Thesis no: MSCS-2016-18

Empirical Study on Quantitative Measurement Methods for Big Image Data

An Experiment using five quantitative methods

Ramya Sravanam

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona Sweden

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:
Author(s):
Ramya Sravanam
E-mail: rasr15@student.bth.se

University advisor:
Dr. Huseyin Kusetogullari
Post- Doctoral Researcher
Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

ABSTRACT

Context. With the increasing demand for image processing applications in multimedia applications, the importance for research on image quality assessment subject has received great interest. While the goal of Image Quality Assessment is to find the efficient Image Quality Metrics that are closely relative to human visual perception, from the last three decades much effort has been put by the researchers and numerous papers and literature has been developed with emerging Image Quality Assessment techniques. In this regard, emphasis is given to Full-Reference Image Quality Assessment research where analysis of quality measurement algorithms is done based on the referenced original image as that is much closer to perceptual visual quality.

Objectives. In this thesis we investigate five mostly used Image Quality Metrics which were selected (which includes Peak Signal to Noise Ratio (PSNR), Structural SIMilarity Index (SSIM), Feature SIMilarity Index (FSIM), Visual Saliency Index (VSI), Universal Quality Index (UQI)) to perform an experiment on a chosen image dataset (of images with different types of distortions due to different image processing applications) and find the most efficient one with respect to the dataset used. This research analysis could possibly be helpful to researchers working on big image data projects where selection of an appropriate Image Quality Metric is of major significance. Our study details the use of dataset taken and the experimental results where the image set highly influences the results.

Methods. The goal of this study is achieved by conducting a Literature Review to investigate the existing Image Quality Assessment research and Image Quality Metrics and by performing an experiment. The image dataset used in the experiment is prepared by obtaining the database from LIVE Image Quality Assessment database. Matlab software engine was used to experiment for image processing applications. Descriptive analysis (includes statistical analysis) was employed to analyze the results obtained from the experiment.

Results. For the distortion types involved (JPEG 2000, JPEG compression, White Gaussian Noise, Gaussian Blur) SSIM was efficient to measure the image quality after distortion for JPEG 2000 compressed and white Gaussian noise images and PSNR was efficient for JPEG compression and Gaussian blur images with respect to the original image.

Conclusions. From this study it is evident that SSIM and PSNR are efficient in Image Quality Assessment for the dataset used. Also, that the level of distortions in the image dataset highly influences the results, where in our case SSIM and PSNR perform efficiently for the used database.

Keywords: Image Quality Assessment (IQA), Image Quality Metric (IQM), LIVE database.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to my supervisor **Dr. Huseyin Kusetogullari** for the support, guidance and encouragement through my thesis. If not for his immense help and feedback during writing my thesis, this thesis would not have been completed.

I thank my parents Mr. S. Ganapathi Rao and Ms. S. L. B. Sarojini and my brother Ravi for their unconditional love and support throughout my thesis. Finally, I thank my friends and well-wishers for their suggestions.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. IQA- Image Quality assessment
2. IQM- Image quality Metric
3. HVS- Human Visual System
4. FR IQA- Full- Reference Image Quality Assessment
5. NR IQA- No- Reference Image Quality Assessment
6. RR IQA- Reduced- Reference Image quality Assessment
7. MSE- Mean Squared Error
8. PSNR- Peak Signal to Noise Ratio
9. UQI- Universal Image Quality Index
10. SSIM- Structural SIMilarity
11. FSIM- Feature SIMilarity
12. VSI- Visual Saliency Index
13. VIFC- Visual Information Criterion
14. JPEG- Joint Photographic Experts Group
15. DMOS- Differential Mean Opinion Score
16. LIVE- Laboratory of Image and Video Engineering
17. CSIQ- Categorical Subjective Image Quality
18. TID- Tampere Image Database

LIST OF EQUATIONS

1 INTRODUCTION

An image is information that is stored visually. Processing an image is required to meet the demands of a particular situation. In scenarios where multimedia communication is of great significance, storing huge image databases and transmitting big sized image data has been challenging from decades. As a result, images need to be compressed (processed) in order to transmit images from a source to destination within the storage limits. This is called image compression. Image processing involves processes such as image compression to meet the demands of storage limitations [1]. Image enhancement is the type of image processing which basically deals with improving the quality of a processed image. Enhancement, noise removal, feature detection are few other image processing applications [1, 2, 3]. The field of image processing involves three main areas [4]. First, image coding to compress and transmit image efficiently. Second, pattern recognition to analyze the image and extract the required information. Third area is related to process of enhancement to improve the image quality for human access. Image compression, contrast enhancement, recognition are few among the challenging problems of image processing [2]. Noise reduction is another major issue with image processing. During the steps of image processing, the image is affected by several kinds of noise. These may be caused because of imperfect image capturing devices or noise in the capturing location.

1.1 Importance of quality assessment in image processing

Image quality assessment is of major importance in the applications of image processing [5]. Digital images are subjected to various kinds of distortions during the time of image acquisition, restoration, enhancement, compression or transmission [6]. These distortions may be due to different image processing applications. These image processing methods result in the disadvantage of degrading the quality of the image after processing. The observer needs to know the accurateness of the image obtained so that image restoration techniques can be applied to retain the maximum quality of the image. Hence, there is a need to assess the image quality degradation due to several image processing problems. The image degradation has to be measured and made known in order to make sure that it can be limited within an acceptable range of values [7]. That is, although the quality of an image is degraded due to an image processing application, the image must be acceptable by the Human Visual System (HVS) as it is humans who access the images at the end of any application on an image. By measuring the image degradation, it could possibly provide an estimate of how closer the degraded image is to the perceptual image quality.

Intuitively, quantitative measurement methods are used to compute the quality of the processed image. Thus, this will give great advantage to compare the measurement methods by evaluating the quantitative results when applied on standard image pairs. Consequently, the performance of the quality measurement methods can be understood and analyzed. The quantitative approaches include statistics error metrics to measure the quality of the image and make them better suitable for human access or machine analysis [4]. These error metrics such as Mean Square Error (MSE) [8], Peak Signal to Noise Ratio (PSNR) [8], Entropy [9], Correlated Coefficient [10], SSIM [11], UQI [12] etc. and many others are presently being used to compute the quality of an image.

It is to be noted that for different image processing problem, different quantitative measurement methods yield different quality. So, we have chosen, evaluated each measurement method based on its performance and suitability during each image processing problem and aim to provide a useful cluster of quantitative measurement methods.

1.2 Problem Description

Research in the area of image quality assessment focuses mainly on comparison between different quality measurement metrics based on several parameters such as accuracy, elapsed time taken etc. In ref [20], authors discuss the importance of objective analysis in image quality and also present automation of quality measurement metrics that can be used to evaluate the image quality attributes. In ref [21], the authors analyse and derive mathematical relationship between PSNR and Structural Similarity Index Measure (SSIM) which are the statistical error metrics that work for different image degradations. In ref [1], image quality measures and their performances are stated with results from experimentation. But we think that it is important to map these performances of quantitative methods by applying on differently processed images.

The current literature on the methods to compute the quality of an image is very wide and diverse. It is further challenging to identify and categorize the different quantitative methods in a structured manner that are suitable to be used for each image processing problem. In this thesis we aimed to identify and classify different quantitative measurement methods of image data based on their suitability for differently distorted images.

1.3 Aim and Objectives

The main aim of this thesis is to identify different existing image quantitative quality measurement metrics to measure the quality of images whose quality has been degraded due to different image processing problems. Also implement these metrics on the selected image set and compare and analyze the performance of each measurement method for an image processing problem (due to different distortions) on the selected image set.
Finally the goal is to provide a cluster of quantitative image quality measurement methods that are suitable for differently distorted images. The suitability of the image quality measurement methods is assessed by performing an experiment to evaluate their performance on a dataset of images with different types of distortions.

Objectives

- Identify existing quantitative image quality measurement methods
- Compute the quality of distorted images using the identified quantitative methods with respect to their original images
- Assess the quantitative measurement methods based on their performance on differently distorted images
- Obtain a useful cluster of suitable quantitative methods for different kinds of distorted images

1.4 Research Questions

R.Q.1 Which quantitative methods have been used to measure the image quality?

Motivation: In order to achieve the aim of the thesis, it is first important to know the quantitative methods of quality measurement being used most recently and that are effectively measuring the quality of an image. Where in these metrics identified could also be useful to analyze and infer the required results. This research question aims at identifying the different existing objective quality measurement techniques from which few metrics would be chosen specifically to perform an experiment further.

R.Q.2 Compute the quality of the selected standard test image dataset (image pair) using each of the quantitative methods identified in R.Q. 1

Motivation: To find which quantitative metric gives efficient result, calculating the quality of the distorted image with respect to the original image using different metrics is obviously necessary. This could yield a competitive result if applied on an image dataset. Then these computations resulted could be analyzed depending on the analysis component as these distorted images are resultants of different image processing applications as mentioned and are used as input to find their quality using five selected quality metrics.

R.Q.2.1 Which quantitative methods on the image dataset performs most efficiently?

Motivation: Based on the image dataset taken, the quantitative measurement metrics that is most effective can be found. This dataset contains the original (reference) images as well are images that are distorted because of various image processing applications i.e. compression, enhancement, reduction. So, with reference to the original image which metrics is most efficient has to be found in order to analyze the most appropriate metric for each distortion. The result obtained is highly dependent on the image dataset used to calculate the image quality.

R.Q.3 Which quantitative methods can be more suitable for different image processing problems?

Motivation: The final aim of this thesis is achieved by answering this question which requires analysis of results obtained in previous research questions.

1.5 Contribution to Literature

There are various algorithms introduced by many authors from decades to overcome the problem of quality evaluation of degraded images. But according to the literature study so far we think that the existing literature lacks to provide a cluster of most suitable quantitative methods to measure the image quality that has been degraded due to different image processing technique (compression, enhancement) etc. This thesis aims at providing the developers working on big image data projects with most suitable quantitative measurement method of image quality for different image processing problems (types of distortions). This could possibly reduce the time and effort for the developers of big image data projects as we provide the performance evaluation of these quality metrics as to what extent they give efficient results by comparing with the subjective evaluation of the selected image dataset.

1.6 Thesis Outline

The structure of this thesis document is as follows. Chapter 2 describes the required background work that pertains to this study and also discusses the existing literature work that is related to the study. Chapter 3 describes the methodology used to accomplish this research study. It is followed by the chapter 4 which presents the results and analysis obtained by following the method. Chapter 5 discusses the findings of the research. Chapter 6 describes the conclusions drawn from the results and analysis which also includes the possible future work to this research.

2 BACKGROUND AND RELATED WORK

This chapter describes the background work pertaining to this study and also the related work that has been done prior to this research is discussed.

2.1 Background

This section describes the basic background that is relevant to the topic of this thesis and that correlates with its objectives.

2.1.1 What is a digital image?

A digital image is a collection of pixels laid in a specific order of 'x' pixels of width and 'y' pixels of height. Each pixel has a specific value to correspond to a color or a grayscale value. A color (RGB) image consists of three channels, which are Red, Green and blue that are almost human eye receptors. A grayscale image consists of a single channel.

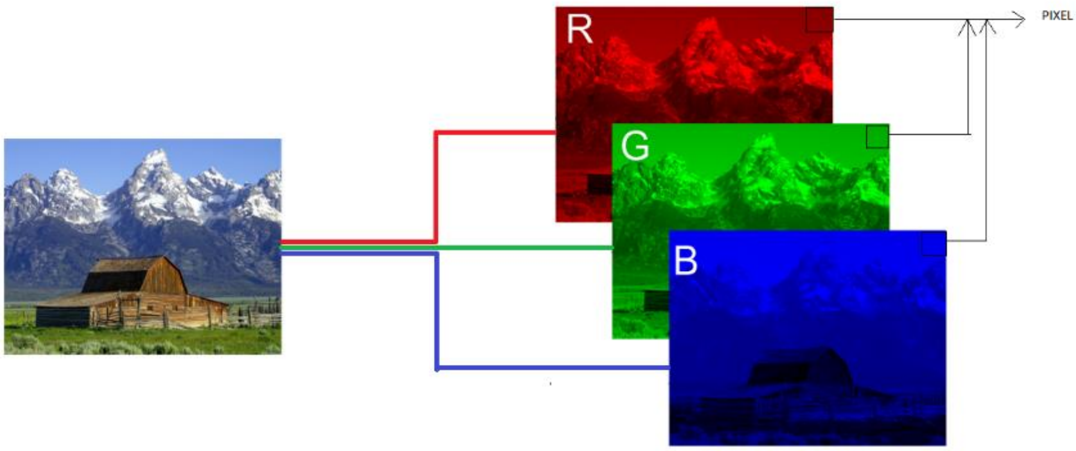

Figure 2.1 RGB image example

Figure 2.2 Grayscale image example

The above figures illustrate an RGB image and a grayscale image where an RGB image consists of three channels (Red, green and blue), each with 8bit pixels if it is a 24 bit image. And the grayscale image has a single channel. *(Source of images is Wikipedia, free encyclopedia)*[15]. Hence, it can be inferred that quality of an image changes depending on the number of channels. That is, quality of a grayscale image differs from the quality of its RGB image.

Quality of an image plays a vital role in the process of visual information acquisition. The quality of an image might be degraded by the time of human access due to several distortions. Better quality of an image is important as it enhances the knowledge acquired from the image. In image processing, compression for transmission and storage, some artifacts or noise might be introduced which degrades the visual quality of an image [8]. Image quality assessment models provide mathematical models to determine the perceptual quality of an image [8]. Image quality assessment models have been categorized into two: subjective and objective models [8, 16]. Subjective models involve humans to rate the quality of an image in a controlled environment [8]. The Mean Opinion Score (MOS) are obtained for the given test images by the multiple subjects [8]. By averaging these scores for each test image by multiple subjects, the mean opinion score and difference mean opinion scores are obtained. Objective quality measurement models are the mathematical models that can predict the quality of the image closer to the perceptual quality.

2.1.2 Drawbacks of subjective quality assessment models:

Subjective evaluation of images may be time consuming and expensive as this involves many number of observers to experiment. The results of the subjects are heavily dependent on their physical environment and emotional state. Moreover, factors such as lighting effect during the experiment and display device used highly influences the result obtained. Also, subjective evaluation cannot be incorporated into real time applications such as multimedia transmission systems [38]. Hence, it is necessary to compute the quality of degraded images using mathematical models that is, objective image quality assessment which are able to predict perceptual image quality in a consistent manner as to the subjective evaluation of image quality.

2.1.3 Use of objective image quality assessment models:

Objective evaluation models are the quantitative approaches that use intensity of the two images that is a reference type and a distorted type to compute the quality of the image [8]. These models are categorized based on the availability level of the referenced image for quality evaluation. Where, the referenced type is the original image and the quality of the distorted image is measured with respect to its original image. The quantitative approaches are classified into Full Reference (FR), Reduced Reference (RR), No Reference (NR) models [8, 16]. The No Reference models which are also called the ‘blind models’, are the mathematical models with quality assessment algorithms that calculate the quality of a distorted image without the help of a reference or an original image. These models can be used in any kind of application where image quality is to be evaluated as they do not need any prior information. In Reduced Reference models, the quality assessment algorithm is provided with partial information regarding the original version of the image to calculate the quality of the distorted image. Full Reference models are the mathematical models where the quality assessment algorithm has access to the original version of the distorted image and the quality of the distorted image is calculated with respect to the original image.

2.2 Related Work

Recently, many techniques to evaluate the image quality have been introduced. Enormous research study has been done which have proposed several computational models of image quality. Also, a lot of research has been done that compare evaluate and combine different statistical metrics in order to identify the most efficient measurement approach.

Different image processing techniques for different image processing applications like compression, enhancement are employed. For instance, there are several image compression methods in order to compress an image where in it is purely dependent on the application it is being used. In order to evaluate the image quality assessment methods, it is important to consider the image processing application (compression, enhancement, etc.) which they are being applied on.

2.2.1 Image Quality Assessment

In ref [8], the authors discussed about different image quality assessment techniques. These include subjective evaluation as well as objective evaluation techniques. The authors finally conclude that quality assessment is more efficient with the use of FR IQA techniques like SSIM and MSSIM as mathematical models like PSNR and MSE become unstable if the image degradation is significant. Also, in ref [9] the authors discuss about different subjective and objective image quality assessment techniques but the main focus of this study was to evaluate FR IQA metrics. 9 methods of this category were thoroughly described. Their performance and the computation times were evaluated. According to [9] there are a number of factors that are to be taken into account while selecting an IQA for a specific application. These factors may include availability of a reference image, computation time etc. Based on the application being used and the requirement in the scenario, the selection of a suitable IQA is of major importance.

In ref [17], authors have analysed image compression techniques using PSNR which is a quantitative image quality assessment method at different compression level. The authors gave a good relative analysis and tabulated the results of transform techniques using this method. According to [1], although some objective measures provide a good correlation with the subjective analysis of image quality of a given image compression technique they are not reliable enough to evaluate across different image processing techniques. Finally, the authors conclude that a useful analysis of image quality can be made with the combination of numerical and graphical measures of image quality assessment. It is difficult to obtain a universal quality evaluation metric that also corresponds to HVS [18]. But, these measures may be useful in order to determine and evaluate the useful image quality metrics [18].

2.2.2 Mapping image quality assessment to distortions

Research in the area of mapping quality assessment methods that are suitable to specific types of distortions is very limited. Mostly the researchers have focused on analysis of image quality of reconstructed images after compression or decompression techniques [19]. In ref [18] quantitative measurement methods have been analyzed and shown that some of the measures are considerably efficient to provide a perceptual quality after image compression. Also, in [1] the authors discuss numerical measures based on their evaluation for compression techniques where a conclusion has been drawn that says such kind of an evaluation is not reliable across other or different techniques of image compression. Hence, an idea of mapping quality assessment methods across different kinds of distortions caused by several image processing problems is necessary as there is a dearth in literature to focus on this.

3 METHODOLOGY

An appropriate research method should be selected in order to achieve the goal of the research. Among the existing available research methods, we selected a literature review and an experiment to perform this research. This is a mixed approach as it uses two methods. Both quantitative and qualitative data is obtained are involved to carry out these research methods. This research plan selected was to conduct a literature review where qualitative data were collected, which was to be followed by an experiment to obtain quantitative results [20]. These results were to be analyzed to obtain final results to the research questions.

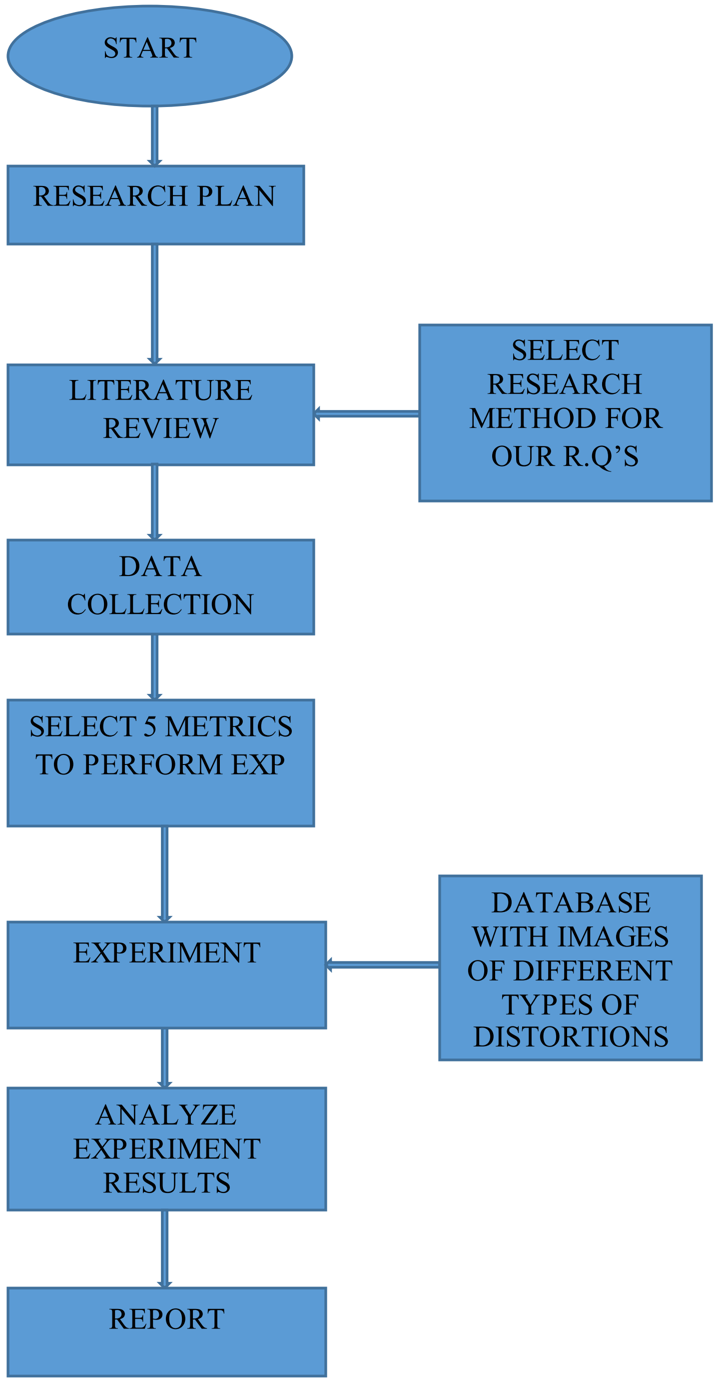

Figure 3.1 Research Plan

The above figure presents the plan used to carry out the research. Initially an LR was conducted to select methods of objective image quality to perform experiment on. Then an experiment was conducted by choosing a database with different distortions in images. The selected metrics were applied on each type of distorted images to evaluate their performance based on their closeness to perceptual visual quality. The results obtained are analyzed to meet the aim of the research.

3.1 Literature Review

An LR was conducted to identify the existing quantitative image quality metrics in image quality assessment research. According to [21], mapping study is a kind of LR, which has been conducted by the author as these are used to identify the existing and available literature prior and mostly rely on tabulating the primary studies in specific categories [21].

3.1.1 Mapping Study

Initially, Inspec database was selected which is directed to Engineering village. Going to the expert search and entering the start keywords which are "quantitative measures" AND "image quality" AND "image processing", we obtained 1757 results which contained articles, journal papers, conference papers etc. Then the fields "image processing" and year "2000-2016" were added which then narrowed the results. After limiting the search we obtained 46 articles. These articles are studied based on their relevance of title, then the abstract and conclusion. Later, keywords like Image Quality Assessment and image quality Metrics are added to find additional papers that might correlate to the search. Articles that are found to be relevant to identify the image quality metrics are selected in the similar way. 10 papers were obtained which were closely related to the thesis topic. After reading the full text of the selected papers, snowball sampling is used to find the image quality metrics and their functionalities.

3.1.2 Snowballing

Snowballing is a search approach which is used to help in identification of additional list of studies through citations, references of selected studies. This approach is used to identify and collect information about different objective image quality metrics that are being employed [21].

Start set identification: After reading the full text of the above chosen 10 papers, papers that are found irrelevant were discarded and the rest of the papers were reviewed to find IQM's using the snowball search. Most of the papers that were surveys from which additional studies could be discovered using their referenced studies.

3.1.3 Identified Quantitative Methods

There are several quantitative image quality measurement metrics that are being employed. Few of such existing quantitative methods that are mostly being used based on their behaviour are identified through a literature review. In the following subsections I_{ref} and I_{tst} are referred to the reference image and the test image i.e. distorted image respectively, where subscript ref denotes reference and tst denotes test image.

- **Mean squared error (MSE)**

Mean Squared Error (MSE) [11] is the simplest and mostly used full reference quality metric which is computed by calculating the average of squared intensity differences of distorted and referenced images [11]. The higher the value of MSE, infers that the error is high.

$$MSE = \frac{1}{WH} \sum_{j=1}^{H} \sum_{i=1}^{W} (I_{ref}(i,j) - I_{tst}(i,j))^2$$

Equation 1 Mathematical equation for MSE

In equation (1), I_{ref} (i,j) is the pixel value of original image at (i, j) position and I_{tst} (i,j) is the pixel value of the test image at (i,j) position. And W is the width and H is the height of the image.
The low complexity and inexpensive method of computation makes this method a good abstraction of image quality measurement. But, from the literature study it is known that it give poor performance when compared to human perceptual quality of the image.

- **Peak Signal to Noise Ratio (PSNR)**

This is an index metric which is defined as the ratio of the maximum power of the signal to the interruption noise in the signal.

$$PSNR = 10\ log_{10} \frac{255^2}{MSE}$$

Equation 2 Mathematical equation for PSNR

PSNR (i) shows the PSNR result of an image in I colour space, where PSNR (1) is red, PSNR (2) is green and PSNR (3) is blue. 255 is the maximum possible value of the pixel of an image when represented 8 bits per sample. MSE is the mean squared error. It is usually measured in decibels (dB). Peak Signal to Noise Ratio (PSNR) [22] is another simple and widely used image quality measurement metric because of its simple computation and physical meaning [11, 22]. But, these are not closely related to perceptual visual quality [8, 22]. So, Structural SIMilarity (SSIM) Index [11] has been developed as it has improved human visual perception capabilities [22].

- **Structural Similarity Index (SSIM)**

SSIM index explores the structural information of image. Traditional IQA quantifies error visibility, where in order to evaluate IQM that correlates well with HVS, it is important to estimate structural information change [11]. Here, the structural information includes luminosity and contrast details of the images. This is because actual purpose of human vision is to extract structural information. The performance of SSIM received great success as to the fact of HVS adaption to the structural information of image [23].It is an index metric that measures the structural similarity between two images. It is a full reference image. It is an improvement to the previous methods which are MSE and PSNR. It is measured between two windows x, y of same size.

$$SSIM(x,y) = \frac{\{(2\mu_x\mu_y + C_1)\ (2\sigma_{xy} + C_2)\}}{\{(\mu_x{}^2 + \mu_y{}^2 + C_2)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)\ \}}$$

Equation 3 Mathematical equation for SSIM

Here, μ_x is the average of x and μ_y is average of y and σ_x and σ_y are standard deviations of referenced and distorted image pixels. C_1 and C_2 are constants.

- **Feature Similarity Index (FSIM)**

It is an index metric that compares the low level features of the referenced and the distorted images. The close correlation of SSIM which explores structural information to HVS makes it an efficient IQM. But, the HVS understands image quality based on edge and zero crossings which are the low level features on an image [23]. Hence, an IQM that compares the low level features of the referenced image and the distorted image could give even more close relation to its subjective evaluation. One such technique is Feature SIMilarity (FSIM) induced FR IQA [23].
The computation of feature similarity index is done in two steps where first the similarity map is generated and it is then mapped to the similarity score. It is ranged between 0 and 1.

$$FSIM = \frac{\sum_{x\in\Omega} S_L(x).PC_m(x)}{\sum_{x\in\Omega} PC_m(x)}$$

Equation 4Mathematical equation for FSIM

The above is the equation to find feature similarity between original and test images where $S_L(x)$ is the similarity at location x, $PC_m(x)$ is the maximum phase congruency of original and test images at location x, Ω is the full image spatial domain.

- **Universal Image Quality Index (UQI)**

It is a mathematical model that computes the quality measure based on the combinations of three factors which are loss of correlation, luminance distortion and contrast distortion [12]. It is a full reference image quality measurement metric. Its values ranges between 0 and 1 where 1 being the best.

$$UQI = \frac{4\sigma_{xy}\overline{xy}}{(\bar{x}^2 + \bar{y}^2)(\sigma_x{}^2 + \sigma_x{}^2)}$$

Equation 5 Mathematical equation for UQI

Here, X is the original image and Y is the test image where X={x_1, $x_{2,...}$} and Y= {y_1, $y_{2,...}$}. $\bar{x}$ is the mean of X, σ_x^2 is the variance of x, σ_{xy} is covariance of xy and σ_x and σ_y are standard deviations of x and y respectively.

- **Visual Information Fidelity Criterion (VIFC)**

Visual Information Fidelity Criterion (VIFC) [13] is a full reference quality metric that explores the measurement of mutual information between input and output with respect to the original image. VIFC quantifies the loss of image information to the distorted signal and explores the relation between image information and visual quality [13, 23]. It is a ration of the two information measurements of the images that relates well with visual quality [23, 13]. Visual information fidelity criterion (VIF) has also come with assumption of HVS model.

- **Visual Saliency Index (VSI)**

Visual Saliency index measures the salient features of the distorted image with respect to the original image [34]. It has been proposed with the idea of using the Visual Saliency (VS) to compute local similarity between the original snd the distorted image where VS quantifies the low-level features of the image [24].Visual importance of the local region is computed, the VSI between two signal f_1 and f_2 is computed as

$$VSI = \frac{\sum_{x \in \Omega} S(x).VS_m(x)}{\sum_{x \in \Omega} VS_m(x)}$$

Equation 6 Mathematical equation for VSI

Here, Ω represents the whole spatial domain of the image. S(x) is the local similarity at position x. $VS_m(x)$ is the maximum of VS at position x, m= 1, 2 …

The existing objective quantitative image quality measurement methods identified through a literature review are tabulated below. Each quantitative method measures the quality based on different quality parameters which is mentioned as a variable in the table below.

IQM	**Variable**
MSE	Squared intensities
PSNR	Signal to noise ration
SSIM	Structural information
FSIM	Low level feature information
VIFC	Mutual information
UQI	Structural Distortions
VSI	Salient features

Table 0-1Identified existing IQM's

After conducting a literature review to identify the existing objective image quality measurement metrics, five of them were selected in order to conduct an experiment to evaluate their performance on different types of distorted images. The following objective image quality evaluation metrics are selected to evaluate their performance on different types of distorted images. These metrics were chosen as each one quantifies the quality of an image with different parameters. Also, each one is applied and compared among each set of distortion type with other metrics based on their consistency with subjective evaluation.

- Peak Signal to Noise Ratio (PSNR)
- Structural Similarity Index (SSIM)
- Feature Structural similarity Index (FSSIM)
- Universal Quality Index (UQI)
- Visual Saliency Index (VSI)

3.2 Experiment

This section describes about the procedures and data considered to carry out the experimental part of this study.

3.2.1 Data collection method

The data collected to carry out this thesis is of two forms:

3.2.1.1 Theoretical data

To carry out the experiment where the efficient IQM is found, few efficient available existing IQM’s were identified to compare their performance on several images that were distorted due to different image processing applications. These IQM’s were chosen from the identified IQM’s through a literature review. Among the identified IQM’s from the literature review, five mostly used IQM’s were chosen to compare with the help of an experiment.

3.2.1.2 Image dataset

The data collected in order to perform the experiment is stored in a database that consist of two major components which are the image dataset which contains the set of images and the Difference Mean Opinion Score (DMOS) (explained in detail in section 3.4) file which contains the scores that help to analyze the result obtained to achieve thesis goal.
The following are the benchmarking publicly available image datasets that are used to evaluate image quality metrics.

- CSIQ: Categorical Subjective Image Quality Database [25]
- LIVE Image Quality Database [26]
- Tampere Image Database (TID) [27]

LIVE database has been chosen for this thesis. The database was chosen in such a way that it contains images pertaining to different image processing applications (distortions) such as image compression, enhancement and noise reduction. The LIVE image quality assessment database provides a dataset of images with different distortion types whose quality has been rated by humans. Quality rating is given by DMOS.

In the end, the goal of the research is to find suitable efficient IQM for different types of distortion, for this needs the subjective opinion of human observers which is given by DMOS that has been provided in the database.

3.2.1.2.1 Data Preparation

The data collected cannot directly be used for the experiment. Instead, only the required and needed data are prepared prior to the start of the experiment which is believed to make it easier and efficient.

- The database prepared to conduct this experiment out of the LIVE image database consists of 20 original images (0-20), 20 JPEG 2000 compressed images (21-40), 20 JPEG compressed images (41-60) and 20 White Gaussian noised images (61-80), 20 Gaussian Blurred images (81-100).
- The DMOS values of the images are collected and stored in dmos_score.mat file.

RGB Image type	Number of images
Original(non-distorted images)	20 (0-20)
JPEG 2000 compressed (distorted) images of the original images	20 (21-40)
JPEG compressed (distorted) images of the original images	20 (41-60)
White Gaussian Noise (distorted) images of the original images	20 (61-80)
Gaussian Blur (distorted) images of the original images	20 (81-100)

Table 0-2Structure of images in the database prepared

The above table describes the structure of the database in which the images are stored. To make it clear, here is an illustration. The image number 21 is the JPEG 2000 compressed image of image number 1 which is an original (distortion less image). Image number 41 is the JPEG compressed image of image number 1 which is an original image. Image number 61 is the white Gaussian noise image of the original image number 1. Image number 81 is the Gaussian blur image of the original image. This way, the 20 original images are followed by its 4 distorted types of images respectively. This is because, it is believed that incorporating the database and inputting the images to apply IQM's could be more efficient while conducting the experiment (in Matlab).

Figure 3.1original RGB image

Figure 3.2JPEG 2000 compressed

Figure 3.3JPEG compressed image

Figure 3.4White Gaussian noise

Figure 3.5Gaussian blur

Figure 3.1-3.5 RGB Images 1, 21, 41, 61, 81 from the database that is original, its JPEG 2000 compressed, JPEG compressed, white Gaussian noise, Gaussian blur distortions.

Figure 3.6Grayscale original image

Figure 3.7Grayscale JPEG 2000 Compressed

Figure 3.8Grayscale JPEG Compressed

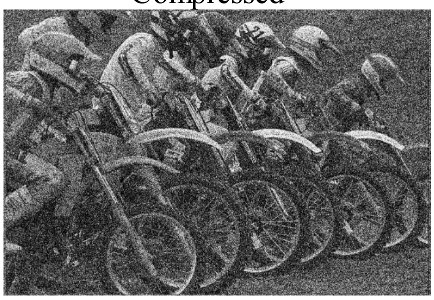

Figure 3.9Grayscale White Gaussian Noise

Figure 3.10Grayscale Gaussian blur Image

Figure 3.6-3.10 Grayscale Images 1, 21, 41, 61, 81 from the database that is original, its JPEG 2000 compressed, JPEG compressed, white Gaussian noise, Gaussian blur distortions.

3.2.2 Experiment design

An experiment was conducted to compare the performance of selected IQM's on images with different distortion types.

Independent variable: image set with different distortions

Dependent variable: image quality measurement methods that are compared based on their performance on the image set used.

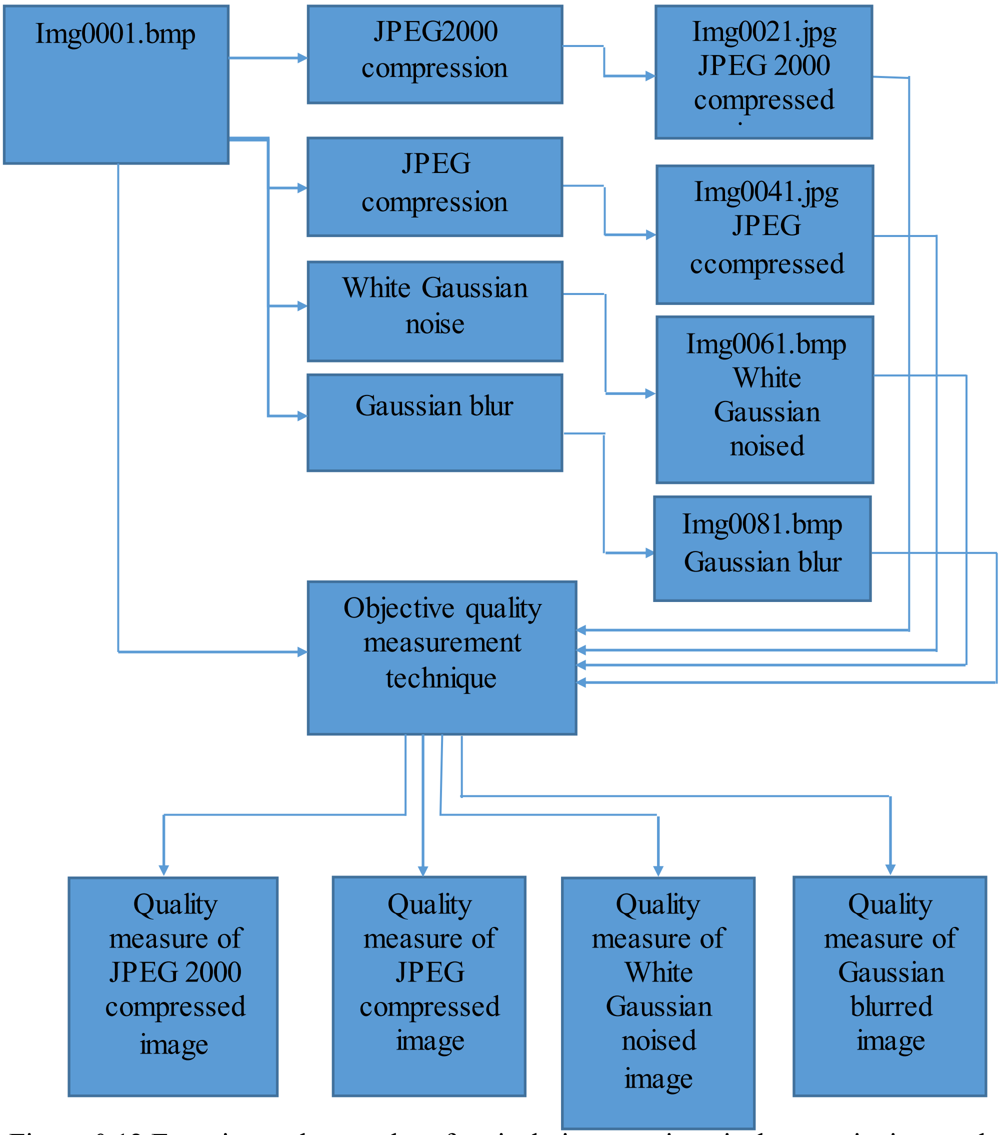

Figure 0.12 Experimental procedure for single image using single quantitative method

The above figure illustrates the experiment carried out. This procedure is applied on all the original images present in the database. Also, the quality is measured using five different objective quality measurement metrics whose performance is compared on each distortion.

System specifications
The hardware system configuration on which the experiment was carried out is a DELL Inspiron PC with an Intel core i5 processor and an 8GB RAM.

System Specification	Environment Variable
Operating system	Microsoft Windows 10
System type	X64
Processor	I5
RAM	8GB
Programming language	Matlab
Database	LIVE
IDE	Matlab environment

Table 0-3System Specifications

Matlab
Image processing applications require a dedicated software where as these packages cannot be easily modified or found by normal users [28]. MATLAB that is derived from matrix laboratory is a matrix oriented computing engine [28]. Thus, Matlab is a software package that is freely available which is used as an engine for image processing applications [28].
This experiment was carried out on Matlab 2015b as a licensed version was made available by BTH's IT Helpdesk.

3.2.2.1 Code generation

- After identifying and analyzing the literature review, five of the IQM's were selected on which the experiment was conducted.
- The image dataset and the selected IQM's are given as input so as to obtain the quality measure given by the IQM for each set of distorted images.
- Code generation is structured as follows:
  1. Initially, the database path was given in order to establish database connection.
  2. The IQM matrices are initialized to NULL to avoid garbage values.
  3. Input the image files; referenced and distorted respectively.
  4. All the IQM's do their job of calculating the quality of each distorted image with respect to the original reference image.

Note: The database prepared consists of only color images. The experiment was also conducted on gray scale images by converting the RGB images to gray scale using ***rgb2gray(colorImg_ref)*** *function to calculate the quality. This is done to identify which among the selected metrics are successful in quantifying the quality measures for gray scale images also.*

3.3 Hypothesis

After gathering the required results through an experimental procedure, statistical inferencing is required to make conclusions through analyzing the results obtained [29]. Testing the significance of claims made for the experiment is referred to hypothesis testing [29]. For this research, as per the experimental design to evaluate how differently each quantitative measurement method performs for differently distorted images, it is important to know if there is significant impact of the performance of these metrics for different distortion types of images. Which then allows us to find and analyze how differently they perform for each distortion type of images. Hence, the following hypotheses are generated:

H1:
Null Hypothesis (H_o): There is no significant impact of the quantitative metrics used to compare their performance on differently distorted images.
Alternate Hypothesis (H_a): There is a significant impact of the quantitative metrics used to compare their performance on differently distorted images.
The alternative hypothesis is two-sided as it claims from the experiment that whether there is a significant impact of the performance of these metrics on different types of distortions or there is no impact where they perform similarly for all types of distortions considered.

3.4 Analysis

In order to analyze the results obtained and find the significant difference in the performance of the quantitative measurement methods, a statistical method of analysis is required. To achieve accurate results an appropriate statistical test is important.
Guidelines in [30] were useful in the process of selection of the appropriate statistical test method. Following are the factors that influenced the selection of a statistical test for analysis:

Step 1- Identifying the variables and requirement for statistical test:

- ***Type of data:*** The experiment results in the quality measures given by each quantitative image quality measurement method. That is, the data to be analyzed is quantitative and in a discrete form as each metric gives quality measure in its measurable range.
- ***How is data organized?*** The results obtained in the experiment are indexed as they are recorded in columns of different types of distortions for each quantitative quality measurement method.
- ***How many samples may be recorded?*** The data to be analyzed statistically has four samples which are the distortion types among with the performance of each quantitative metrics is recorded.
- ***Independent/ dependent variables:*** For the statistical test, the independent variables are different quantitative methods and the types of distortions considered in the experiment. The dependent variable is the Difference in Quality Measure (DQM) that is obtained from the quality measure given by each quantitative metric to the standard subjective evaluation (which is DMOS, briefed below) for each type of distortion image set.
- ***Paired/ unpaired groups:*** The samples are independent as each one is a type of distorted version of original image set. Also, the samples on the group of 'methods' and 'distortions' are not related to each other which infers that these are unpaired groups.

To determine if parametric or non-parametric statistical tests are suitable - is data distributed normally?

- Based on normality of distribution we find whether the data is normally distributed or not. If the data is normally distributed, parametric test are suitable and if data is not distributed normally, non-parametric tests are suitable. The criterion of assumptions to be satisfied to opt for a suitable text are given in [31].
- Test for homogeneity of variance is conducted to know of the variance of data is homogeneous or not. Levene's test of homogeneity is conducted to test the homogeneity of variance.

Taking into consideration, the above factors, a Kruskal – Wallis statistical test of analysis is suitable for this experimental results. In order to perform this test IBM SSPS [32] statistical analysis tool is used.

The analysis of the results obtained from the experiment is done using statistical testing (Kruskal –Wallis test) to test the hypothesis. After testing the hypothesis, the significant difference is calculated using the effect size of the variables. Further, in order to find the metric which gives significant impact of its performance for each distortion type, differences of means from the standard subjective evaluation are calculated. Image quality is subjective in nature as it is ultimately viewed by humans [6]. Hence, its evaluation based on subjective evaluation is widely acceptable [6]. The quality measures by quantitative methods for each distortion are compared with the subjective evaluation by humans. Hence, an extensive experiment which was conducted at LIVE to obtain Mean Opinion Score (MOS) from human subjects for a number of images distorted that are present in this database was studied. These scores stored in a dmos_score.mat file were drawn from the LIVE database in order to analyze the results. These DMOS values are obtained from the LIVE database which is publicly available for image quality assessment research. DMOS is obtained directly by humans where humans are subjects to distorted images to rate the quality of the images. These values for each type of distorted images are collected from the database and are used for analyzing the performance of the quality measurement metrics which achieves the goal of the research.

4 RESULTS AND ANALYSIS

The results of literature review which is conducted in order to identify image quality assessment metrics to answer the R.Q.1 are presented in section 3.2. The results obtained from experiment and analyzing the results obtained from experiment are presented in this section.

4.1 Experiment

4.1.1 Results for R.Q.2

The quality of images in the dataset are found by performing an experiment in MATLAB. To calculate the quality measure by different quantitative methods, their algorithmic codes were gathered and generated according to the experimental design by inputting the database prepared. The final result obtained for each metric is tabulated below.

4.1.1.1 PSNR

The PSNR quality measure obtained by conducting an experiment on different types of distorted images both for grayscale and RGB images is tabulated as follows

Images (ref, test)	PSNR of grayscale image	PSNR of RGB image
1,21	22.20	26.55
2,22	18.55	23.09
3,23	26.63	27.07
4,24	28.98	33.10
5,25	25.04	28.88
6,26	22.77	26.82
7,27	23.65	27.95
8,28	21.82	25.96
9,29	22.31	26.45
10,30	20.81	25.13
11,31	27.04	31.15
12,32	24.33	28.93
13,33	25.68	30.14
14,34	24.20	28.60
15,35	23.80	27.49
16,36	27.92	32.50
17,37	23.78	28.13
18,38	30.00	33.52
19,39	29.84	34.05
20,40	25.06	29.20

Table 4-1PSNR values for JPEG 2000 compressed grayscale and RGB images

In the above table, the PSNR values for JPEG 2000 compressed images with respect to their original images both for grayscale and color images are presented. The value of PSNR is in Decibels (dB). (Here, (1, 21) represents (original image, JPEG 2000 compressed image) and so on)

Images (ref, test)	PSNR of grayscale image	PSNR of RGB image
1,41	22.71	26.29
2,42	20.72	24.70
3,43	20.69	24.62
4,44	29.42	31.93
5,45	26.15	29.11
6,46	23.28	26.58
7,47	23.62	27.11
8,48	24.26	27.59
9,49	20.15	23.97
10,40	19.32	23.30
11,51	24.05	26.84
12,52	23.70	26.52
13,53	24.94	28.31
14,54	25.68	29.47
15,55	29.56	31.75
16,56	29.60	33.38
17,57	24.81	28.51
18,58	30.97	32.94
19,59	30.39	33.99
20,60	25.70	28.84

Table 4-2PSNR values for JPEG compressed grayscale and RGB images

In the above table, PSNR values for JPEG compressed images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 41) represents (original image, its JPEG compressed image))

Images (ref, test)	PSNR of grayscale image	PSNR of RGB image
1,61	14.95	16.42
2,62	13.21	14.92
3,63	13.44	14.92
4,64	21.75	23.12
5,65	7.90	10.21
6,66	16.36	17.75
7,67	32.63	33.92
8,68	9.50	11.52
9,69	13.68	15.31
10,70	15.44	17.03
11,71	19.08	20.45
12,72	10.25	11.87
13,73	19.47	20.79
14,74	31.07	32.38
15,75	32.58	33.88
16,76	22.84	24.14
17,77	16.40	17.82
18,78	10.25	11.75
19,79	8.72	11.22
20,80	10.05	11.73

Table 4-3PSNR values for White Gaussian noise grayscale and RGB images

In the above table, PSNR values for white Gaussian noised images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 61) represents (original image, its white Gaussian noised image))

Images (ref, test)	PSNR of grayscale image	PSNR of RGB image
1,81	23.59	28.43
2,82	18.64	23.46
3,83	18.67	23.52
4,84	25.25	30.11
5,85	25.59	30.27
6,86	25.39	30.14
7,87	17.45	22.17
8,88	25.14	29.83
9,89	24.32	29.01
10,90	18.21	22.96
11,91	24.14	28.78
12,92	18.54	23.45
13,93	25.05	29.86
14,94	22.15	26.84
15,95	26.29	31.05
16,96	28.40	33.20
17,97	26.13	30.76
18,98	24.14	28.39
19,99	22.12	26.97
20,100	20.57	25.13

Table 4-4PSNR values for Gaussian blur grayscale and RGB images

In the above table, PSNR values for Gaussian blurred images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 81) represents (original image, its Gaussian blurred image))

4.1.1.2 SSIM

The SSIM quality measure obtained by conducting an experiment on different types of distorted images both for grayscale and RGB images is tabulated as follows

Images (ref, test)	SSIM of grayscale image	SSIM of RGB image (MSSIM)
1,21	0.72	0.70
2,22	0.61	0.60
3,23	0.80	0.79
4,24	0.83	0.81
5,25	0.82	0.79
6,26	0.69	0.65
7,27	0.82	0.81
8,28	0.61	0.59
9,29	0.81	0.79
10,20	0.80	0.78
11,31	0.74	0.72
12,32	0.71	0.77
13,33	0.85	0.83
14,34	0.74	0.74
15,35	0.81	0.77

16,36	0.77	0.76
17,37	0.71	0.70
18,38	0.87	0.85
19,39	0.90	0.88
20,40	0.74	0.72

Table 4-5SSIM values for JPEG 2000 compressed images

In the above table, SSIM values for JPEG 2000 compressed images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 21) represents (original image, its JPEG 2000 compressed image))

Images (ref, test)	SSIM of grayscale image	SSIM for RGB images
1,41	0.78	0.74
2,42	0.82	0.78
3,43	0.75	0.73
4,44	0.83	0.79
5,45	0.87	0.82
6,46	0.76	0.69
7,47	0.81	0.76
8,48	0.77	0.71
9,49	0.70	0.67
10,50	0.72	0.69
11,51	0.59	0.54
12,52	0.67	0.69
13,53	0.83	0.79
14,54	0.85	0.83
15,55	0.92	0.87
16,56	0.87	0.84
17,57	0.82	0.79
18,58	0.86	0.80
19,59	0.93	0.90
20,60	0.80	0.75

Table 4-6SSIM values for JPEG compressed images

In the above table, SSIM values for JPEG compressed images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 41) represents (original image, its JPEG compressed image))

Images (ref, test)	SSIM of grayscale image	SSIM for RGB images
1,61	0.56	0.40
2,62	0.57	0.42
3,63	0.52	0.38
4,64	0.54	0.38
5,65	0.08	0.05
6,66	0.58	0.45
7,67	0.95	0.91
8,68	0.16	0.09
9,69	0.50	0.38
10,70	0.60	0.47

11,71	0.53	0.38
12,72	0.13	0.13
13,73	0.55	0.42
14,74	0.96	0.91
15,75	0.94	0.88
16,76	0.66	0.49
17,77	0.49	0.34
18,78	0.08	0.05
19,79	0.10	0.05
20,80	0.14	0.08

Table 4-7SSIM values for white Gaussian noise images

In the above table, SSIM values for white Gaussian noised images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 61) represents (original image, its white Gaussian noised image))

Images (ref, test)	SSIM of grayscale image	SSIM for RGB images
1,81	0.87	0.87
2,82	0.67	0.67
3,83	0.64	0.64
4,84	0.76	0.77
5,85	0.93	0.93
6,86	0.93	0.93
7,87	0.47	0.47
8,88	0.87	0.87
9,89	0.94	0.94
10,90	0.61	0.62
11,91	0.65	0.65
12,92	0.49	0.51
13,93	0.88	0.87
14,94	0.76	0.76
15,95	0.94	0.94
16,96	0.87	0.87
17,97	0.94	0.94
18,98	0.78	0.77
19,99	0.77	0.76
20,100	0.44	0.44

Table 4-8 SSIM values for Gaussian blur images

In the above table, SSIM values for Gaussian blur images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 81) represents (original image, its Gaussian blurred image))

4.1.1.3 FSIM

The FSIM quality measure obtained by conducting an experiment on different types of distorted images both for grayscale and RGB images is tabulated as follows

Images (ref, test)	FSIM of RGB image	FSIM of grayscale image
1,21	0.85	0.86

2,22	0.82	0.83
3,23	0.89	0.89
4,24	0.87	0.87
5,25	0.87	0.88
6,26	0.85	0.85
7,27	0.88	0.89
8,28	0.79	0.79
9,29	0.88	0.89
10,20	0.89	0.90
11,31	0.87	0.87
12,32	0.79	0.79
13,33	0.88	0.88
14,34	0.86	0.87
15,35	0.83	0.84
16,36	0.87	0.87
17,37	0.85	0.85
18,38	0.91	0.91
19,39	0.93	0.93
20,40	0.87	0.87

Table 4-9 FSIM values for JPEG 2000 compressed images

In the above table, FSIM values for JPEG 2000 compressed images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 21) represents (original image, its JPEG 2000 compressed image))

Images (ref, test)	FSIM of RGB image	FSIM of grayscale image
1,41	0.87	0.88
2,42	0.89	0.91
3,43	0.82	0.83
4,44	0.86	0.97
5,45	0.90	0.91
6,46	0.86	0.88
7,47	0.86	0.87
8,48	0.85	0.86
9,49	0,77	0.78
10,50	0.83	0.84
11,51	0.66	0.66
12,52	0.73	0.73
13,53	0.84	0.85
14,54	0.90	0.90
15,55	0.93	0.93
16,56	0.92	0.93
17,57	0.89	0.89
18,58	0.88	0.89
19,59	0.95	0.95
20,60	0.88	0.89

Table 4-10 FSIM values for JPEG compressed images

In the above table, FSIM values for JPEG compressed images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 41) represents (original image, its JPEG compressed image))

Images (ref, test)	FSIM of RGB image	FSIM of grayscale image
1,61	0.77	0.81
2,62	0.77	0.82
3,63	0.72	0.76
4,64	0.78	0.80
5,65	0.49	0.52
6,66	0.80	0.84
7,67	0.98	0.98
8,68	0.56	0.60
9,69	0.70	0.74
10,70	0.83	0.87
11,71	0.78	0.81
12,72	0.49	0.52
13,73	0.83	0.85
14,74	0.97	0.98
15,75	0.98	0.98
16,76	0.85	0.86
17,77	0.73	0.76
18,78	0.37	0.39
19,79	0.36	0.39
20,80	0.55	0.59

Table 4-11 FSIM values for white Gaussian noise images

In the above table, FSIM values for white Gaussian noised images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 61) represents (original image, its White Gaussian noised image))

Images (ref, test)	FSIM of RGB image	FSIM of grayscale image
1,81	0.92	0.92
2,82	0.82	0.83
3,83	0.78	0.78
4,84	0.79	0.79
5,85	0.94	0.94
6,86	0.96	0.96
7,87	0.59	0.59
8,88	0.92	0.92
9,89	0.96	0.96
10,90	0.72	0.72
11,91	0.76	0.76
12,92	0.58	0.58
13,93	0.89	0.89
14,94	0.85	0.85
15,95	0.93	0.93
16,96	0.92	0.92
17,97	0.96	0.96
18,98	0.81	0.81
19,99	0.80	0.80
20,100	0.61	0.61

Table 4-12 FSIM values for Gaussian blur images

In the above table, FSIM values for Gaussian blurred images with respect to their original images both for grayscale and RGB images are presented. (Here, (1, 81) represents (original image, its Gaussian blurred image))

4.1.1.4 VSI

The VSI quality measure obtained by conducting an experiment on different types of distorted images for RGB images is tabulated as follows. The VSI quality measures for the dataset are calculated only for RGB images because computation of VSI for grayscale images either involves graph based machine learning algorithms [33] or spatial residual algorithms [34, 35] which is considered complex and hence, these were not reported in this study. However, as the VSI is obtained for RGB images, it has been compared with others metrics as per main goal of this study.

Images (ref, test)	VSI of RGB image
1,21	0.94
2,22	0.92
3,23	0.96
4,24	0.97
5,25	0.96
6,26	0.94
7,27	0.96
8,28	0.93
9,29	0.95
10,20	0.96
11,31	0.96
12,32	0.94
13,33	0.97
14,34	0.95
15,35	0.93
16,36	0.97
17,37	0.95
18,38	0.97
19,39	0.98
20,40	0.96

Table 4-13Visual Saliency index for JPEG 2000 compressed images

In the above table, VSI values for JPEG 2000 compressed images with respect to their original images for RGB images are presented. (Here, (1, 21) represents (original image, its JPEG 2000 compressed image))

Images (ref, test)	VSI of RGB image
1,41	0.94
2,42	0.95
3,43	0.92
4,44	0.96
5,45	0.96
6,46	0.94
7,47	0.94

8,48	0.94
9,49	0.90
10,50	0.92
11,51	0.89
12,52	0.91
13,53	0.95
14,54	0.96
15,55	0.97
16,56	0.98
17,57	0.96
18,58	0.97
19,59	0.98
20,60	0.96

Table 4-14 VSI values for JPEG compressed images

In the above table, VSI values for JPEG compressed images with respect to their original images for RGB images are presented. (Here, (1,421) represents (original image, its JPEG compressed image))

Images (ref, test)	VSI of RGB image
1,61	0.88
2,62	0.88
3,63	0.84
4,64	0.91
5,65	0.70
6,66	0.89
7,67	0.99
8,68	0.75
9,69	0.84
10,70	0.88
11,71	0.89
12,72	0.73
13,73	0.90
14,74	0.99
15,75	0.99
16,76	0.94
17,77	0.86
18,78	0.68
19,79	0.65
20,80	0.76

Table 4-15 VSI values for white Gaussian noise images

In the above table, VSI values for white Gaussian noised images with respect to their original images for RGB images are presented. (Here, (1, 61) represents (original image, its white Gaussian noised image))

Images (ref, test)	VSI of RGB image
1,81	0.97
2,82	0.92

3,83	0.92
4,84	0.95
5,85	0.98
6,86	0.98
7,87	0.87
8,88	0.97
9,89	0.98
10,90	0.91
11,91	0.94
12,92	0.86
13,93	0.97
14,94	0.95
15,95	0.97
16,96	0.98
17,97	0.98
18,98	0.95
19,99	0.94
20,100	0.88

Table 4-16 VSI values for Gaussian blur images

In the above table, VSI values for Gaussian blurred images with respect to their original images for RGB images are presented. (Here, (1, 81) represents (original image, its Gaussian blurred image))

4.1.1.5 UQI

The UQI quality measure obtained by conducting an experiment on different types of distorted images for grayscale images is tabulated as follows. UQI of an image quantifies the three factors of loss of correlation, chrominance and contrast estimation. But for color image quality, a metric called Color Image Quality Index [36] has been proposed which is an improvement to UQI. Hence, UQI for RGB images has not been reported in this study. Further this has been listed as a limitation for this study (see section 5.3).

Images (ref, test)	UQI of grayscale image
1,21	0.9974
2,22	0.9955
3,23	0.9996
4,24	0.9999
5,25	0.9999
6,26	0.9997
7,27	0.9988
8,28	0.9567
9,29	0.9959
10,20	0.9999
11,31	0.9999
12,32	0.9998
13,33	0.9999
14,34	0.9906
15,35	0.9999
16,36	0.9999
17,37	0.9999
18,38	0.9999
19,39	0.9955

20,40	0.9999

Table 4-17 UQI values for JPEG 2000 compressed images

In the above table, UQI values for JPEG 2000 compressed images with respect to their original images for grayscale images are presented. (Here, (1, 21) represents (original image, its JPEG 2000 compressed image))

Images (ref, test)	UQI of grayscale image
1,41	0.9965
2,42	0.9987
3,43	0.9985
4,44	0.9999
5,45	0.9999
6,46	0.9999
7,47	0.9996
8,48	0.9577
9,49	0.9883
10,50	0.9999
11,51	0.9999
12,52	0.9999
13,53	0.9999
14,54	0.9918
15,55	0.9999
16,56	0.9999
17,57	0.9999
18,58	0.9999
19,59	0.9954
20,60	0.9935

Table 4-18 UQI values for JPEG compressed images

In the above table, UQI values for JPEG compressed images with respect to their original images for grayscale images are presented. (Here, (1, 41) represents (original image, its JPEG compressed image))

Images (ref, test)	UQI of grayscale image
1,61	0.9997
2,62	0.9997
3,63	0.9999
4,64	0.9998
5,65	0.9999
6,66	0.9999
7,67	0.9999
8,68	0.9937
9,69	0.9992
10,70	0.9999
11,71	0.9999
12,72	0.9999
13,73	0.9998
14,74	0.9910
15,75	0.9999

16,76	0.9998
17,77	0.9997
18,78	0.9999
19,79	0.9999
20,80	0.9999

Table 4-19 UQI values for white Gaussian noise images

In the above table, UQI values for white Gaussian noised images with respect to their original images for grayscale images are presented. (Here, (1, 61) represents (original image, its white Gaussian noised image))

Images (ref, test)	UQI of grayscale image
1,81	0.9990
2,82	0.9882
3,83	0.9997
4,84	0.9999
5,85	0.9999
6,86	0.9999
7,87	0.9992
8,88	0.9884
9,89	0.9999
10,90	0.9998
11,91	0.9999
12,92	0.9999
13,93	0.9999
14,94	0.9996
15,95	0.9999
16,96	0.9999
17,97	0.9999
18,98	0.9999
19,99	0.9960
20,100	0.9992

Table 4-20 UQI values for Gaussian blur images

In the above table, UQI values for JPEG 2000 compressed images with respect to their original images for grayscale images are presented. (Here, (1, 21) represents (original image, its JPEG 2000 compressed image))

4.1.2 Results for R.Q.2.1

All the quantitative methods were successful in measuring the quality of images with different distortion types present in our database. Each quantitative method takes into account different parameters to measure the quality of an image. For each quantitative metric performs differently on differently distorted images.

As per our experiment the measure of quality obtained by each quantitative method is tabulated as follows. The quality measures reported are the mean values for each set of distortion type images. These are the means of normalized quality measures which are calculated using unity based normalization [37]. This is done as metrics that quantify images using different parameters can be ranged in similar order in order to compare.

4.1.2.1 PSNR

The quality measure obtained using the PSNR quality metric (that are tabulated in the result of previous R.Q) ranged between 10-50 dBs But, in order to evaluate the performance we normalized the value to result in a value between 0-1. Thereby, value obtained closer to perceptual visual quality tends to be more efficient.

Quality Measurement	JPEG 2000 Compression	JPEG compression	White Gaussian Noise	Gaussian Blur
PSNR	0.4875	0.4661	0.3516	0.5024
DMOS	0.6305	0.6124	0.5171	0.5804

Table 4-21Mean of normalized PSNR values for each distortion type

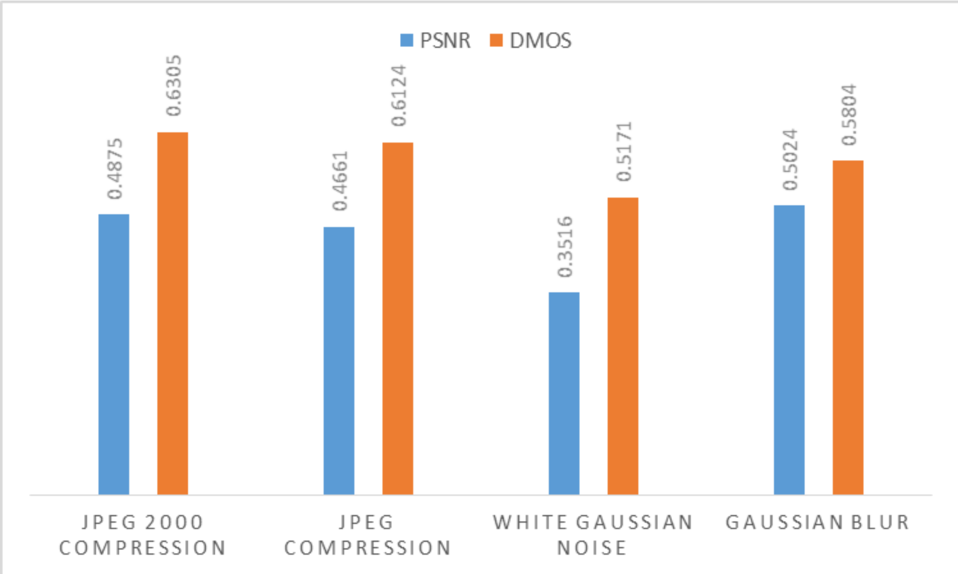

Figure 4.1Comparing the PSNR values with DMOS for each type of distortion

The above figure describes the performance of PSNR in measuring the image quality in comparison with the DMOS of the images in the dataset. From the figure it is clear that PSNR gives closer correlation with the DMOS for Gaussian Blur images. That is, PSNR performs better for Gaussian Blur type of distortion compared to the other distortion types.

4.1.2.2 SSIM

The mean SSIM values obtained for each distorted image set with respect to their original images are as follows.

Quality Measurement	JPEG 2000 Compression	JPEG compression	White Gaussian Noise	Gaussian Blur
SSIM	0.7525	0.759	0.383	0.761
DMOS	0.6305	0.6124	0.5171	0.5804

Table 4-22Mean SSIM values for each distortion type

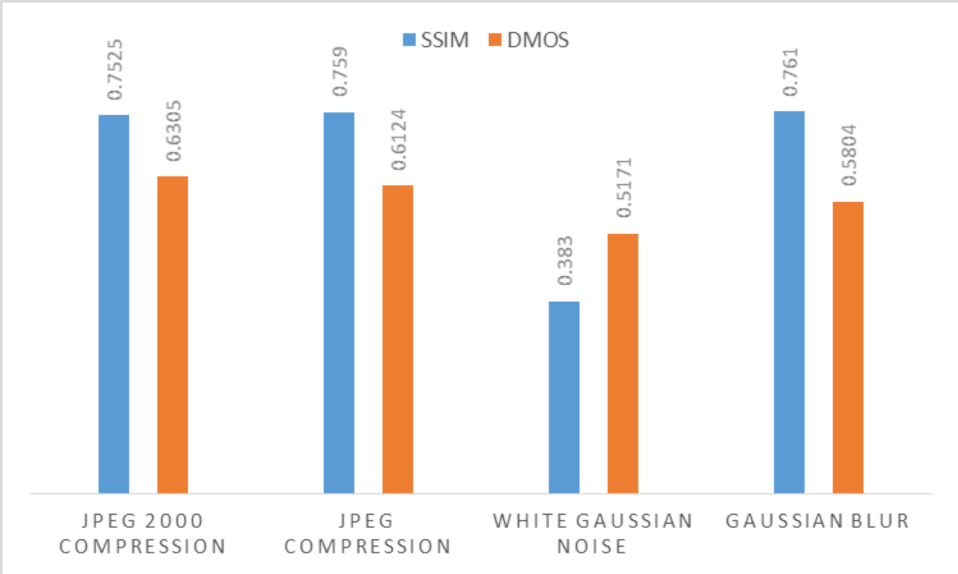

Figure 4.2Comparing the SSIM values with DMOS for each distortion type

The above figure presents the performance of SSIM in comparison with the DMOS for each distortion type. From the figure, it is evident that SSIM give closer correlation with DMOS for JPEG 2000 compressed images.

4.1.2.3 FSIM

The mean FSIM values obtained for each distortion type images with respect to their original images are tabulated below

Quality Measurement	JPEG 2000 Compression	JPEG compression	White Gaussian Noise	Gaussian Blur
FSIM	0.8625	0.8589	0.7155	0.8255
DMOS	0.6305	0.6124	0.5171	0.5804

Table 4-23Mean FSIM for each distortion type

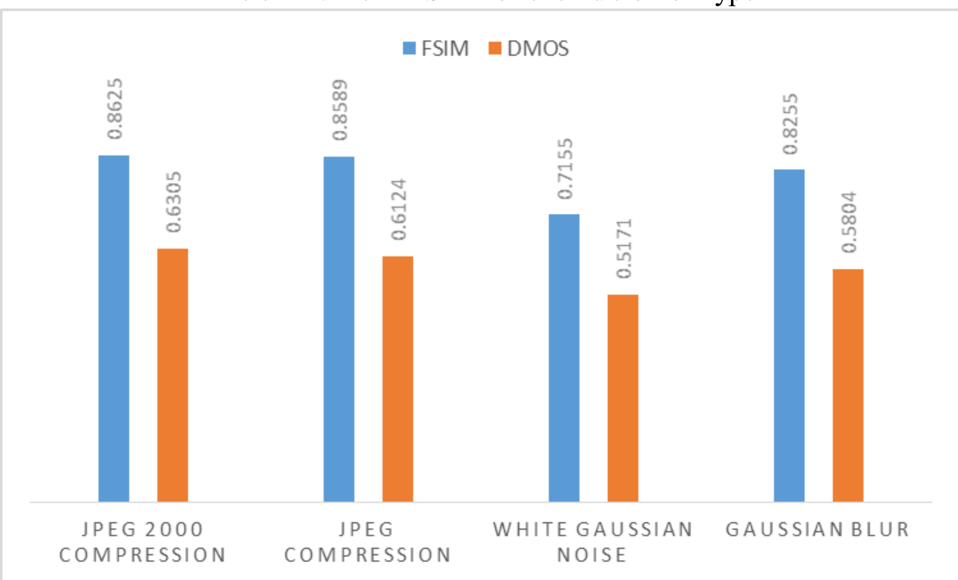

Figure 4.2Comparing the FSIM values with the DMOS for each distortion type

The above figure depicts the graphical representation of the performance of FSIM in comparison with DMOS for each distortion type. From the figure, it is clear that FSIM gives close correlation with DMOS for white Gaussian noise images. That is, FSIM performs better for white Gaussian noise type of distortion images.

4.1.2.4 VSI

The mean VSI values for each distortion type image set with respect to their original image are tabulated as follows

Quality Measurement	JPEG 2000 Compression	JPEG compression	White Gaussian Noise	Gaussian Blur
VSI	0.9525	0.945	0.8475	0.9435
DMOS	0.6305	0.6124	0.5171	0.5804

Table 4-24Mean VSI values for each distortion type

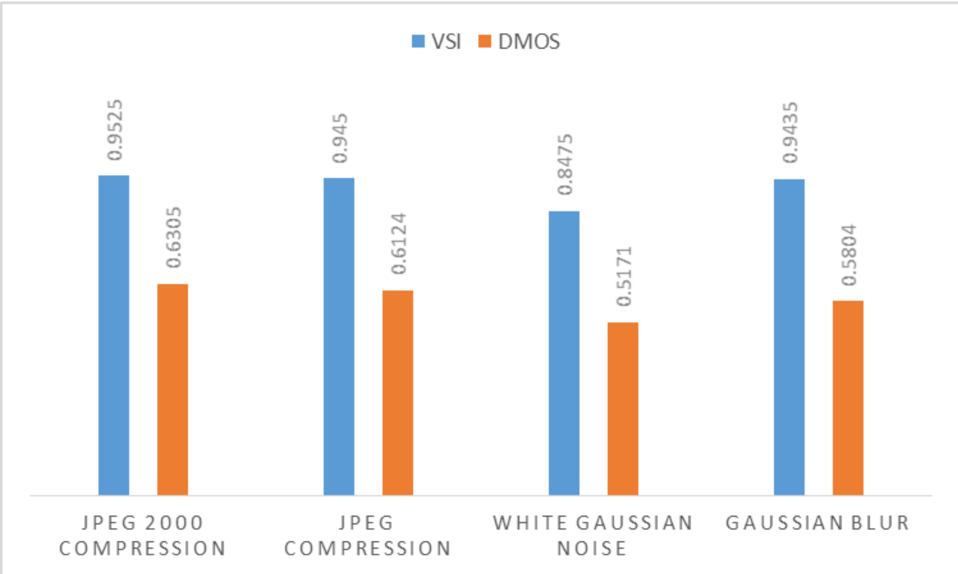

Figure 4.3Comparing the VSI values with DMOS for each distortion type

The above figure presents the performance of VSI in comparison with the DMOS for each distortion type. The mean VSI value is closer to the DMOS value for a white Gaussian noise images. Hence, VSI performs closer to perceptual visual quality for white Gaussian noised images.

4.1.3 Results for R.Q.3

4.1.3.1 JPEG 2000 compression

The quality measures obtained from each quantitative method from the experiment are tabulated below.

DISTORTION TYPE	DMOS	PSNR	SSIM	FSIM	VSI
JPEG 2000 COMPRESSION	0.6305	0.4875	0.7525	0.8625	0.9525

Table 4-25Quality measures obtained by different quantitative methods for JPEG 2000 compressed images

The above table shows quality measures by different quantitative methods for the same set of JPEG 2000 compressed images with respect to their original images. The quantitative method that gives the closest value to the DMOS value of the dataset is said to give the most accurate result. This is analyzed and presented in section 4.2.4.1.

4.1.3.2 JPEG compression

The quality measures obtained from each quantitative method from the experiment are tabulated below.

DISTORTION TYPE	DMOS	PSNR	SSIM	FSIM	VSI
JPEG COMPRESSION	0.6124	0.4661	0.759	0.8589	0.945

Table 4-28Quality measures obtained by different quantitative methods for JPEG compressed images

The above table shows the quality measures resulted by each quantitative measurement method for the set of JPEG compressed images. The suitable metric for JPEG compressed images is analyzed by calculating differences in section 4.2.4.2.

4.1.3.3 White Gaussian Noise

The quality measures obtained from each quantitative method from the experiment are tabulated below.

DISTORTION TYPE	DMOS	PSNR	SSIM	FSIM	VSI
WHITE GAUSSIAN NOISE	0.5171	0.3516	0.383	0.7155	0.8475

Table 4-29Quality measures obtained by different quantitative methods for white Gaussian noise images

The above table reports the quality scores obtained by different quantitative image quality measurement metrics for white Gaussian noised images. The quantitative measure that gives the closest value to DMOS is said to give quality measure closer to human visual system (see section 4.2.4.3).

4.1.3.4 Gaussian blur

The quality measures obtained from each quantitative method from the experiment are tabulated below.

DISTORTION TYPE	DMOS	PSNR	SSIM	FSIM	VSI
GAUSSAIN BLUR	0.5804	0.5024	0.761	0.8255	0.9435

Table 4-30Quality measures obtained by different quantitative methods for Gaussian blur images

The above table shows the quality measures obtained by different quantitative metrics in comparison with the DMOS value for Gaussian blur images. The quantitative metric that is most suitable among the compared methods is analyzed and reported in section 4.2.

4.2 Analysis

This section presents the analysis of the experimental results obtained to answer R.Q.2, R.Q.2.1 and R.Q.3. It also addresses the statistical testing and analyses the results corresponding to the statistical test used.

To calculate the differences of means given by each image quality measurement method from the subjective evaluation of each distortion type image set, it is first important to know if these methods actually give a significant impact for each type of distortion. This is achieved by testing the hypothesis stated in section **X.** Each selected quantitative measurement method corresponding to images with different kinds of distortions is presented. The statistical test Kruskal-Wallis is used to find if these methods make a significant impact.

Variables	Groups	Levels	Description
Independent	Methods	4	PSNR
			SSIM
			FSIM
			VSI
	Distortions	4	JPEG 2000 compression
			JPEG compression
			White Gaussian Noise
			Gaussian Blur
Dependent	Difference in Quality Measure (DQM)	-	It is the difference between the mean of quality measure given by each quantitative method and the DMOS value for each distortion type.

Table 4-26 Variables used for statistical analysis

The above table presents the independent and dependent variable of the statistical test. The independent variables have more than 3 levels and as the data is non homogenous, a non-parametric statistical test that is Kruskal-Wallis is used to test the significance of quantitative methods on differently distorted images.

4.2.1 Result obtained from statistical test

4.2.1.1 Descriptive statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Methods	16	1.00	4.00	2.5000	1.15470	1.333	.000	.564	-1.411	1.091
Distortion	16	1.00	4.00	2.5000	1.15470	1.333	.000	.564	-1.411	1.091
DQM	16	.07	.36	.2077	.08858	.008	.428	.564	-1.010	1.091
Valid N (listwise)	16									

Table 4-27 Descriptive statistics obtained for experimental results

Kruskal-Wallis Test

Ranks

	Methods	N	Mean Rank
DQM	PSNR	4	4.63
	SSIM	4	4.38
	FSIM	4	10.50
	VSI	4	14.50
	Total	16	

Test Statistics[a,b]

	DQM
Chi-Square	12.768
df	3
Asymp. Sig.	.005

a. Kruskal Wallis Test

b. Grouping Variable: Methods

Table 4-28 Results of Kruskal-Wallis statistical test

From the above table, it is observed that the significance (p-value) is 0.005 which is less than 0.05. Hence, it can be inferred that these quantitative measurement methods have a statistically significant impact on different kinds of distortions. For the experiment conducted, the quantitative methods compared for types of distortions have significantly different behavior and hence have an impact on the type of distortion.

4.2.2 Hypothesis

From the above analysis results from the statistical test it is indicated that the quantitative methods have a statistically significant difference in performance for different types of distortions. Hence, the null hypothesis is rejected.

4.2.3 Effect size

Effect size is a measure of strength of an effect. It measures to what extent the significant difference is present on the dependent variable. That is, to what extent the performance of the quantitative methods have an impact on the different kinds of distortions for the experimental results.

$$Effect\ size\ for\ Kruskal - Wallis\ Sample = \frac{Chi - Square}{n - 1}$$

That is, for this statistical test, 0.8512 is the effect size which indicates that it these quantitative methods have a large impact to perform significantly in a different manner for the dataset of 4 types of distortions. This difference is calculated by comparing with the standard subjective evaluation of specific distorted image set.

4.2.4 Calculating minimum differences by each metric from DMOS for each distortion type

The quality measure obtained by each quantitative measurement methods is analyzed by comparing it with the subjective evaluation as it is closer to perceptual image quality. Analysis is done by comparing the quality values with the DMOS values of each image set with different distortions.

The DMOS scores of each distortion type present in the database are calculated as follows: For each distortion type a random number of subjects were involved in the experiment to rate the quality of the images presented to them. The images are rated between 0-100. 0 being least quality and 100 being highest quality of the image [23]. The mean of DMOS values for each distortion type is calculated as it is be compared to the mean quality scores obtained from each quantitative method. The DMOS values are normalized to 0-1 as all the other quantitative metrics' scores are normalized to 0-1 so that it would make it simple to compare among them. For the image set used the DMOS scores obtained are tabulated as follows:

Distortion type	DMOS (normalized to 0-1)
JPEG 2000 compression	0.6305
JPEG compression	0.6124
White Gaussian noise	0.5171
Gaussian Blur	0.5804

Table 4-29 DMOS scores for each distortion type from the database

These DMOS values are only recorded for the RGB images in the database as the database consists of only the RGB images whereas the quality measurement for grayscale images was possible by converting the RGB images to grayscale using ***rgb2gray(colorimg)*** command in MATLAB. Hence, the following results are tabulated only to distorted RGB images with respect to their original RGB images.

4.2.4.1 JPEG 2000 compression

In order to find the closest value to DMOS, differences from each method to the standard DMOS are calculated. This is found by the following equation.

$$\begin{aligned}|\mathrm{MIN}\ ((&\mathrm{DMOS}_{JPEG2000} - \mathrm{PSNR}_{JPEG2000}),\\ &(\mathrm{DMOS}_{JPEG2000} - \mathrm{SSIM}_{JPEG2000}),\\ &(\mathrm{DMOS}_{JPEG2000} - \mathrm{FSIM}_{JPEG2000}),\\ &(\mathrm{DMOS}_{JPEG2000} - \mathrm{VSI}_{JPEG2000}))|\end{aligned}$$

Equation 7to find the most accurate quantitative metric for JPEG 2000 compressed images

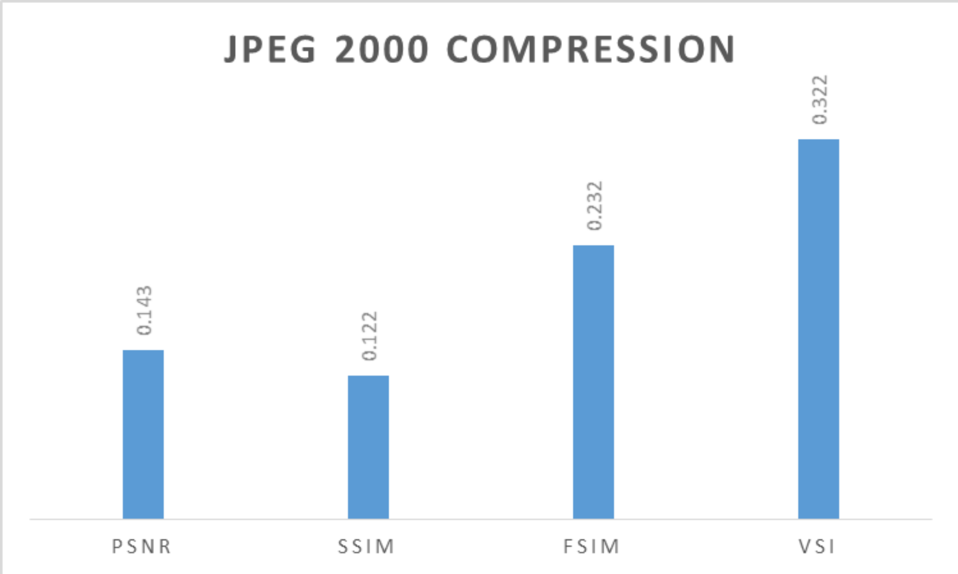

Figure 4.4Differences obtained for each quantitative method with DMOS of JPEG 2000 compression

The values obtained by applying the above equation are represented graphically in the figure above which makes the quantitative method that gives the minimum difference prominently visible. That is, for JPEG 2000 compression, it is visible from the figure above that, SSIM gives the least difference. Hence, SSIM measures the quality of JPEG 2000 compressed images most accurately when compared to other metrics for our dataset.

4.2.4.2 JPEG compression

Similar to JPEG 2000 compression, the quantitative method that resulted in an accurate measure is found by the following statistical equation.

$$
\begin{aligned}
|MIN\ ((&DMOS_{JPEG} - PSNR_{JPEG}),\\
&(DMOS_{JPEG} - SSIM_{JPEG}),\\
&(DMOS_{JPEG} - FSIM_{JPEG}),\\
&(DMOS_{JPEG} - VSI_{JPEG}))|
\end{aligned}
$$

Equation 8to find the most accurate quantitative metric for JPEG compressed images

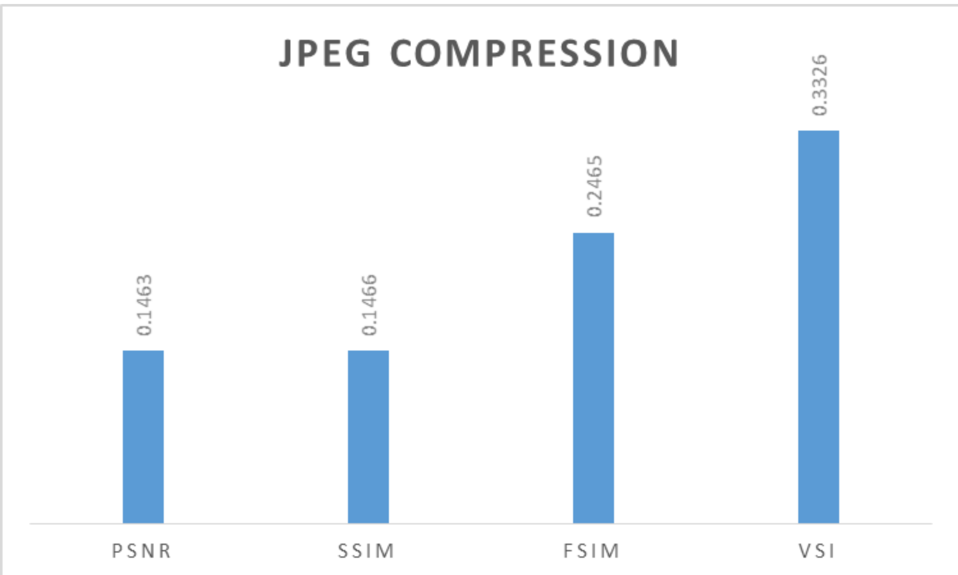

Figure 4.5Differences obtained for each quantitative method with DMOS of JPEG compression

The minimum value is given by the most accurate quantitative method and it is said to be most efficient to measure the quality of JPEG compressed images. The above figure represents the values obtained by above equation, where it is evident that PSNR gives the least difference with DMOS. That is, PSNR is the most efficient quantitative metric to measure the quality of JPEG compressed images.

4.2.4.3 White Gaussian Noise

This minimum difference can be calculated by the following equation.

$$
\begin{aligned}
|MIN\ ((&DMOS_{WN} - PSNR_{WN}),\\
&(DMOS_{WN} - SSIM_{WN}),\\
&(DMOS_{WN} - FSIM_{WN}),\\
&(DMOS_{WN} - VSI_{WN}))|
\end{aligned}
$$

Equation 9to find the most accurate quantitative metric for white Gaussian noise images

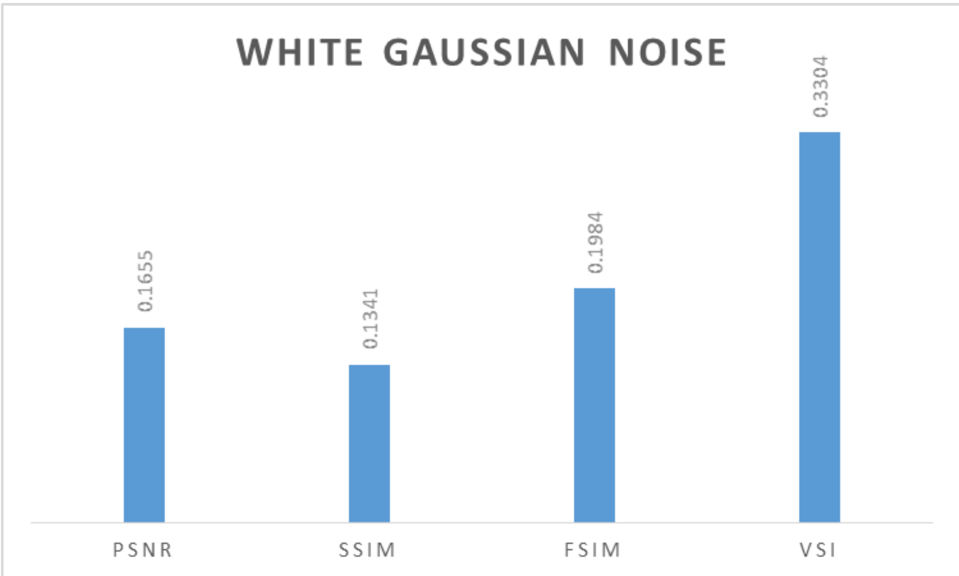

Figure 4.6Differences obtained for each quantitative method with DMOS of White Gaussian noise

The value obtained by applying the above equation is resulted by the most accurate quantitative method and is the most efficient method to measure the quality of white Gaussian noised images. In order to find the method that gives minimum difference, a figure that represents a graph with differences from each method with DMOS is presented. From the above graph it is seen that SSIM gives the least difference. That is, SSIM is the most efficient quantitative method to measure the quality of white Gaussian noise images.

4.2.4.4 Gaussian Blur

The quantitative metric that gives closest result to subjective evaluation of Gaussian blurred images can be found by calculating the differences from each metric to subjective evaluation. The metric that gives the least difference is said to perform in a consistent manner to the subjective evaluation compared to other metrics. These differences can be calculated using the following equation.

$$
\left| \mathrm{MIN}\left((DMOS_{GB} - PSNR_{GB}),\ (DMOS_{GB} - SSIM_{GB}),\ (DMOS_{GB} - FSIM_{GB}),\ (DMOS_{GB} - VSI_{GB}) \right) \right|
$$

Equation 10to find the most accurate quantitative metric for Gaussian blur images

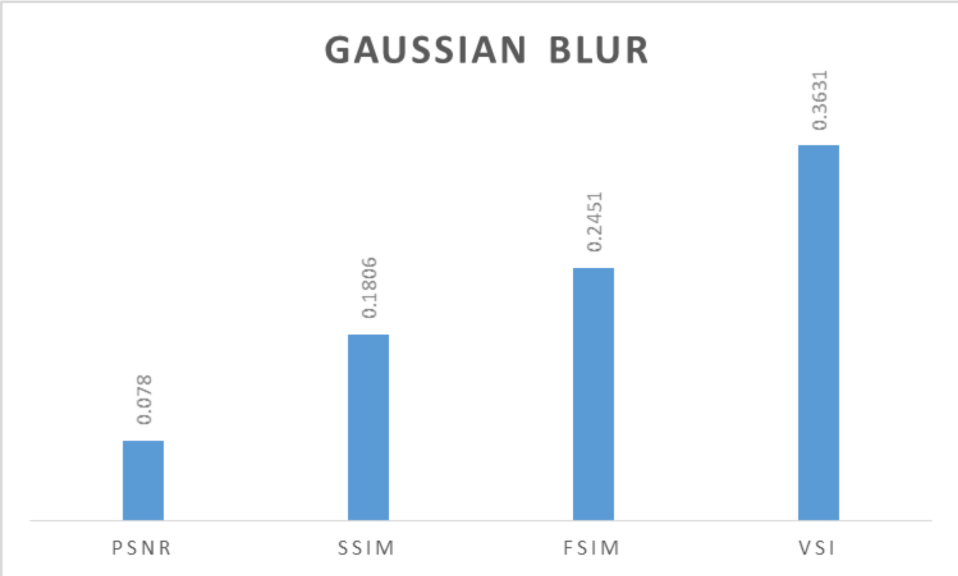

Figure 4.7Differences obtained for each quantitative method with DMOS of Gaussian blur images

The above equation results in the minimum difference of the quality measure from the DMOS value. It is represented through a graph in the figure above. Thus, the quantitative method that gives this minimum difference is the closest one to DMOS. Hence it is the most accurate method to measure the quality of Gaussian blur images. Least difference is given by PSNR. Hence, PSNR is the most efficient quantitative method to measure the quality of images distorted due to Gaussian blur.

5 DISCUSSIONS

This section includes discussion on considered distortion types, quantitative metrics that corresponds to the study. It also contains discussions on findings from the experiment. Also, The threats to validity, limitations of the study and finally addressing the answers to the research questions.

JPEG 2000 is an image compression standard that has been created by the Joint Photographic Experts Group in 2000. It has been created with the idea of replacing the original JPEG by improving flexibility and scalability of the standard. Often, compression or any kind of image processing application results in an image that is not original but with changes in its properties. That is, in case of compression the size of the image is reduced in order for multimedia transmission and storage limitations. White Gaussian noise type of distortion is caused mostly during image acquisition that is due to poor illumination or transmission due to circuit noise. Due to these reasons, if the image is affected the type of distortion is called Gaussian noise. When an image is distorted due to an image processing application, the quality of the image is degraded. In order to reduce the distortion such as noise, a Gaussian function is applied to the image. The resultant image is the Gaussian blurred image that is not the original image but is blurred by the application of a Gaussian function.

As this is the case, for each kind of image distortion there is a specific reason for the quality degradation. Also, to measure the quality of these resultant distorted images various image quality assessment methods are employed. For our research, we have chosen five of the objective quantitative quality measurement methods to measure the quality of the dataset of images with above mentioned type of distortions. They are first, PSNR which is the traditional, simple and inexpensive quality measurement method that measures the distorted image quality with respect to the original image by comparing the maximum power of the signal in each case. Second, SSIM which measures the image quality with respect to the original image by comparing the structural similarity properties. Third, FSIM which is also a full reference objective image quality measurement method that measure quality based on the feature similarity properties of the original and distorted images. Fourth, UQI measures the image quality based on the combination of factors like loss or correlation, luminance and contrast distortion. Fifth, VSI which measures the image quality based on the salient features as humans percept to these more significantly. Hence, all these selected methods were applied on the mentioned distortion types by preparing a database according to the experiment designed. The quality measure obtained by each metric for each distortion type are tabulated in the results section and are analyzed.

5.1 Discussion on findings from experiment

By observing the results obtained and analyzed it is once again proved that SSIM measures the quality of a distorted image accurately as it results in a quality measure that is much closer to the DMOS. As DMOS is a subjective quality measurement technique which is obtained directly from human subjects, a quality measure that is closer to the DMOS is said to be highly accurate when compared to the other metrics. Likewise, PSNR also resulted in a quality measure that is nearer to the DMOS value for some distortion types. It is interesting to see that PSNR has accomplished over the metrics like FSIM, VSI in measuring the quality of the distorted image. This may be because the dataset contains distorted images with high maximum power of signal or low salient and similarity features that can be measurable. However, for the experiment conducted with the dataset prepared (which although highly influenced the results), the results are valid to a great extent as long as the database is considered.

It is intriguing to observe that none of the objective quantitative methods gave a quality measure that is same as subjective evaluation. This may be because each image quality assessment technique evaluates the image differently and they cannot be similar. But it is important to evaluate the quality measurement metrics that can closely correlate with the human visual system as the knowledge acquired by the human visual system judges the image on the whole spatial domain.
In order to analyze the results obtained from the experiment we used statistical approach which involved calculation of difference of quality from each quantitative method to the standard DMOS value for each distortion type.

5.2 Threats to validity

- ***Internal Validity*** A literature review and an experiment were conducted in this research. Possibilities of effecting internal validity may arise due to wrongly interpreting these research methods.
  - The metrics chosen to perform the experiment were selected keenly based on their behavior as each one measures the quality using different parameters and as per the literature these are said to give competitively accurate results consistent to perceptual visual quality.
  - The quality of the images is calculated using Matlab which is so far gave good interpretation for image processing applications [27]. And the means of quality measure given by each metrics are compared to subjective evaluation. Also, these values are normalized to a range of 0-1 using unity based normalization in order to compare in an even basis.

- ***External Validity***

The image dataset used for this study is prepared only from one database which is LIVE image database which is a publicly available image quality assessment research database. There could be a possibility of threat to external validity as the study might not be generalized for other databases. But, the database was chosen as it contains different kinds of image distortions. And as the main idea was to see how differently these quantitative metrics perform on differently distorted. The study can be generalized as long as the distortions used for this research are considered.

- ***Construct Validity***

The research was carried out by conducting an experiment using inputs from literature review. The findings from the literature review might influence the selection of specific quality metrics as it was a huge review involving subjective and objective quality measurement metrics, image processing applications and its types of distortions. Also, there is a possibility of researcher bias in choosing these metrics. There may be a scenario where the method used for experimenting resulted in biased measures as it is an external software. The results obtained from the analysis might be misinterpreted by the researcher in concluding the research. But, these are the most preferable methods of research and analysis that correspond to research design as each one is justified sufficiently in respective sections. Hence, construct validity is achieved.

- ***Statistical Conclusion Validity***

The conclusions made are highly influenced by the research method used. The results drawn and analyzed may be misinterpreted while concluding the study. The results obtained from the experiment are analyzed to draw conclusions to this research. The method of analysis used is the most crucial to find conclusions where there is a possibility of selecting an inappropriate analysis method. A statistical test which is found to be most appropriate corresponding to dependent and independent variables and the type of data was chosen to draw conclusions from the results.

5.3 Limitations

- ***Literature Review:***

Only five image quality assessment methods were considered in the experiment to compare which were selected by a literature review. There is a possibility of getting results only in comparison within these selected methods. There may be other quantitative metrics that perform better than these methods for a particular type of distorted image. Also there is a chance of a biased selection of metrics on which the experiment is conducted. These reasons might affect the results obtained for each image processing problem (type of distortions).
Argument: Due to limited time and scope of the research, the experiment was conducted only on five of the metrics chosen form the literature review. The main idea was to see how differently a metric performs on different types of distorted images. These five metrics chosen are the most commonly used and are said to give competitive results according to the literature based on their performance.

- ***Experiment Design:***

- Quality of all the images present in the database is measured using the five chosen methods. Where, the database consists of all RGB images for each distortion type. Nevertheless, quality is even measured for grayscale images using the ***rgb2gray (color-_img)*** function in MATLAB. But, efficient quality metric for grayscale images was not analyzed as the DMOS values for the distortions types are not available. As, DMOS values are the standard scores that are to be compared with the quality measure obtained for each metric if they are performing well for each distortion type.
  Argument: This is because the DMOS values for grayscale images were not made available in the LIVE database and obtaining these values is beyond the scope of this research. The quantitative metrics on which the experiment was conducted on were compared with the subjective evaluation for RGB images. Hence, the desired validation for RGB images is achieved.

- Universal Image Quality Index (UQI) is calculated only for grayscale images and hence could not be included for analyzing in comparison with the other quantitative metrics for each distortion type correlating to subjective evaluation
  Argument: For computing quality index for color images an improved version called CIQI has been proposed [36]. It was encountered during the study which is left as a limitation as it would make the study complex.

- ***Dataset used for experiment:***

The experiment conducted to acquire the quality of images taken from the LIVE image database which is an image quality assessment research database. There are other image databases which have been considered in order to validate the results in a broader way. The experiment was limited to just one database.
Argument: One of the research questions (R.Q.2) was to compute the quality of the standard image set with different kinds of distortions. Computing the quality for more number of

databases or other images and reporting the results would not be feasible within the given scope of the research. However, the idea was to compare and analyze which among the selected metrics perform better for each distortion type which is still achieved within a smaller scope by limiting to one database.

5.4 Answering Research questions

R.Q.1 Which quantitative methods have been used to measure the image quality?

Answer:

Several FR IQA methods are being employed to measure the quality of images. By conducting a literature review, few among them are identified which are MSE, PSNR, SSIM, FSIM, UQI, VIFC, VSI. Each metric quantifies the quality of an image using different parameters, these have been tabulated in section 3.1.3. From these identified metrics, 5 of the quantitative measurement methods were chosen to carry out this research which are PSNR, SSIM, FSIM, UQI, and VSI. These are chosen as each one exhibits different behavior in the way of calculating the quality and as the literature are known to perform competitively better compared to few other metrics.

R.Q.2 Compute the quality of the selected standard test image dataset (image pair) using each of the quantitative methods identified in R.Q. 1

Answer:

The quality of the distorted images in the dataset is calculated using the 5 selected metrics with respect to their original images. Quality is calculated using Matlab. The quality measures obtained from each quantitative metric for each type of distortion is tabulated in section 4.1.1.

R.Q.2.1 Which quantitative methods on the image dataset performs most efficiently?

Answer:

The quality measures given by each quantitative metric are analyzed to find the method that performs efficiently. Here, a metric is said to perform efficiently if it performs in a consistent manner with the subjective evaluation. As the dataset consists of images with kinds of distortions, the quantitative metrics are evaluated to find on which kind of distortion image set each metric performs efficiently. The result is reported as follows.

QUANTITATIVE METHOD	DISTORTION TYPE
PSNR	Gaussian Blur
SSIM	JPEG 2000 compression
FSIM	White Gaussian Noise
VSI	White Gaussian Noise

Table 5-1 Efficient performance of each metric corresponding to different distortions in the dataset

R.Q.3 Which quantitative methods can be more suitable for different image processing problems?

Answer:

The selected quantitative image quality measurement metrics which were compared based on their performance on images with different types of distortions due to different image processing applications, perform differently for each kind of distorted image set used in this research. Based on the analysis done on the results from experiment where a suitable or an efficient metric is the one that gives closer correlation with subjective evaluation of the particular distortion type image set. The efficient metric to measure the quality for each image distortion type are tabulated as follows.

DISTORTION TYPE	EFFICIENT QUANTITATIVE METRIC
JPEG 2000 compression	SSIM
JPEG compression	PSNR
White Gaussian noise	SSIM
Gaussian Blur	PSNR

Table 5-2Efficient quantitative metric for each distortion type of images for the database on which the experiment was conducted

6 CONCLUSION AND FUTURE WORK

In this study five different quantitative image quality assessment methods were applied on four different types of distortion image sets. The quantitative methods of image quality included are PSNR, SSIM, FSIM, UQI and VSI. And, the distortion types considered are JPEG 2000 compression, JPEG compression, white Gaussian noise and Gaussian blur. These distorted images were obtained from the publicly available database, LIVE image quality assessment database and a database is prepared according to the design of the experiment. Each quantitative method performs differently for each distortion type. So, an experiment was conducted to evaluate the performance of each of the quantitative image quality measurement methods for each distortion type. Their performance was evaluated by comparing the results with the standard subjective evaluation DMOS of each distortion type that are made available in the LIVE database. According to the analyzed results SSIM performs efficiently for JPEG 2000 compression and white Gaussian noised images when compared to the other quantitative methods. PSNR performs efficiently for JPEG compression and Gaussian blur images when compared to other quantitative methods. Further, the most suitable quantitative quality measurement method for each type of distortion are presented.

6.1 Future Work

This thesis can be extended towards finding efficient quality measurement method for grayscale images by obtaining their DMOS values. Further, it would be interesting to know which quantitative measurement methods work better for other image processing applications like contrast enhancement, noise reduction etc. That is for other types of distortions caused due to other image processing applications. Also, implementing other existing and available quantitative measurement methods and providing a more informative cluster of suitable metrics would be of added knowledge. This would provide a large scope to the future researchers to add contribution to this research.

REFERENCES

[1] A. M. Eskicioglu and P. S. Fisher, “Image quality measures and their performance,” *Commun. IEEE Trans. On*, vol. 43, no. 12, pp. 2959–2965, 1995.

[2] T. S. Huang and K. Aizawa, “Image processing: Some challenging problems,” *Proc. Natl. Acad. Sci.*, vol. 90, no. 21, pp. 9766–9769, 1993.

[3] R. C. Gonzalez and R. E. Woods, “Digital image processing,” *Nueva Jersey*, 2008.

[4] M. S. Corson and R. H. Moss, “Computer vision: An in-depth look at the technique and applications of digital image processing,” *IEEE Potentials*, vol. 5, no. 4, pp. 31–34, Dec. 1986.

[5] Z. Wang, A. C. Bovik, and L. Lu, “Why is image quality assessment so difficult?,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, vol. 4, pp. IV–3313–IV–3316.

[6] V. L. Jaya and R. Gopikakumari, “IEM: a new image enhancement metric for contrast and sharpness measurements,” *Int. J. Comput. Appl.*, vol. 79, no. 9, 2013.

[7] D. Wolin, K. Johnson, and Y. Kipman, “The importance of objective analysis in image quality evaluation,” in *NIP & Digital Fabrication Conference*, 1998, vol. 1998, pp. 603–606.

[8] A. Lahouhou, E. Viennet, and A. Beghdadi, “Combining and selecting indicators for image quality assesment,” in *Information Technology Interfaces, 2009. ITI’09. Proceedings of the ITI 2009 31st International Conference on*, 2009, pp. 261–266.

[9] S. Kumari, V. S. Meel, and R. Vijay, “Image quality prediction by minimum entropy calculation for various filter banks,” *Int. J. Comput. Appl.*, vol. 7, no. 5, 2010.

[10] J. Li and W. Dai, “Image quality assessment based on the correlation coefficient and the 2-D discrete wavelet transform,” in *Automation and Logistics, 2009. ICAL’09. IEEE International Conference on*, 2009, pp. 789–793.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[12] N. Thakur and S. Devi, “A new method for color image quality assessment,” *Int. J. Comput. Appl.*, vol. 15, no. 2, pp. 10–17, 2011.

[13] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *Image Process. IEEE Trans. On*, vol. 15, no. 2, pp. 430–444, 2006.

[14] A. Hore and D. Ziou, “Image quality metrics: PSNR vs. SSIM,” in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 2366–2369.

[15] “YCbCr,” *Wikipedia, the free encyclopedia*. 06-Sep-2016.

[16] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, “Subjective and objective quality assessment of image: A survey,” *ArXiv Prepr. ArXiv14067799*, 2014.

[17] N. Sharma, I. Garg, P. K. Sharma, and D. Sharma, “Analysis of Transform Techniques for 2D Image Compression,” *Int. J. Eng. Innov. Technol. IJEIT*, vol. 1, no. 5, May 2015.

[18] J. Silvestre-Blanes, “Compress-image quality measures in image-processing applications,” in *2009 IEEE Conference on Emerging Technologies & Factory Automation*, 2009, pp. 1 – 4.

[19] R. Matsuoka, M. Sone, K. Fukue, K. Cho, and H. Shimoda, “Quantitative analysis of image quality of lossy compression images,” *entropy*, vol. 3, no. 4.13, pp. 4–06, 2004.

[20] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2013.

[21] B. Kitchenham *et al.*, “Systematic literature reviews in software engineering – A tertiary study,” *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792 – 805, 2010.

[22] C.-H. Chen, Y. Yao, D. L. Page, B. Abidi, A. Koschan, and M. Abidi, “Objective Image Quality Evaluation for JPEG, JPEG 2000, and Vidware VisionTM,” in *Pacific-Rim Symposium on Image and Video Technology*, 2006, pp. 751–760.

[23] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: a feature similarity index for image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.
[24] L. Zhang, Y. Shen, and H. Li, “VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment,” *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.
[25] E. C. Larson and D. M. Chandler, “Categorical image quality (CSIQ) database,” *Online Httpvision Okstate Educsiq*, 2010.
[26] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE image quality assessment database release 2*. 2005.
[27] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, “TID2008-a database for evaluation of full-reference visual quality assessment metrics,” *Adv. Mod. Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
[28] C. Zuria, J. Ramirez, D. Báez-López, and G. E. Flores-Ventad, “MATLAB based image processing lab experiments,” in *Frontiers in Education Conference*, 1998, vol. 3, pp. 1255–1258.
[29] “Tests of Significance.” [Online]. Available: http://www.stat.yale.edu/Courses/1997-98/101/sigtest.htm. [Accessed: 06-Oct-2016].
[30] M. Marusteri and V. Bacarea, “Comparing groups for statistical differences: how to choose the right statistical test?,” *Biochem. Medica*, pp. 15–32, 2010.
[31] “Testing of Assumptions,” *Statistics Solutions*. .
[32] “IBM SPSS - IBM Analytics,” 17-Mar-2016. [Online]. Available: http://www.ibm.com/analytics/us/en/technology/spss/. [Accessed: 06-Oct-2016].
[33] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *Advances in neural information processing systems*, 2006, pp. 545–552.
[34] “Can we detect visual saliency on gray scale image/video?” [Online]. Available: https://www.researchgate.net/post/Can_we_detect_visual_saliency_on_gray_scale_image_video. [Accessed: 09-Oct-2016].
[35] P. Sharma, “Towards three-dimensional visual saliency,” 2014.
[36] A. Medda and V. DeBrunner, “Color Image Quality Index Based on the UIQI,” in *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, 2006, pp. 213–217.
[37] “normalization - How to normalize data to 0-1 range? - Cross Validated.” [Online]. Available: http://stats.stackexchange.com/questions/70801/how-to-normalize-data-to-0-1-range. [Accessed: 09-Oct-2016].
[38] Nisha and Sunil Kumar, “Image Quality Assessment Techniques,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 7, Jul. 2013.