 images generated from 25 source images using 24 types of distortions.
2.2 Objective Image Quality Assessment

The methods which are automated to predict the quality as it would be perceived by a human observer are referred as objective perceptual quality metrics. Objective quality evaluation can be divide into two broad types: signal fidelity metrics and Perceptual Visual Quality Metrics (PVQMs). The signal fidelity measures are MSE and PSNR which evaluate quality based on pixel information.

As the fidelity metrics often do not align well with the subjective quality assessment, the aim of PVQMs is to combine the advantages of automated evaluation, skipping human interaction, with precise prediction performance. These metrics can be classified into three types [4] based on metric design process. A brief explanation of the three types of metrics will be presented in the following sections.

2.2.1 Metrics based on the HVS

The aim of PVQMs is to achieve a quality prediction close to the perception of a human observer. Therefore, it is necessary to include characteristics of HVS in the design process. Based on the assumptions and approximations of HVS properties, the model can be simple or complex. But a better quality prediction is achieved by using complex approximations of the HVS.

HVS-based metrics are either bottom-up or top-down design based on the inclusion of HVS components in the design. In the bottom-up approach, all the components of HVS are integrated to design a metric. In the top-down approach, only high level assumptions about quality processing in the HVS are simulated.

2.2.2 Metrics based on the Distortions in Visual Content

In this case, the metrics are designed based on the types of distortions accounted while making assumptions. For example, general purpose models do not make any assumptions with respect to the distortions in visual content. Whereas, certain application specific metrics make specific assumptions about distortions in visual content. This information about distortions helps to improve the prediction of quality of visual content. On the other hand, the application specific metrics perform poorly in other applications for which it has not been intended.

2.2.3 Metrics based on the Reference Information

The source image or video content is crucial for the prediction performance of metrics. Based on the availability of source/original content of image/video, various metrics are modelled. When the information of original content is available, then it results in better prediction performance. Therefore, FR metrics are prin-
incipiently used for quality assessment as the entire original image/video content is available.

In No-Reference (NR) metric design, the quality is assessed with no information regarding source/original visual content. The NR approach is based only on the distorted image/video. Thus, it is troublesome to predict the quality for the algorithms in the case of NR approach.

The Reduced-Reference (RR) quality metrics are modelled such that they are a compromise between FR and NR metrics. The RR metrics have only a subset of information from source/original and distorted visual content. Hence, RR metrics have advantages of both FR and NR metrics.

2.3 Full-Reference Quality Assessment

In FR quality assessment, reference is available for the evaluation of the quality of visual content. These methods predict the quality by assessing the degradation of the distorted medium on comparison with the reference. Thus, the FR method provides superior quality prediction performance.

2.3.1 Literature Review

The Picture Quality Scale (PQS) by Yamashita [13] is a technique where spatial and temporal features such as jitter, flicker, noise and blur are extracted from the video. Tan and Ghanbari [14] proposed a blockiness detector for MPEG video.

The Structural Similarity ($SSIM$) index by Wang et al. [15] proposes that HVS is adapted to extraction of structural information. The $SSIM$ index is a compromise between complexity and quality prediction which is the reason for its wide usage as an image quality metric. Sheikh and Bovik proposed the Visual Information Fidelity ($VIF$) criterion [16] for quality prediction from an information theoretic viewpoint. The $VIF$ criterion is superior to $SSIM$ which comes at the cost of higher computational complexity.

The Visual Signal-to-Noise Ratio (VSNR) by Chandler and Hemami [17] is based on a two stage approach, where the first stage involves determining a distortion detection threshold. The VSNR is computed based on perceived contrast and global precedence attributes of the HVS if the distortions are suprathreshold.

The Most Apparent Distortion (MAD) by Larson and Chandler [18] proposes different strategies for determining image quality, based on the near-threshold and suprathreshold visual distortions. The Temporal Trajectory Aware Video Quality Measure (TetraVQM) by Barkowsky et al. [19] focuses on temporal issues such as frame rate reduction, frame freezes and skips, tracking of the visibility of distorted objects and influence of scene cuts.
The FR image quality assessment maybe grouped into various subgroups as shown in Fig. 2.1. A brief description of each subgroup is as follows:

- **Mathematical metrics**
  This type of metrics considers an image as a 2D signal where the similarity or dissimilarity between reference and distorted images is treated as distortion of quality. The Minkowski metric, for example, calculates the distance between samples of reference and distorted images. One such metric is PSNR based on the MSE. The advantage of these metrics are mathematical simplicity and tractability. As mentioned above, the correlation of these metrics with the perceived quality measurement is poor as the HVS characteristics [20] are not considered in the framework of this model.

- **HVS based metrics**
  These metrics are based on the error signal obtained from the difference between reference and test images. The signal is normalized according to its visibility determined by the psychophysics of human perception. The framework of the HVS based metrics consists of pre-processing, error normalization and masking stage. Few research models based on this framework are “visible differences predictor” and Just-Noticeable-Distortion (JND) model by Sarnoff [5].

- **Other Metrics**
  A new framework for measuring the image quality is proposed by Wang et al. [15], the SSIM approach. Another such metric is VIF proposed in [16].
2.3.2 Visual Information Fidelity

Visual information fidelity quality metric is based on the information theoretic framework. The distortion in the image is measured by the degradation of visual quality, in other words by quantification of two mutual information quantities [16]. They are the information between input and output of the HVS channel in the absence of distortion and mutual information between input of distortion channel and output of the HVS channel. In simple terms, the \( VIF \) metric measures the loss of information between reference and distortion images. Natural scene statistics, specifically, Gaussian scale mixtures are used to model the images. The \( VIF \) metric is given by

\[
VIF = \frac{\sum_{j \in \text{subbands}} I(\overrightarrow{C}^N_j, \overrightarrow{F}^N_j | s^N_j)}{\sum_{j \in \text{subbands}} I(\overrightarrow{C}^N_j, \overrightarrow{E}^N_j | s^N_j)}
\] (2.1)

In (2.1), \( I(\cdot,\cdot) \) denotes the mutual information, \( \overrightarrow{C} \) denotes Gaussian Scale Mixtures (GSM), \( N \) denotes number of GSMs utilized, \( s \) is a random field of scalars, and \( \overrightarrow{E}, \overrightarrow{F} \) represent the visual output of the HVS model for reference and distorted images, respectively.

2.3.3 Structural Similarity

The human visual system tends to extract structural information from the viewing field. Therefore, a good approximation of perceived image distortion can be obtained by measuring the structural information change. In the \( SSIM \) context, perceived changes in structural information variation is considered as image degradations. Applying the \( SSIM \) index locally rather than globally is beneficial while measuring the image quality. The \( SSIM \) index does not employ an explicit model of the HVS but it is based on the high-level properties of the HVS. They account for HVS properties such as masking and light adaptation including perception of image structure [21].

The HVS has evolved in such a way that it extracts the structure of images. Based on this fact, the \( SSIM \), which emphasize the structure over the lighting effects of scenes, is a useful perceptual quality metric. The \( SSIM \) approach is sensitive to distortions that breakdown spatial correlation of an image, such as noise, blur and block compression artifacts. However, it is insensitive to distortions due to lighting changes. In simple words, the \( SSIM \) index [4] predicts the structural degradations between images based on simple intensity and contrast measures. The \( SSIM \) index is defined as

\[
SSIM = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}
\] (2.2)
Chapter 2. Background and Related Work

In (2.2), \( C_1 \) and \( C_2 \) are constants used to avoid instability that may occur for mean contrast and intensity combinations. The terms \( \mu_x, \mu_y \) and \( \sigma_x, \sigma_y \) are mean intensity and contrast of the signals \( x \) and \( y \) whereas \( \sigma_{xy} \) denotes covariance of \( x \) and \( y \).

2.4 Visual Attention Framework

Visual attention is based on higher cognitive processing implemented to reduce the overall complexity of scene analysis. Even though the number of visual quality metrics that have been proposed thus far is large, the visual attention, which impact the perception of overall perceived quality, is not considered.

Therefore, to reduce the complexity, a subset of visual information is selected by choosing the most salient part of the image/video content. Since the artifacts in highly salient regions are more likely to attract the viewer, it is assumed that inclusion of VA into quality assessment might be very beneficial. While assessing the quality using VA, distortions in highly salient regions may be more annoying compared to the distortions in regions of low saliency.

A large number of low-level and high-level attributes may impact visual attention. Low-level attributes include colour, shape, motion of objects and size. High-level attributes are based on semantic information and include written text and faces [4].

2.4.1 Review of Visual Attention with Explicit Design Methods

These models aim to foresee the behaviour of human visual attention when viewing a scene. These models are limited to predict the objects and locations that humans focus on. The theory by Treisman and Gelade [22] inspired many VA models, the most widely used bottom-up VA model is based on the neuronal architecture of the visual system [4]. The top-down approach is focused only by a few models as it is less understood compared to that of the bottom-up approach.

Regions or objects which receive higher level of importance are referred to as regions-of-interest. Segmentation of the image or video frames into various regions of interest is performed in this model. Thus, a visual scene is divided into background and ROI. Then, each region is assigned a level of interest. The modelling and analysis of these models is implanted in different ways. Osberger et al. [23] validates the model using faze patterns recorded from eye tracking.

Pinneli and Chandler [24] proposed to rate the perceived levels of interest by a number of observers. All the proposed models infer that there is a correlation between ROI selections and eye movements. Therefore, Saliency Maps (SMs) are used to determine the ROI by defining appropriate thresholds. U. Engelke et al. [2] presented a saliency awareness model for video frames to consider
varying saliency of video frames. The proposed model for video quality metrics has considerably improved the quality prediction performance.

2.5 Motivation for Visual Attention based Quality Assessment

In spite of extensive works in the area of quality prediction metrics, the prediction accuracy of the metrics is limited. In the previous works, quality metrics compute quality scores considering that the whole content of scene is of equal interest to the observer. But it is observed that the VA consists of higher cognitive processing which reduces the complexity of analysis.

As VA is considered as one of the most important features of the HVS, it cannot be neglected while assessing the quality of visual scene. Considering the fact that humans focus on highly salient regions, the impact of distortions outside these regions has a negligible effect on overall perceived quality.

To determine the impact of VA on IQM, a high priority is given only to a sub-set of the visual information, i.e. to the ROI. Therefore, the distortions particularly in the ROI may affect the quality prediction of an image. This integration of visual saliency and perceptual distortion features in quality assessment may provide better prediction performance of metrics rather than the traditional IQM.

In this context, the metrics are measured independently over ROI and BG of an image to obtain quality measures. Then, the ROI and BG metrics are pooled together by applying respective weightage resulting in a new ROI aware quality metric. The quality metric chosen in this thesis is an FR metric. As higher amount of reference information promises better quality prediction performance. The FR metrics provide superior prediction performance compared to other metrics. The FR metrics chosen for this thesis are $VIF$ and $SSIM$. 


3.1 Overview of the Method Implemented

This section briefly explains the research design and steps implemented to obtain the desired data. The method of obtaining results is setup in a step-by-step process where the results obtained answer the research questions stated in Section 1.1.

3.1.1 Framework for FR Quality Assessment in a Wireless Communication System

A fundamental model of an FR visual quality assessment in a wireless communication system is presented in the Fig. 3.1. This model has been designed to provide an overview of steps involved in the process of achieving required data for analysis.

The above model is designed in such a way that it facilitates the deployment of the ROI aware metric. Firstly, it involves extraction of ROI of the source and test images. Then, the extracted ROI and BG images are given as a input to the FR quality assessment where the ROI and BG of test image is compared to that of the source image. Then a metric value is computed for the ROI and BG of test image. Finally, a pooling function is deployed to combine the ROI and BG
metrics producing an ROI aware image quality metric. An illustration of image quality assessment based on ROI is presented in the Section 3.1.4.

### 3.1.2 Reference and Distorted Images

The required reference and distorted images for experimenting are obtained from the Wireless Imaging Quality (WIQ) database [25] [26]. MOS of images obtained from the subjective tests are also provided in the WIQ database. The ROI coordinates of required images are obtained from the ROI database [27] [28].

The ROI of test images obtained from the ROI database is used to obtain mean ROI where mean ROI is considered for all the seven reference images. The mean ROI is computed by considering all the 30 ROI selections where mean of x and y coordinates are calculated separately.

The main reason behind the calculation of mean ROI is twofold. Firstly, almost all the ROI selections are overlapping and include each other. Secondly, to reduce the computational complexity and overhead while transmission of an image over the channel. The BG of images is obtained by setting the ROI of image to zero, i.e. by cutting out ROI area of the image.

### 3.1.3 Objective Metric Model

As we already know that the objective approach of quality assessment is automated, every metric is assessed using an algorithm. VIF and SSIM metrics are implemented in Matlab to estimate the distortion of an image.

The output of algorithm is a metric value where the inputs are reference image and distorted image. The estimated metric value indicates the distortion of distorted image on the scale of 0 to 1.

The process of estimating the metric value is implemented on three different categories where the metric value is computed over the whole region, the ROI and the BG. The ROI and BG of an image is presented in Fig. 3.2 to illustrate the metric computations.

- **Whole region**

  In this category, VIF is computed for whole region region of the distorted image. It is obtained by considering the whole region of reference image and distorted image as inputs to the VIF algorithm.
Chapter 3. Methodology

• ROI

Here, \( VIF_{ROI} \) is obtained by considering only ROI of reference and distorted images as inputs. The output metric value indicates the distortion concerning only ROI of the distorted image.

• BG

In this case, \( VIF_{BG} \) is estimated by considering only BG of reference and distorted images as inputs. The output value quantifies the distortion for BG of the distorted image.

Figure 3.2: Image sample ‘Elaine’ with ROI and BG (ROI: region inside black frame; BG: region outside black frame).

3.1.4 Evaluating ROI aware Objective Quality Metric

In this section, an ROI aware quality metric is obtained using a pooling function. In the pooling stage, the ROI and BG are combined to evaluate the ROI aware metric. It is to be noted that ROI and BG are weighted according to the pooling function employed to obtain a final ROI aware metric value.

The idea behind the pooling function is to apply weightage to the quality metrics computed independently based on the ROI and BG. The process of obtaining an ROI aware metric can be seen in Fig. 3.3.
The mathematical expression to compute the $VIF_{RA}$ is given by

$$VIF_{RA} = \alpha \cdot VIF_{ROI} + (1 - \alpha) \cdot VIF_{BG}$$  \hspace{1cm} (3.1)$$

where

- $VIF$ = Metric computed for whole image.
- $VIF_{ROI}$ = $VIF$ metric computed for ROI.
- $VIF_{BG}$ = $VIF$ metric computed for BG.
- $VIF_{RA}$ = ROI aware $VIF$ metric.
- $\alpha$ = Perceptual metric weight.

In the above expression, the prediction performance of the ROI aware metric, $VIF_{RA}$ is observed by varying $\alpha$ from -1 to 1. The value of $\alpha$ regulates the weight of $VIF_{ROI}$ and $VIF_{BG}$, thereby changing the contribution of respective metric to the final value. With respect to our assumption that artifacts in ROI maybe more annoying compared to that of artifacts in BG, the value of $\alpha$ would have a value greater than 0.5.

Similarly, an ROI aware metric for SSIM can be derived as follows:

$$SSIM_{RA} = \alpha \cdot SSIM_{ROI} + (1 - \alpha) \cdot SSIM_{BG}$$  \hspace{1cm} (3.2)$$

A comprehensive performance analysis of the $VIF_{RA}$ and $SSIM_{RA}$ metrics is performed to find out an optimal value of $\alpha$.

### 3.2 Perceptual Metric Weight Optimisation

The optimisation framework consists of evaluating the $VIF_{RA}$ metric based on performance indicators, analysing and obtaining an optimal perceptual relevance weight. The performance indicators for this analysis are Root Mean Squared Error (RMSE), Pearson linear correlation $P_{corr}$ and Spearman rank order correlation $S_{corr}$.

RMSE is a measure of the differences between values predicted by a model and the observed values. It serves to aggregate the effect of errors in estimations
into a single measure. In other terms, one can say that the prediction accuracy is determined using the RMSE. The RMSE is given by

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]  \hspace{1cm} (3.3)

where \(y_i\) is the observed value, \(\hat{y}_i\) is the predicted value and \(n\) is the number of observations.

Pearson linear correlation \(P_{corr}\) is a measure of linear dependence between two variables. The correlation value is between +1 and -1, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. The prediction accuracy as recommended in [29] is determined using the Pearson correlation. \(P_{corr}\) can be computed as

\[
P_{corr} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]  \hspace{1cm} (3.4)

where \(x_i\) and \(y_i\) are the \(i\)th values of data sets \(x\) and \(y\), respectively. The variables \(\bar{x}\) and \(\bar{y}\) are means of the respective data sets containing \(n\) values.

Spearman rank order correlation \(S_{corr}\) is a measure of rank correlation between two variables which indicates the statistical dependence between ranks of two variables. In other words, it assesses the monotonicity between two variables. Hence the prediction monotonicity is measured by the Spearman correlation. \(S_{corr}\) is given as

\[
S_{corr} = \frac{\sum_{i=1}^{K} (U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^{n} (U_i - \bar{U})^2} \sqrt{\sum_{i=1}^{n} (V_i - \bar{V})^2}}
\]  \hspace{1cm} (3.5)

where \(U_i\) and \(V_i\) are ranks of the data sets \(U\) and \(V\), respectively. The variables \(\bar{U}\) and \(\bar{V}\) are midranks of the respective data sets containing \(K\) values.

The goal of optimising the perceptual metric weight is attained after thorough examination of the performance indicators mentioned above. To obtain an optimal metric weight, \(\alpha\) is varied between -1 and 1 to observe the variation in the performance of quality metrics, \(VIF\) and \(VIF_{RA}\).

To evaluate the performance of the metric, \(PMOS\) and \(PMOS_{RA}\) are computed using a mapping function. Here, \(PMOS\) is the Predicted Mean Opinion Score of \(VIF\) and \(PMOS_{RA}\) is the Predicted Mean Opinion Score of \(VIF_{RA}\). A comprehensive analysis is performed to find out a suitable mapping function to map \(VIF\) and \(VIF_{RA}\) onto \(PMOS\) and \(PMOS_{RA}\), respectively.
3.2.1 Mapping Function

The purpose of mapping function is to map the image quality metrics onto the subjective mean opinion scores obtained from the WIQ database. This prediction function maps the range of quality metric onto the range of subjective MOS, thereby serving the purpose of predicting the MOS. Due to the non-linear processing in the HVS, the relationship between a quality metric and MOS does not follow a linear relationship. Thus, the mapping function is a function of the quality metric which can be represented as follows:

\[ MOS_\theta = f(\theta) \]  

(3.6)

where \( \theta \) is the image quality metric.

Various functions like polynomial, exponential, logistic and power functions are considered to satisfy the non-linear relation between quality metric and MOS. The function parameters are determined using the image quality metric data and MOS of the distorted images. The quality metric data of images is computed using the VIF and SSIM algorithms. The required MOS of images is obtained from the WIQ database.

The Matlab Curve Fitting Toolbox is used to deduce parameters of the mapping function. The Matlab Toolbox specifies the parameters which describe the best relationship between MOS and quality metric. VIF and VIF\(_{RA}\) are considered separately as image quality metric while mapping a function onto MOS. Similarly, the curve fitting toolbox is also applied to SSIM and SSIM\(_{RA}\) for mapping them onto MOS.

Fig. 3.4 and Fig. 3.5 shows the curve fitting of VIF and VIF\(_{RA}\), respectively. After visual examination of various curve fitting functions, a favourable fitting is obtained with an exponential function. The performance of the exponential function is also confirmed with the goodness of fit measures. In Fig. 3.6 and Fig. 3.7, curve fittings of SSIM and SSIM\(_{RA}\) are presented, respectively. Visual inspection along with goodness of fit measures showed that a power function is the best performing mapping function for SSIM and SSIM\(_{RA}\).
Chapter 3. Methodology

Figure 3.4: Curve fitting of $VIF$ for exponential function.

Figure 3.5: Curve fitting of $VIF_{RA}$ for exponential function.
Chapter 3. Methodology

Figure 3.6: Curve fitting of $SSIM$ for power function.

Figure 3.7: Curve fitting of $SSIM_{RA}$ for power function.
Chapter 3. Methodology

The function chosen for curve fitting of VIF and VIFRA is given as

\[ f(\theta) = a \cdot e^{b \theta} \]  

(3.7)

where a and b are constants. The resultant mapping functions of VIF and VIFRA after applying curve fitting are

\[
PMOS = 23.8235 \cdot e^{(1.2289 \cdot VIF)}
\]  

(3.8)

\[
PMOS_{RA} = 23.9015 \cdot e^{(1.2325 \cdot VIF_{RA})}
\]  

(3.9)

Similarly, the curve fitting function for SSIM and SSIMRA is given as

\[ f(\theta) = a \cdot \theta^b + c \]  

(3.10)

where a, b and c are constants. The resultant mapping functions of SSIM and SSIMRA obtained in curve fitting are

\[
PMOS = 55.78 \cdot \theta^{5.87} + 25.34
\]  

(3.11)

\[
PMOS_{RA} = 52.7 \cdot \theta^{13.41} + 32.28
\]  

(3.12)

The goodness of fit between metric value and MOS is identified by measures such as RMSE and Sum of Squared Errors (SSE). To understand the performance of mapping functions, the curves are plotted with 95% Confidence Interval (CI).

The RMSE and SSE measures of both SSIM and SSIMRA are approximately around 17.92 and 1.9 \times 10^4, respectively. In the cases of VIF and VIFRA, the RMSE is less than 15.56 and the SSE is less than 1.7 \times 10^4. After analysing the performance of the metrics, it is observed that VIF and VIFRA performs better than SSIM and SSIMRA. Therefore, only VIF and VIFRA are considered for further analysing the impact of VA on prediction performance.

An optimization procedure for \( \alpha \) is performed with the help of performance indicators RMSE, \( P_{corr} \) and \( S_{corr} \). The perceptual weight \( \alpha \) is varied between -1 and 1 to observe its performance where the RMSE, \( P_{corr} \) and \( S_{corr} \) are computed between predicted scores and subjective scores. In Section 4.1, the plots of RMSE versus \( \alpha \), \( P_{corr} \) versus \( \alpha \) and \( S_{corr} \) versus \( \alpha \) are examined to optimize the perceptual weight, \( \alpha \).
3.3 Generation of Distorted Images

A set of seven reference images are chosen to cover a variety of contents, complexities, and textures. All the images are of dimension $512 \times 512$ pixels represented in grey scale. Fig. 3.8 shows the reference images utilised to generate a set of eighty distorted images in the WIQ database. The distorted images of WIQ database are obtained by transmitting the reference images over a wireless Rayleigh fading channel.

Figure 3.8: Reference images used for analysis.
### 3.3.1 Wireless Channel Modelling

This section describes the wireless channel utilized to obtain the distorted images of WIQ database. It is a two-step process where the first step involves simulating a wireless radio channel and the second step is to transmit the reference images through the wireless channel.

In order to create a wireless radio channel, an uncorrelated flat Rayleigh fading was implemented [4]. Fading along with Additive White Gaussian Noise (AWGN) was implemented to simulate severe transmission conditions. As the Rayleigh fading model simulates the non-line of sight scattering of urban environment, it was chosen for fading the wireless channel.

The fading model was implemented using Jakes model, based on summing sinusoids. In this model, \(N\) equal strength rays arrive at the receiver with uniformly distributed angles [4]. As a receiver in motion was considered, each ray has a Doppler shift. To produce real time wireless transmission conditions, the Signal-to-Noise Ratio (SNR) was varied between zero and forty decibels.

The Joint Photographic Experts Group (JPEG) image format was chosen to source encode the reference images before transmission through the channel [4]. Impairments may occur while source encoding the images itself as JPEG implements a block Discrete Cosine Transform (DCT) algorithm. A Bose-Chaudhuri-Hocquenghem (BCH) code was used for error protection purposes. The images were modulated using Binary Phase Shift Keying (BPSK) for transmission.

The artifacts observed in the distorted images are described as follows:

- **Ringing**

An image with ringing is shown in Fig. 3.9 (a). This type of impairment appears as periodic pseudo edges near sharp transitions of the image segments. It occurs due to the loss of precision in high frequency components. It occurs particularly during JPEG encoding as it splits image into \(8 \times 8\) blocks during DCT. Ringing may also occur in JPEG-2000 which is a wavelet based source encoding.

- **Blocking**

It occurs particularly in any block-based coding scheme which causes artifacts in pixel blocks and block boundaries. This type of distortion is presented in Fig. 3.10 (a) and 3.10 (b). Blocking distortion is inherent with the compression techniques like JPEG source encoding. It appears as surface discontinuity at the pixel block boundaries. As the JPEG encoding is based on DCT and quantization of blocks, an error in quantization of blocks causes low-resolution blocks. This effect is particularly visible in flat areas, where there is little detail to reduce the blocking effect.
Chapter 3. Methodology

Figure 3.9: Distorted images: (a) Elaine with ringing, (b) Goldhill with blurring.

Figure 3.10: Distorted images: (a) Barbara with blocking, (b) Lena with blocking and ringing.
• **Blurring**

This impairment is a result of Gaussian smoothing, an effect to reduce the image noise. It occurs as the semantic information about the shapes and details of objects is lost in the process of applying Gaussian function. Blurring is also a consequence of the quantisation errors of frequency components introduced in JPEG source encoding. It may also occur in JPEG2000 encoding which is implemented based on Discrete Wavelet Transform (DWT). Blurring is usually prevalent within $8 \times 8$ blocks compared to that on a global scale. It can be observed in the image shown in Fig. 3.9 (b).
Chapter 4

Results and Analysis

To evaluate the prediction performance of the ROI aware metric, $VIF_{RA}$, the analysis is segmented into three sections. Firstly, the plots of RMSE, $P_{corr}$ and $S_{corr}$ against $\alpha$ are examined to derive an optimized perceptual weight. Then, a brief comparative analysis is performed to establish a better image quality metric. Lastly, a comparative analysis based on the scale of distortion is discussed. All the required data for the analysis is acquired from the reference and distorted images of the WIQ database [26].

4.1 Perceptual Weight Optimization

The necessary plots of performance indicators for optimization are presented in Fig. 4.1 and 4.2 to discuss and arrive at an $\alpha$ value. The value of $\alpha$ in (3.1) is chosen in such a way that it achieves best possible performance with all the performance indicators.

Firstly, in Fig. 4.1, the relation between $\alpha$ and RMSE is studied and an $\alpha$ value is identified to attain minimal RMSE between MOS and $PMOS_{RA}$. In Fig. 4.1, it is observed that as the value of $\alpha$ increases from -1 to 1, the RMSE between MOS and $PMOS_{RA}$ decreases. This phenomenon indicates that as the weightage of ROI metric increases, the error between predicted MOS, $PMOS_{RA}$ and subjective score, MOS is reduced. Increasing the contribution of $VIF_{ROI}$ compared to the background metric, $VIF_{BG}$ improves the prediction performance of the ROI aware metric, $VIF_{RA}$. Therefore, the performance results of RMSE as a function of $\alpha$ demonstrate that a minimum error is achieved when $\alpha$ is 0.88.

As has already been mentioned in Chapter 3, correlation between $PMOS_{RA}$ and MOS represents the performance of $VIF_{RA}$. With this conjecture, we can observe in Fig. 4.2 that as the $\alpha$ value increases the performance of $P_{corr}$ and $S_{corr}$ increases.
Figure 4.1: RMSE as a function of $\alpha$.

Figure 4.2: Pearson and Spearman correlations as a function of $\alpha$. 
In Fig. 4.2, it is observed that $P_{corr}$ increases as the value of $\alpha$ is increases from -1 to 1. A maximum $P_{corr}$ is found to occur at 0.82 and then it decreases slightly. A similar behaviour is observed in the case of Spearman correlation where a maximum correlation is achieved when $\alpha$ is 0.84.

The results from the three plots show that better prediction performance is achieved when $\alpha$ is greater than 0.5. This result itself shows that a greater perceptual weightage of ROI leads to better prediction performance of the image quality metric. These observations are in accordance with the initial conjecture that artifacts in the ROI are perceived to be more annoying than that of the background.

Finally, a consolidated $\alpha$ value is obtained by taking a mean of the best $\alpha$ values obtained for the three cases. The results from the performance indicators allow us to draw an optimised $\alpha$ value of 0.84. The obtained $\alpha$ is used to calculate the $VIF_{RA}$ thereby obtaining the predicted score, $PMOS_{RA}$ using a mapping function. Further analysis of the performance of the ROI aware metric is demonstrated in the following section.

### 4.2 Comparative Analysis based on Predicted MOS

This section provides an overview of the performance of the image quality metrics based on the predicted mean opinion scores, $PMOS$ and $PMOS_{RA}$. A better metric among the $VIF$ and $VIF_{RA}$ is determined in terms of prediction accuracy and monotonicity.

Table 4.1: Comparison of $VIF$ and $VIF_{RA}$ over all distorted images

<table>
<thead>
<tr>
<th>Metric</th>
<th>$\alpha$</th>
<th>RMSE</th>
<th>$P_{corr}$</th>
<th>$S_{corr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIF</td>
<td>N/A</td>
<td>15.5679</td>
<td>0.7135</td>
<td>0.6918</td>
</tr>
<tr>
<td>$VIF_{RA}$</td>
<td>0.84</td>
<td>14.7909</td>
<td>0.7638</td>
<td>0.7211</td>
</tr>
</tbody>
</table>

Table 4.1 presents the performance of $VIF_{RA}$ and $VIF$ in terms of RMSE, $P_{corr}$ and $S_{corr}$. These results are obtained considering all the 80 distorted images of the reference images. It can be observed that the RMSE of the ROI aware metric is lower compared to that of the original metric, $VIF$. Similarly, a higher correlation is observed in the case of $VIF_{RA}$ than its counterpart.

In Table 4.1, it can be seen that the RMSE is lower in the case of $VIF_{RA}$. We can also observe that $P_{corr}$, $S_{corr}$ are higher in the case of $VIF_{RA}$ compared to $VIF$. Therefore, the prediction accuracy and prediction monotonicity has been improved with $VIF_{RA}$. 
Chapter 4. Results and Analysis

We can observe that there is an improvement of about 4-5% in the case of $VIF_{RA}$ compared to $VIF$. This improvement is consistent for all the performance indicators of $VIF_{RA}$. The enhanced performance of $VIF_{RA}$ with all the performance indicators illustrates that the ROI aware metric performs better compared to $VIF$. To further examine the performance of metrics, a comprehensive analysis of metrics on each reference image and analysis based on the scale of distortion and location of artifacts is demonstrated in the following sections.

4.2.1 Analysis on Each Reference Image

In this section, the performance of $VIF_{RA}$ is studied on each reference image based on the performance indicators. This analysis helps us to observe the behavior of $VIF_{RA}$ on various images with different ROI and BG. This analysis is performed by taking into account several distorted images of the respective reference image. Then, the performance of $VIF_{RA}$ is compared to the performance of $VIF$ to identify the better metric.

Table 4.2: Comparison of metrics based on individual reference images

<table>
<thead>
<tr>
<th>IMAGE</th>
<th>METRIC</th>
<th>RMSE</th>
<th>Pcorr</th>
<th>Scorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara</td>
<td>$VIF$</td>
<td>13.0012</td>
<td>0.9478</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>9.4605</td>
<td>0.9530</td>
<td>0.98</td>
</tr>
<tr>
<td>Goldhill</td>
<td>$VIF$</td>
<td>13.9425</td>
<td>0.9039</td>
<td>0.8667</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>13.3370</td>
<td>0.9240</td>
<td>0.8833</td>
</tr>
<tr>
<td>Lena</td>
<td>$VIF$</td>
<td>16.4490</td>
<td>0.8798</td>
<td>0.8878</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>15.6725</td>
<td>0.8870</td>
<td>0.9034</td>
</tr>
<tr>
<td>Elaine</td>
<td>$VIF$</td>
<td>9.2027</td>
<td>0.82</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>8.3251</td>
<td>0.8345</td>
<td>0.6571</td>
</tr>
<tr>
<td>Mandrill</td>
<td>$VIF$</td>
<td>22.1587</td>
<td>0.6990</td>
<td>0.6636</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>18.2853</td>
<td>0.7006</td>
<td>0.7909</td>
</tr>
<tr>
<td>Peppers</td>
<td>$VIF$</td>
<td>10.9635</td>
<td>0.6722</td>
<td>0.4799</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>10.9056</td>
<td>0.7245</td>
<td>0.7810</td>
</tr>
<tr>
<td>Tiffany</td>
<td>$VIF$</td>
<td>15.0608</td>
<td>0.7801</td>
<td>0.8297</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>14.9931</td>
<td>0.7854</td>
<td>0.8681</td>
</tr>
</tbody>
</table>

In Table 4.2, a similar behavior of $VIF_{RA}$ can be observed for all reference images. The RMSE is lower for $VIF_{RA}$ compared to $VIF$ in all the reference images. Also, the correlations $P_{corr}$ and $S_{corr}$ are higher for $VIF_{RA}$ compared to $VIF$ in all cases.

After the analysis of ROI aware metric individually on various reference images, we can observe that in all the cases there is an improvement in quality.
prediction with the $VIF_{RA}$ compared to that of the $VIF$. The detailed analysis on each reference image with various distortions lead to the importance of visual attention in the quality metric design. It also allows us to generalize the impact of visual attention on all the types of images based on the results observed. Thus, the ROI aware metric provides a better prediction performance over the existing prediction techniques in practice. With the improved performance of prediction in images, the ROI aware prediction technique can also be applied on video applications to measure the video quality.

### 4.2.2 Analysis based on Location of Artifacts

In this section, we take a closer look on the performance of metrics with the help of analysis based on location of artifacts. For this purpose, reference image ‘Lena’ is taken as an example for analysing the impact of artifact’s location. To understand the impact of ROI on the image quality and quality prediction, it is necessary to observe the results in this section.

In Fig. 4.3, the images mainly contain artifacts within ROI whereas in Fig. 4.4, the images exhibit artifacts primarily in the BG. In Fig 4.3(a), a row of distortion is observed across the face of Lena whereas in Fig 4.3(b), an artifact is located beside the nose of Lena.

In Fig 4.4(a), a distortion is observed on the left corner just above the hat of Lena. In Fig 4.4(b), rows of distortion can be observed on the right corner of the image along with small distortions at lower left corner in the hair of Lena.

These images contain similar amount of distortion either in ROI or BG. Thus, it enables us to examine the impact of ROI on quality prediction. The prediction performance of the metrics is evaluated based on the respective predicted mean opinion scores and subjective scores of the images.

Table 4.3: Comparison of the ROI aware quality metric and original metric for Lena with artifacts in ROI

<table>
<thead>
<tr>
<th>Image</th>
<th>Metric</th>
<th>Distortion</th>
<th>Predicted MOS</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 4.3(a)</td>
<td>VIF</td>
<td>0.9537</td>
<td>77.7069</td>
<td>71.6333</td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>0.9139</td>
<td>73.7837</td>
<td></td>
</tr>
<tr>
<td>Fig. 4.3(b)</td>
<td>VIF</td>
<td>0.8836</td>
<td>71.0225</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VIF_{RA}$</td>
<td>0.8208</td>
<td>66.7477</td>
<td>49.6</td>
</tr>
</tbody>
</table>
Chapter 4. Results and Analysis

Figure 4.3: Lena with artifacts in ROI.

Figure 4.4: Lena with artifacts in BG.
Table 4.4: Comparison of the ROI aware quality metric and original metric for Lena with artifacts in BG

<table>
<thead>
<tr>
<th>Image</th>
<th>Metric</th>
<th>Distortion</th>
<th>Predicted MOS</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 4.4(a)</td>
<td>VIF</td>
<td>0.9918</td>
<td>81.1489</td>
<td>91.3</td>
</tr>
<tr>
<td></td>
<td>VIF(_RA)</td>
<td>0.9963</td>
<td>81.5335</td>
<td></td>
</tr>
<tr>
<td>Fig. 4.4(b)</td>
<td>VIF</td>
<td>0.9107</td>
<td>73.4373</td>
<td>59.8333</td>
</tr>
<tr>
<td></td>
<td>VIF(_RA)</td>
<td>0.8710</td>
<td>70.7529</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4.3(b) and 4.4(a) show images with similar scale of distortions. In Fig 4.3(b), a small speckle can be observed in the ROI whereas in Fig. 4.4(a), a small speckle is in the BG of image. From the Tables 4.3 and 4.4, a significant difference of subjective scores can be seen in the considered cases even if the distortion is similar. The MOS of Fig. 4.3(b) is lower compared to that of the Fig. 4.4(a) as the distortion is in the ROI which is annoying compared to the distortion in the BG.

Another interesting characteristic can be observed in Fig. 4.3(a) and 4.4(b), where we can locate rows of distortions in ROI and BG, respectively. From Tables 4.3 and 4.4, we can observe that the MOS is lower in Fig. 4.4(b) compared to Fig. 4.3(a). But, a greater quality degradation is expected when distortion is in ROI. This phenomenon is due to the several rows of distortions located in Fig 4.4(b) along with small distortions near the hair which is in the ROI of image. From a pure visibility point of view, these small distortions in the ROI of Fig 4.4(b) might have affected the MOS.

From Tables 4.3 and 4.4, an improvement can be observed in predicted MOS with VIF\(_RA\) compared to that of the VIF. The improvement is consistent in both tables which indicate a superior quality prediction of VIF\(_RA\) metric compared to the VIF.

### 4.3 Analysis based on Scale of Distortion

In this section, the performance of the VIF\(_RA\) is examined based on the scale of distortion in the image. This analysis is primarily done to examine the impact of size of distortions on the quality prediction.

For this study, the distorted images are segregated based on the distortion scale. The distortions which are primarily spread in less than fifty 8 × 8 blocks and limited to either ROI or BG region of an image is categorized as just-noticeable distortions. Whereas, the distortions which are predominantly spread in more than fifty 8 × 8 blocks and are located in both ROI and BG regions are termed as large scale distortions. Images with large scale and small scale distortions utilized for this analysis are presented in Appendix A and B.

To study the prediction performance of VIF\(_RA\) based on the scale of distor-
tions, Table 4.5 presents the performance indicators for images with large scale distortions.

Table 4.5: Performance of quality metrics in the case of images with large scale distortions

<table>
<thead>
<tr>
<th>Metric</th>
<th>RMSE</th>
<th>Pcorr</th>
<th>Scorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIF</td>
<td>15.0661</td>
<td>0.6438</td>
<td>0.6633</td>
</tr>
<tr>
<td>VIF$_{RA}$</td>
<td>14.6061</td>
<td>0.6690</td>
<td>0.6695</td>
</tr>
</tbody>
</table>

From Table 4.5, it can be observed that the RMSE is lower in the case of VIF$_{RA}$. We can also observe that the correlations $P_{corr}$, $S_{corr}$ are higher in the case of VIF$_{RA}$ compared to the VIF. From the table of results, we can judge that the performance of the VIF$_{RA}$ metric is comparable to VIF. In fact, the RMSE is better in the case of VIF rather than VIF$_{RA}$ and we can observe an improvement of about 2% in $P_{corr}$ of VIF$_{RA}$. Whereas, in the case of $S_{corr}$ the performance of VIF$_{RA}$ is almost equal to VIF. Finally, the results demonstrate that VIF$_{RA}$ shows no significant improvement in prediction performance compared to that of original metric, VIF.

If the distortions predominantly cover a large area of the image, then it might cause a greater annoyance to the viewer. As the distortions are widespread over the ROI and BG, even if we consider visual attention in these images, it might not have a significant impact on the overall quality prediction. This phenomenon is evident from Table 4.2 where there is no difference in performance of VIF$_{RA}$ and VIF.

Table 4.6: Performance of quality metrics in the case of images with just-noticeable distortions

<table>
<thead>
<tr>
<th>Metric</th>
<th>RMSE</th>
<th>Pcorr</th>
<th>Scorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIF</td>
<td>16.1372</td>
<td>0.7887</td>
<td>0.7077</td>
</tr>
<tr>
<td>VIF$_{RA}$</td>
<td>14.0204</td>
<td>0.8412</td>
<td>0.7954</td>
</tr>
</tbody>
</table>

In Table 4.6, the RMSE for VIF$_{RA}$ shows an improvement of about 2% compared to the VIF. Therefore, it can be stated the prediction accuracy of VIF$_{RA}$ is better than that of the VIF metric. In particular, the performance of correlations $P_{corr}$ and $S_{corr}$ could be increased by 10% with the ROI aware metric,
The results show that prediction accuracy and monotonicity could be significantly improved with the help of $VIF_{RA}$.

The results in this case demonstrate the impact of visual attention on quality prediction performance. As the distortions in this scenario are either spread in ROI or BG and are small scale distortions, it cause higher impact on performance of $VIF_{RA}$. If these just-noticeable distortions are in the ROI then it results in greater quality degradation which is evident from Section 4.2.2. But if these distortions are in the BG, it has a little impact on quality of an image. Therefore, the effect of visual attention is predominant in this scenario which results in better performance of $VIF_{RA}$ compared to $VIF$. 
Chapter 5

Summary

With the advent of communication systems, the transmitted data suffers from impairments through source coding and transmission over error prone channels. The significant demand for a good quality of multimedia is a major challenge for service providers in communication systems. With the existing resource constraints of image and video communication systems, it is necessary to develop an accurate objective metric design for assessing the QoE of the end user.

This thesis aims at improving the quality prediction performance of a full-reference objective quality metric. A FR quality metric called Visual Information Fidelity (VIF) is specifically chosen for measuring the quality of images. As the performance of this metric is higher and complexity is lower compared to other metrics, it is used for the quality measurement.

A new technique is implemented in this thesis to further improve the performance of the VIF metric. A set of seven reference images and eighty distorted images are used to analyse the performance of this technique. An ROI aware metric, $VIF_{RA}$ is proposed for predicting the quality of an image which is based on pooling the weighted ROI and BG metrics.

Specifically, an optimised perceptual weight is used for pooling ROI and BG metrics to get $VIF_{RA}$. The perceptual weight is consolidated by evaluating the performance of $VIF_{RA}$ based on the performance indicators such as RMSE, $P_{corr}$, $S_{corr}$. A mapping function is derived in Matlab to predict the mean opinion scores of images. Then, the predicted scores are compared to the mean opinion scores of images to assess the performance of a metric.

To assess the prediction performance, $VIF_{RA}$ is compared to VIF with the help of performance indicators. Finally, the distorted images are segregated based on location of artifacts and scale of distortions to further study the performance of $VIF_{RA}$ and VIF.

After thorough analysis of results, the findings of the thesis are as follows:

1. The consolidated $\alpha$ value is 0.84 which is in accordance with the initial conjecture that artifacts in the ROI are perceived to be more annoying than that of the background.

2. A comparative analysis in Section 4.2 confirms the better prediction performance of $VIF_{RA}$ than VIF.
3. The superior performance of $VIF_{RA}$ is understood with a separate analysis on each reference image and with a study based on location of artifacts.

4. Although, a better performance of $VIF_{RA}$ than $VIF$ is seen for images with large scale distortions. It is observed that the prediction performance of $VIF_{RA}$ over $VIF$ is significant, particularly for the images with just-noticeable distortions.

5. After a brief analysis of $VIF_{RA}$’s performance, it is confirmed that the VA contributes significantly towards a better quality prediction.

### 5.1 Limitations and Future work

The assessment of QoE of the end user is a complex process as it must oblige the behaviour of the human visual system. Incorporating all attributes and parameters to define a perceptual quality metric would derive an extremely complex metric. However, with the advancement in computational capabilities, it is possible to further improve the prediction accuracy of the objective quality metrics.

The proposed technique is performed over a full-reference objective metric. Hence, the quality prediction can be performed only when the reference information is available. The method of perceptual relevance weights proposed in this thesis is studied only to a set of images available from a database. Hence, the scope of this thesis is limited only to a set of available distortions. Also, the ROI aware metric derived performs better only with the just-noticeable/small scale distortions. Hence more research should be done with large scale distortions.

A research using the proposed model could be performed over video processing applications.

To further improve the prediction performance over a wide variety of distortions, extensive research over vast number of images with several distortions could be done. The thesis work is performed by transmitting the ROI and BG information over a single strong link. A new study could be done on the basis of link adaptation and various relay structures. Further analysis of the visual attention could be done by considering various modulation schemes and fading channels.

### 5.2 Conclusions

The tremendous deployment of quality assessment in image processing applications insist the development of reliable metric design. Despite the efforts to develop a reliable metric, there are no quality metrics that are applicable under wide range of scenarios. This thesis work demonstrates the importance of VA in the quality metric design. A new model to measure the quality of an image using an FR metric, $VIF$ is proposed. An ROI aware metric, $VIF_{RA}$ is designed
using a pooling function which is based on perceptual weightage of ROI and BG metrics.

The proposed model proved to be beneficial for better quality prediction compared to the VIF approach in practice. The improvement in prediction is around 5% with the help of $VIF_{RA}$. However, an increase of about 10% in prediction performance of $VIF_{RA}$ over $VIF$ is observed only when the distortions are just noticeable.

The results consistently support the impact of VA in image quality assessment. But, these results would not support the work of [30], where it reports that an improvement in quality assessment could not be established with VA. Although, the results of this thesis agree with the importance of VA, further studies are required to be done on various image quality metrics to understand the limits of metric pooling in VA. In spite of the enhanced performance of prediction with this model, a large scope for improvement exists in the field of images and videos over a broad range of distortions.
Appendices
Appendix A

Images with Large Scale Distortions

The images with large scale distortions which have been used for the analysis in Section 4.3 are shown below.

Figure A.1  Figure A.2  Figure A.3

Figure A.4  Figure A.5  Figure A.6
Appendix A. Images with Large Scale Distortions

Figure A.7  
Figure A.8  
Figure A.9  

Figure A.10  
Figure A.11  
Figure A.12  

Figure A.13  
Figure A.14  
Figure A.15
Appendix A. Images with Large Scale Distortions

Figure A.34
Figure A.35
Figure A.36

Figure A.37
Figure A.38
Figure A.39

Figure A.40
Figure A.41
Figure A.42
Appendix A. Images with Large Scale Distortions

Figure A.43

Figure A.44

Figure A.45

Figure A.46
Appendix B

Images with Just-Noticeable Distortions

The images with just-noticeable distortions which have been used for the analysis in Section 4.3 are shown below.

Figure B.1  Figure B.2  Figure B.3

Figure B.4  Figure B.5  Figure B.6
Appendix B. Images with Just-Noticeable Distortions

Figure B.7

Figure B.8

Figure B.9

Figure B.10

Figure B.11

Figure B.12

Figure B.13

Figure B.14

Figure B.15
Appendix B. Images with Just-Noticeable Distortions

Figure B.16

Figure B.17

Figure B.18

Figure B.19

Figure B.20

Figure B.21

Figure B.22

Figure B.23

Figure B.24
Appendix B. Images with Just-Noticeable Distortions

Figure B.34
References


References


References


