A Novel Full Reference Metric for Image Quality Assessment
Based on Human Vision System

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science and Engineering

Submitted By
Sumit Mittal
(Roll No. 800932023)

Under the supervision of:
Dr. Seema Bawa
Professor, Computer Science & Engineering
Dean (Student Affairs)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

June 2011

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "***A Novel Full Reference Metric for Image Quality Assessment Based on Human Vision System***", in partial fulfillment of the requirements for the award of degree of Master of Engineering in Computer Science and Engineering submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Seema Bawa and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Sumit Mittal
Signature:

(Sumit Mittal)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Seema Bawa 01/07/2011

Dr. Seema Bawa
Dean of Student Affairs
Professor, Computer Science and Engineering Department
Thapar University
Patiala

Countersigned by

(Dr. Maninder Singh)
Head 6/7/11
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. K. Mohapatra)
Dean (Academic Affairs)
Thapar University
Patiala

ACKNOWLEDGEMENT

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life.

This work would not have been possible without the encouragement and able guidance of my supervisor Dr. Seema Bawa. I thank my supervisor for their time, patience, discussions and valuable comments. Their enthusiasm and optimism made this experience both rewarding and enjoyable.

I am equally grateful to Dr. Maninder Singh, Associate Professor & Head, Computer Science & Engineering Department, a nice person, an excellent teacher and a well credited researcher, who always encouraged me to keep going with work and always advised me with his invaluable suggestions.

I will be failing in my duty if I don't express my gratitude to Dr. Abhijit Mukherjee, Director of the University, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection, which made my stay at Thapar University memorable.

Last but not least, I would like to thank my family whom I dearly miss and without whose blessings none of this would have been possible. To my parents, I own thanks for their wonderful love and encouragement. I would also like to thank my sister, since she insisted that I should do so. I would also like to thank my close friends for their constant support.

ABSTRACT

With the tremendous growth in field of digital image processing, there is a need to store and maintain digital images efficently. Digital images are captured, processed, compressed, stored and transmitted through various devices. During these stages of processing, there is a need to maintain quality of the images because various distortions may take place and affect the quality of the image. To assess the quality of an image there are subjective and objective methods available. Subjective methods are generally time consuming and difficult to use in autonomous systems. Objective image quality assessment methods are more used these days to evaluate the quality of the image. These objective image quality assessment methods are based on the situation that whether the reference image is available, partially available or just not available. Today, most of evaluations of images are based on the assumption that reference image is available this is called full reference method. There are many metrics based on full reference approach to determine the quality of an image. Some metrics are mathematical like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) etc. These metrics are not able to correlate well with human perception. Some metrics are based on Human Vision System (HVS) like Structural Similarity Index (SSIM), Visual Image Fidelity (VIF), Visual Discrimination Model (VDM), Perceptual Difference Method (PDiff) etc. These metrics also have some limitations in evaluating the quality of the image. There are many full reference metrics implemented for assessing the quality of an image based on high level and low features of an image based on HVS. These are implemented by different metrics. In this thesis effort is made to implement a new full reference metric by integrating the human vision based metrics including SSIM, VIF, VDM and PDiff. Further the quality of image has been tested on available Laboratory for Image & Video Engineering (LIVE) Image Quality Assessment Database. The quality index of the image has been observed and reported in this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

1.1 Image Quality

Recent advances in digital imaging technology, computational speed, storage capacity and networking have resulted in the proliferation of digital data, both images and videos. As the digital images are captured, stored, transmitted, and displayed in different devices, there is a need to store and maintain images efficently [2].

Image quality is a characteristic of an image that measures the perceived image degradation. Imaging systems may introduce some amounts of distortion or artifacts in the signal that leads to problem in assessing the quality of an image. This allows to find out various methods available to assess the quality of an image [2].

Image Quality is defined in many ways as described below:

Engeldrum defined image quality as “Image quality is the integrated perception of the overall degree of excellence.” This makes a minimal number of assumptions and restrictions. It does not take into account the context or application. It neither restricts the image quality to be dependent on the observer’s relation to the image; photographer or end-user, expert or non-expert [10].

According to Camera Phone Image Quality (CPIQ) group of the International Imaging Industry Association (I3A) :
“Image quality is the perceptually weighted combination of all visually significant attributes of an image when considered in its marketplace or application.” It suggests that image quality is dependenet on the application. Although Engeldrum’s proposal for general image quality can be seen as an ultimate goal, application specific image quality is a more realistic approach from the viewpoint of fully computational models.

Keelan defined "the quality of an image as to be an impression of its merit or excellence, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted."

Keelan divides quality assessment into three types: first, second and third-party.

The first-party quality assessment refers to an image evaluation by photographer who took the image. The second-party quality assessment is made by a person involved in the subject of the image and the third-party assessment covers the individuals who are not involved in the picture taking or subject. Keelan's definition of image quality focuses only on the third-party assessment. The main reason for this stems from the practical viewpoint of performing subjective tests. Unlike in the first and secondparty assessments, in the third-party assessment the same image samples can be shown to all observers and the observers need not to meet certain requirements [12].

Janssen and Blommaert defined the quality of an image to be the degree to which the image is both useful and natural.

The usefulness of an image is defined as to be the precision of the visual representation of the image, and the naturalness is defined as as the degree of correspondence between the visual representation of the image and knowledge of reality as stored in memory.

The processing of image information is divided into three different processes:

i. Perception: forming a internal representation of the image using mainly low level knowledge of the visual world.
ii. Interpretation: matching the internal representation to memory representations.
iii. Semantic: processing of interpretation based on the task [11].

1.2 Aspects of Image Quality

The concept of image quality can be approached by first defining good quality.
According to Johansson good quality has always two sides: subjective opinions and an image itself.
Subjective opinions are affected by complex physical and psychological parameters, while image

goodness is simpler to define. For a good quality image, properties like optimal photography, technical excellence, and natural color reproduction are required. Furthermore, when assessing many images at the same time, the balanced and equal output of the images is essential for good quality experience.

Janssen expands the idea of a good image by presenting two requirements for visual information in the image: precision and reliability. For meeting the requirements, Janssen emphasizes the importance of successful image interpretation by an observer, leading thoughts to a secured observation situation and the suitable high-level characteristics of the observed image. For general image quality, Janssen lists two defining measures: image usefulness and naturalness, where usefulness refers to the precision of the internal representation of an image and naturalness to the correspondence between the internal precision and observer's knowledge of reality.

Ridder and Endrikhovski, who divide the quality of a reproduced image into three categories: fidelity, usefulness, and naturalness.

With the fidelity category, Ridder and Endrikhovski refers to the reproduction accuracy of an observed image in comparison to the original, perfect image.

The second category, usefulness, indicates image suitability for the designed task, while naturalness category is defined similarly as in Janssen , the match between a reproduced image and the mental impression of an observer, affected e.g. by memory traces [18].

According to Silverstein and Farrell, image fidelity refers to the ability to discriminate two images from each other. An image with high fidelity is reproduced accurately, without any visible information loss or distortion. Image quality, on the other hand, explains an ability to form preferences between images. While subjective conditions affect both the cases, image quality observation is based on the own preferences and impressions of an observer [19].

1.3 Image Quality Attributes

To describe the quality of an image, specific image quality attributes are needed.

Leisti divide quality attributes to two categories: low-level and high-level.

For low-level attributes, e.g. colorfulness and sharpness, there exists a physical, measurable counterpart in quality evaluation. [2].

High-level attributes, e.g. naturalness and usefulness, are strongly connected to observer's visual perception and experience from image content, making them harder to quantify. There lies, however, an interesting connection between high and low-level attributes.

High-level attributes guide an observer in quality sensation by clarifying the meaning of low-level attributes for general quality. Examples for high-level attributes like naturalness, clearness, realism, usefulness, and depth.

1.4 Image Quality Assessment

Image Quality assessment (IQA) plays a fundamental role in the design and evaluation of imaging and image processing systems. Quality assessment techniques can be used to systematically evaluate the performance of different image compression algorithms that attempt to minimize the number of bits required to store an image, while maintaining sufficiently high image quality .

The goal of image quality assessment is to supply quality metrics that can predict perceived image quality automatically. Very often the quality of an image needs to be quantified. This can be done by subjective testing sessions, or by objective computational metrics [6].

In past, many vision functions have been used to understand how the Human Vision System responds to the visual world. Among many vision functions, comparison is a natural capability that everyone takes for granted. With this ability, people can notice the differences between two similar images. Training a system to perceive the differences in a human manner is frustrating.

There are several reasons that limited success has been achieved.

i. First, image comparison by a human entity is a subjective action. The judgment will be acted by individual physiological limits, observation behavior, personal preference, and knowledge limitation. All of these imply that simple distance metrics are not able to evaluate the perceptual differences.

ii. Second, distortion variety poses difficulties for subjective evaluation. Research shows that observers could become familiar with the stimuli (such as image distortion) in order to make judgment results more repeatable. But interpretation of certain distortions with varied image contents is highly subjective. Also people usually do not quantify the difference but evaluate the difference in a fuzzy manner.

iii. Third, accumulating individual perceptible differences and combining them is a subjective process.

The whole procedure of quantication and accumulation to a single score is referred to as image quality assessment. There has been so much effort done in developing reliable methods for evaluating image quality. Because quality assessment metrics are widely used in every image processing area where human perception is involved.

1.5 Subjective and objective quality

Fundamentally, image quality is always an outcome from human sensation. Human observers make the final decisions about quality based on their own visual preferences that naturally are not only affected by the psychophysical aspects of the observer, but also by the fidelity of the image and the observation situation. For evaluating image quality, testing with human observers, subjective evaluation is often considered the most reliable way to estimate the quality of images.

From subjective evaluation measures, the Mean Opinion Score (MOS) is the most widely used.

MOS is often regarded as the most reliable image quality measure, but it requires numerous human observations and a specific test arrangement. MOS is also slow and expensive method in real world situations [1].

Another option for quality evaluation, objective assessment, relies on computational models that can predict the image quality observations of humans.

According to Wang and Bovik, an accurate objective image quality model predicts the image quality sensation of an average human observer. In other words, strong correlations to subjective observations are essential when defining a good objective quality model.

In designing objective quality methods, three types of knowledge can be exploited in development:

i. The knowledge of the original image.
ii. The knowledge of the process causing image distortion.
iii. The knowledge of HVS and the subjective processes affecting quality observation.

Based on the knowledge reference, measures can be described with three commonly used categories:

i. Full Reference
ii. Reduced Reference
iii. No Reference

i. Full-reference methods present the most popular category, describing methods where evaluation is based on the comparison of an original, perfect image and a distorted sample image. The image quality methods desribed in literature are full-reference methods.
ii. In many practical applications, however, an original reference is not available and No Reference methods must be applied. These no reference methods are very complex to develop with software, but for a human, who already has the knowledge what the image should look like quality observation without a reference is a rather uncomplicated task.
iii. In the category between full reference and no-reference methods are reduced-reference methods.

1.5.1 Subjective Image Quality Assessment

The straightforward way of measuring image quality is to consult human opinions because people are the ultimate viewers in most image applications. This is known as subjective image quality assessment. The basic idea of subjective quality assessment is to ask several observers to give their personal judgments for a given image test sample and then compute the mean opinion score (MOS) of the human judgments as the subjective evaluation.

Subjective testing for visual quality assessment has been defined in ITU-R Recommendation BT.500-6. The three most commonly used procedures are:

i. Double Stimulus Continuous Quality Scale (DSCQS)
ii. Double Stimulus Impairment Scale (DSIS)
iii. Single Stimulus Continuous Quality Evaluation (SSCQE)

The databases LIVE use double stimulus continuous quality scale method for subjective testing, since DSCQS is the most reliable method [1]

Several issues should be considered during the subjective quality assessment experiment.

i. First, suffcient observers should be included in the experiment. At least 15 observers are recommended. Even though the observers are urged to base their score only on how different or dissimilar they perceive the images to be, the personal preference effect on evaluation results cannot be totally eliminated. More observers will help to achieve a quality score that most people agree with. Also nonexpert observers and expert observers need to be included in the observer group. A non expert viewer may pay more attention to the larger context, but a trained viewer may focus on the details or evaluate image quality from professional perspective [2].
ii. Second, to conduct appropriate subjective assessment, it is necessary to set up the proper viewing conditions which include display device requirements and viewing manner. For instance, the ratio of luminance of inactive screen to peak luminance should be less than 0.02. The viewing distance is recommended to be 1.5m, which corresponds to a viewing distance/display height ratio of 6. In addition, other procedural elements are worth noting, such as: timing of presenting test samples and selection of samples. These weaknesses result in the request for a fast, computational and reliable quality assessment method. The practicality of such tests is limited since they are time consuming, expensive, difficult to repeat, and require specialised viewing conditions [13].

These problems have resulted in the use of objective quality metrics. Simple objective metrics such as Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are commonly used due to their simplicity and to the absence of any standardised or widely accepted alternative.

1.5.2 Objective Image Quality Assessment

Objective Image Quality Assessment aims to design quality measures that can automatically determine percieved image quality. These quality measures play important roles in a broad range of applications such as image acquisition, compression, restoration and enhancement. To assess the quality of an image using objective methods we have objective quality metrics. Objective image quality metrics can be classified according to the availability of an original image, with which the distorted image is to be compared.There are three ways under which comparison can be done as listed below [1].

i. Full Reference Metric

In full reference metric the reference image is present with which distorted image is to be compared.

ii. Reduced Reference Metric

In reduced reference metric the reference image is partially present , a form of set of extracted features available with which distorted image is to be compared.

iii. No Reference Metric

In No reference metric the reference image is not present with which distorted image is to be compared

1.6 Image Processing

In electrical engineering and computer science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or, a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. It is motivated by two major applications:

i. Image Enhancement: Improvement of pictorial information for human perceptions means whatever image you get we wants to enhance the quality of the image so that image will have better look.

ii. Efficient storage and transmission for example if we want to store the image on our computer the image will need certain amount of space in hard disk so to reduce this space requirement [10].

1.7 Human Vision system

Most digital images are ultimately viewed and appraised by humans. It is therefore important for imaging systems to be designed with consideration of the operation of the HVS. This ensures that only information relevant for the required task is received by the viewer, thus avoiding unnecessary expense in image storage, transmission, and manipulation.. The improved understanding of the operation of the HVS which has been gained since the 1960s, and the increased interaction between engineers and vision scientists has resulted in more widespread usage of HVS properties in digital image processing systems. This trend appears likely to continue as improved visual models are developed, and as more applications requiring computational vision models are recognised. Understanding the operation of the HVS is particularly important for the development of accurate and robust picture quality assessment techniques. The goal of image quality assessment is to simulate human assessment of image dissimilarities**.** Although the full functions of the human visual system, especially high level functions, are not thoroughly investigated, knowledge of the early stage of the human visual system is still helpful in understanding fundamental visual behaviors. Some human vision behaviors are modeled to simulate the human assessment of visual quality.

Four of them are popular for the design of quality assessment [8].

1.7.1 Luminance Sensitivity

In early research, experiments results showed that the human vision system responds to luminance change in a logarithmic way. Consider a light area with luminance of I+4I surrounded by a background with luminance of I. The just noticeable difference 4I is proportional to the background luminance as in Equation

Later research [3] shows that human visual system luminance sensitivity not only depends on the overall background luminance but also the local neighborhood luminance.

Fig 1.1

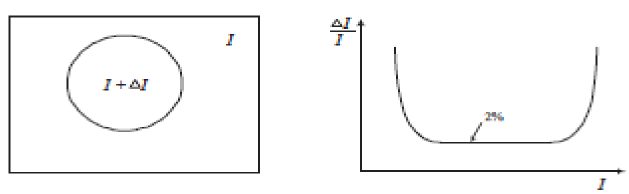


Fig 1.1 Luminance sensitivity- The just noticeable luminance difference 4I is proportional to the background luminance I within a wide range.The experiment results show that the ratio is nearly constant at 0.02.

1.7.2 Contrast Sensitivity

The famous Campbell-Robson chart was proposed to illustrate that the human visual system has different responses to visual signals at dierent frequencies, which is referred to as contrast sensitivity.

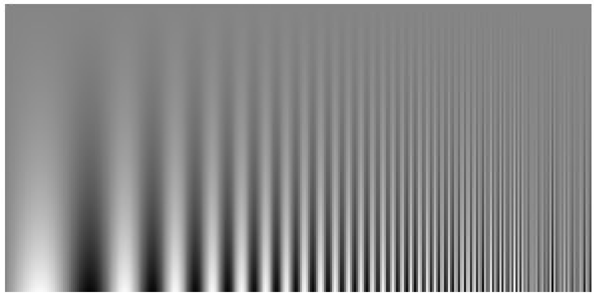


Fig 1.2 Contrast sensitivity chart (Campbell-Robson Chart)

Spatial frequency increases (the bars get thinner) in the horizontal direction and contrast decreases (the black/white bars merge to plain gray area) in the vertical direction [16].

1.7.3 Contrast Masking

Contrast Masking happens when different image components are present at the same spatial location. A simple example below gives a clear illustration. Figure 1.3(a) is a raw image. The top half of Figure 1.3(b) has extra white noise, whereas the bottom half of Figure 1.3(c) has the same amount of white noise. Generally people are more likely to notice the noise in the middle gure than the right. The presence of one component (grass) decreases the visibility of another component (noise) in Figure 2.3(c). Usually, the contrast masking eect is strongest when two components have similar frequency and orientation [65]. Based on this statement, masking is typically modeled within each channel. This vision property is also widely adopted in many quality assessment methods.

a b c

Figure 1.3 Masking effect.

Figure (a) is the raw image.

Figure (b) has extra white noise in the top half.

Figure (c) has the same amount of white noise in the bottom half.

1.7.4 Visual Channel Decomposition

This suggests the visual area functions similarly to filters with orientation and frequency selectivity. Along with contrast sensitivity in different frequencies. Mannos and Sakrison suggested that the human visual system contains a number of independent channels with narrow bandwidths and specific center frequencies.

Chapter 2

Literature Review

There are varoius subjective and objective methods available for assessing the quality of an image.

2.1 Subjective Image Quality Assessment Methods

Subjective methods for digital image quality assessment are defined in ITU-R.BT.500-6.

Four assessment methods are recommended in the standard, with the choice of which particular method to use being left to the experimenter, depending on the purpose of the test

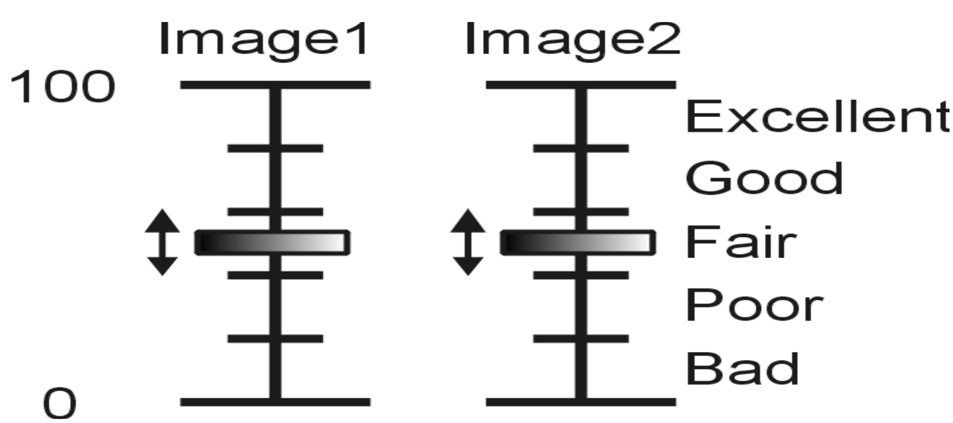
The methods described in ITU-R BT.500-6 are:

i. Double Stimulus Continuous Quality-Scale Method (DSCQS)

Double Stimulus Continuous Quality-Scale Method (DSCQS) is useful when it is not possible to provide stimulus test conditions that exhibit the full range of the quality.

Eg,.Two versions of each picture are presented. The observers are asked to assess the overall quality of each picture by inserting a mark on a continuous vertical scale [1].

Fig. 2.1 Double Stimulus Continuous Quality-Scale Method



The results are analysed as follows: Positions on the vertical scale are converted to normalized scores in the range 0 to 100. Each pair of scores is then converted to rating difference. The overall difference in quality is given as DMOS (differential mean opinion score), which is computed as the mean value the differences from all observers related to one image pair. The higher the DMOS, the more distortion in the image is visible [3].

The continuous quality scale allows greater precision in judgments to be made by the observers. This kind of test is best suited to situations where the impairments are relatively small.

ii. Double stimulus impairment scale method

This method is intended for images or sequences which cover a wide range of impairments. Subjects are shown the original picture followed by the impaired picture and are asked to rate the quality of the impaired picture with respect to the original. Results are indicated on a discrete ve-grade impairment scale: imperceptible, perceptible but not annoying, slightly annoying, annoying, and very annoying.

iii. Single stimulus methods

These techniques are useful when the effect of one or more factors needs to be assessed. The factors can either be tested separately or can be combined to test interactions. Subjects are shown the test image or sequence and are asked to rate its quality. A number of different ways of recording observer results are possible. These include: a discrete ve-grade impairment scale (as per the double stimulus impairment scale method); an 11-grade numerical categorical scale; continuous scaling, where the subject indicates the score on a line drawn between two semantic labels; and numerical scaling, where subjects are free to choose their own numerical scale, thus avoiding scale boundary effects.

iv. Stimulus comparison methods

When two impaired images or sequences are required to be compared directly, these techniques are most useful. Subjects are shown the two distorted scenes in a random order and are asked to rate the quality of one scene with respect to the other.

Subjects' opinions are typically recorded using adjectival categorical judgments, where the subject compares the scenes using a discrete seven level scale (much worse, worse, slightly worse, the same, slightly better, better, and much better).

Continuous scales or performance methods may also be used [4].

2.2 Subjective Image Quality Factors

The Rec. BT.500-6 standard also species several other features to be considered in the testing, which are listed below.

Number of subjects: at least 15, and preferably more. They should have normal or corrected-to-normal vision, and preferably should be non-expert.

Test scenes: these should be critical to the impairment being tested. Although useful results can still be obtained using only two dierent pictures, a minimum of four is recommended for most tests.

Viewing conditions: specications have been established for the room environment, monitor, ambient lighting conditions, viewing distance, and viewing angle.

Stimulus presentation: random presentation of sequences is recommended, with no sequence being displayed twice in succession. Test sessions should not last longer than half an hour.

There are several problems associated with performing subjective quality tests. [5].

i. Subjective tests are extremely time consuming and costly. Many groups that research compression techniques do not possess the required laboratory viewing conditions, and must either perform their testing at other locations or conduct tests under non-standard conditions. It is also difficult to obtain a large number of subjects to perform the tests. The process of subjective testing may take in the order of weeks or months, thus becoming a major bottleneck in the development of a coder. In order to obtain accurate results, a large number of subjects are required. This is because there can be a large variation in individual viewing opinions, depending on the subject's age, sex, experience, motivation,or other personal factors.
ii. Subjective test data is only valid for the particular set of viewing conditions that the test was performed under. New tests need to be done to extend the results to other conditions.

2. 3 Objective Full Reference Metrics For Image Quality Assessment

The most widely used full reference objective metrics for image quality assessment metrics used are:

i. Mean Squared Error (MSE) & Peak Signal To Noise Ratio (PSNR)

The Mean Squared Error (MSE) has been the dominant quantitative performance metric in the field of signal processing. It remains the standard criterion for the assessment of signal quality and fidelity.

The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of similarity/ fidelity or, conversely, the level of error/distortion between them. Usually, it is assumed that one of the signals is a pristine original, while the other is distorted or contaminated by errors.

Suppose that **x**= {*xi*|*i*= 1,2,· · ·, *N*} and **y**= {*yi*|= 1,2,· · ·, *N*} are two finite-length, discrete signals (e.g visual images where *N* is the number of signal samples (pixels, if the signals are images) and *xi* and *yi* are the values of the *i*th samples **i**n x and y, respectively.

The MSE between the signals is [6].

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2.$$

In the MSE, we will often refer to the error signal
ei = xi − yi which is the difference between the original and distorted signals.
MSE is often converted into a peak signal-to-noise ratio (PSNR) measure
PSNR = 10 log10 (L*L)/MSE

where L is the dynamic range of allowable image pixel intensities.

The PSNR is useful if images having different dynamic ranges are being compared, but otherwise contains no new information relative to the MSE [4].

Features of MSE

I. It is simple. It is parameter free and inexpensive to compute, with a complexity of only one multiply and two additions per sample. It is also memoryless- the squared error can be evaluated at each sample, independent of other samples.

II. The MSE is additive for independent sources of distortions.

III. Finally, the MSE is widely used simply because it has been employed extensively for optimizing and assessing a wide variety of signal processing applications, including filter design, signal compression, restoration, reconstruction, and classification [6].

Failure of MSE and PSNR

I. These metrics are not able to correlate well with the human perception quality.

II. Digital pixel values, on which the MSE is typically computed, may not exactly represent the light stimulus entering the eye.

III. Two distorted image signals with the same amount of error energy may have very different structure of errors, and hence different perceptual quality [3].

ii. Weighted Signal to Noise Ratio (WSNR)

A different approach to PSNR was presented: As the human visual system (HVS) is not equally sensitive to all spatial frequencies, a contrast sensitivity function (CSF) is taken into account.

The CSF is simulated by a low pass or band pass frequency filter.

First of all, the difference of the reference and the distorted image is computed. Then the difference is transformed into frequency domain using 2-dimensional fast Fourier transform. The obtained error spectrum is weighted by the CSF resulting in weighted error spectrum. The last thing to do is to compute the power of the weighted error spectrum and the power of the signal (also transformed into frequency domain) [1].

iii. Structural Similarity Index (SSIM)

One recently proposed approach to image fidelity measurement, which may also prove highly effective for measuring the fidelity of other signals, is the SSIM .

This method is based on comparing the structures of the reference and the distorted images.

Let x and y be two non-negative signals corresponding to the reference and distorted

images, and let μx ,μy ,σx ,σy and σxy be the mean of x, the mean of y, the variance of x, the variance of y, and the covariance of x and y, respectively.

Here the mean and the standard deviation (square root of the variance) of a signal are roughly considered as estimates of the luminance and the contrast of the signal respectievly.

the measure called Structural Similarity (SSIM) index is given by [3]

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

Features of SSIM

I. Despite its simplicity, the SSIM index performs remarkably well across a wide variety of image and distortion types. Without much effort, it can be seen that the SSIM scores are much more consistent than the MSE scores relative to visual perception.

II. Luminance shifting and contrast-stretching, which generally does not degrade image structure, lead to very high SSIM values,while noise contamination and excessive compression lead to low SSIM values.

III. SSIM actually takes a variety of forms, depending on whether it is implemented at a single scale over multiple scales , or in the wavelet domain[6].

Drawback of SSIM

I. A drawback of the basic SSIM index is its sensitivity to relative translations, scaling and rotations of an image.

II. It does not provide better results when it is tuned to translation and scaling features to an image.

iv. Complex Wavelet Structural Similarity Index(CWSSIM)

CWSSIM,a wavelet domain version of SSIM, called the complex wavelet SSIM (CW-SSIM) index was developed. It is an alternative to SSIM which is not tuned to scaling and rotation features applied to an image.

The CWSSIM index is also inspired by the fact hat local phase contains more structural information than magnitude in natural images , while rigid translations of image structures leads to consistent phase shifts [4].

Advantage of CWSSIM

CW-SSIM delivers high scores to luminance-shifted, contrast stretched, space-shifted, scaled, and rotated images as compared to SSIM , and low scores to the images containing structural distortions.

v. Visual Image Fidelity

There are two variations of VIF. One is based on Pixel Domain and other one is based on Wavelet Domain.Pixel Based is easier to use as comapred to Wavelet Based VIF [3].

VIF quality

I. The VIF has a distinction over traditional quality assessment methods, a linear contrast enhancement of the reference image that does not add noise to it will result in a VIF value larger than unity, thereby signifying that the enhanced image has a superior visual quality than the reference image.

II. No other quality assessment algorithm has the ability to predict if the visual image quality has been enhanced by a contrast enhancement operation.

III. VIF produces more coorelated values as compared to other objective metrics but its cost with respect to time is very high. To overcome this problem a new quality metric based on full reference appraoch is proposed that is Contast Error Distribution(CED).

vi. Contast Error Distribution

It provides a new way for defining the quality of an distortd images with a quiet simple form and good stability across various types of degradation of an image [7].

vii. Harris Response

The Harris response was proposed by Harris and Stephens [22] for detecting corner points. To detect interest points under luminance variations, the contrast Harris response was presented. The Harris response is computed with three steps. First, an image is filtered by a high pass filter to obtain directional derivative images and . Second, eigenvalues or the determinant and trace of the gradient information matrix are computed to obtain the Harris response. These two steps are accomplished pixel by pixel over the entire pixels.

The Harris response describes the geometric structure of an arbitrary image pixel. For example, if is negative (positive), the pixel is a corner (edge) pixel. Because the structure of an image pixel is changed by the image distortion, is also changed. If IQM is close to one, the image is similar to the reference image, in other words, the image has a good quality. It is necessary for them to be related with the subjective quality assessment, the differential mean opinion score (DMOS), because the goal of quality measurement algorithms is to quantify the subjective quality of an image.

viii. HVS error-based methods

A distorted image can be divided into an undistorted reference signal and an error signal. A typical HVS image quality assessment method is based on the assumption that the loss of perceptual quality is directly related to the visibility of the error signal [11].

These image quality assessment method operate by weighting different aspects of the error signal based on their visibility. The approach was first introduced by Mannos and Sakrison [21].

Other popular HVS error based methods are the Visual Difference Predictor (VDP) and Perceptual

Image Difference (Pdiff) by Daly and the Sarnoff.

I. Visual Discrimination Model

A typical scheme for computing HVS error-based QA algorithms consists of the following steps: preprocessing, contrast sensitivity function (CSF) filtering, channel decomposition, error-normalisation and error pooling [16]. The preprocessing step includes, for example, colour space transforms and low-pass filtering to simulate the point-spread function of the eye optics. In the CSF filtering step, the image is weighted according to the sensitivity of the HVS to different spatial and temporal frequencies. In the channel decomposition step, the image is separated into subbands (channels) using, for example, the discrete cosine transform or a wavelet transform. In the next step, the error (the difference between the reference and input image) is computed for each channel and weighted to convert the errors into units of just noticeable difference (JND). Finally, the errors in different channels are combined into a single scalar using, for example, the Minkowski distance.

The HVS model of VDM is a typical example containing three main steps: amplitude non-linearity, CSF weighting and a series of detection mechanisms. First, each image is passed through a non-linear response function to simulate the adaptation and response of retinal neurons. Second, the images are weighted with the CSF in the frequency domain and converted to local contrast information. Next, the images are divided into 31 channels (five spatial frequency bands combined with six orientation bands and one orientation-independent band) and transformed back to the spatial domain. A masking function is applied to each channel separately, and finally, error pooling is performed to sum up the probabilities of visible differences in all channels to a single map of detection probabilities, which characterises the regions in the input image that are visually different from the original image [16].

II. Perceptual Image Difference

The perceptual Image Difference (PDiff) was originally developed for the needs of image rendering. The method is largely based on Daly's VDP algorithm. Both the reference and input images are first transformed into a CIE L*a*b* colour space, and the CSF is applied. The visual masking is performed similarly to the VDP algorithm, and finally, the difference between reference and input image is computed. The difference is categorised as unperceivable if certain conditions are filled for both lightness and chroma components.

Several limitations of the HVS error-based QA algorithms are listed in [17]. All HVS error-based QA algorithms are founded on two fundamental assumptions: 1) the error visibility is equal to the loss of quality and 2) the vision models derived from psychophysical experiments using simple test patterns are generalisable to the image quality assessment of complex natural images. It is not clear whether these assumptions are correct.

ix. Image quality assessment algorithms using Natural Scene Statistics(NSS)

The other way to approach the quality assessment problem is the statistical viewpoint. Natural scene statistics (NSS) refers to the statistical properties of natural images as a distinction to the statistics of artificial images such as text, paintings or computer generated graphics [8].

x. Information Fidelity Criterion

The information fidelity criterion (IFC) based on the NSS. In the criterion, quality is evaluated by using the NSS and distortion models to findstatistical information shared by the original and input images. The NSS model used is a Gaussian scale mixture in the wavelet domain. Since the NSS and HVS modelling are assumed to be dual problems, some parts of the HVS are already involved in the NSS model of IFC. However, e.g. contrast sensitivity and the point spread function are missing. In visual information fidelity which is an extension of the IFC, the HVS model is added to include these aspects [8].

xi. Universal Image Quality Index

This quality assessment approach does not depend on the images being tested, the viewing conditions or the individual observers. More Importantly, it must be applicable to various image processing applications and provide meaningful comparison across different types of image distortions. This work attempts to develop a new index to replace MSE and PSNR roles. MSE is sensitive to energy of errors instead of structural distortions [5].

2.4 Performance Comparison of MSE, SSIM and CWSSIM

I) j) k) l)

Fig 2.2 Einstein image altered with different types of distortions [4].

In above Figure 2.2 we have:

a)Reference Image	e)Impulsive noise contamination	i)Shift(right)
b)Mean Contrast Stretch	f) JPEG Compression	j)Shift(Left)
c)Luminance Shift	g)Blurring	k)Rotate(anti)
d)Gaussian Noise Contamination	h)Spatial Scaling(zooming out)	l)Rotate Clock

2.4.1 Results Obtained

I. As shown in the above figure2, where the original "Einstein" image is altered with different distortions, each adjusted to yield nearly identical MSE relative to the original image. Despite this, the images can be seen to have drastically different perceptual quality. This leads to failure of MSE.

II. It can be seen that SSIM scores are much more consistent than MSE scores relative to visual perception.

III. Luminance shifting and contrast-stretching, which generally does not degrade image structure, lead to very high SSIM values, while noise contamination and excessive JPEG compression lead to low SSIM values.

IV. But there is problem in SSIM when it is tuned to scaling and rotation as in fig 2(h)-(l) .in which image is shifted right,left and rotate clockwise and anticlockwise and led to give low SSIM scores.

V. It may be seen that CWSSIM delivers high score to luminance shifted,contrast stretched,spatial shifting,scaled and rotated images and low scores to images containing structural distortions.

Chapter 3

Problem Statement

In this chapter, gaps in the existing work, problem statement, objectives to achieve and methodology to achieve the objectives are discussed.

3.1 Gap Analysis

In the previous chapter of Literature Survey, various full reference metrics are discussed to assess the quality of an image. Following are the gaps in the existing work:

I. Existing techniques like MSE and PSNR are not coorelated to human vision.

II. Some of existing metrics based on HVS like SSIM do not respond well to translation and rotation of images.

III. Existing metrics do not determine high level features of an image like VIF, SSIM, VDM.

IV. Most of the full reference quality metrics do not perform well for finding the luminance intensity of images well.

V. Existing Metrics like CED do not proide good results with respect to time and cost factor.

VI. Existing metric like CWSSIM , PDiff do not perform well for luminance of an images.

3.2 Problem Statement

There are many full reference metrics implemented for assessing the quality of an image based on high level and low level features of an image based on HVS. These are implemented by different metrics. In this thesis effort is made to implement a new full reference metric by integrating the human vision based metrics including SSIM, VIF, VDM and Pdiff.

3.3 Objectives

I. To study the existing metrics for image quality assessment based on full reference approach

II To propose a novel full reference image quality metric.

III To implement the proposed technique

IV To test and validate the use of metric on various images available in LIVE Image Quality Assessment Database.

3.4 Methodology

I. Literature Survey

II. Propose the technique using concepts of human vision based image quality metrics.

III. Implementation process using the concepts of SSIM , VDM, PDiff and VIF.

IV. Test the metric using LIVE database available for various reference images and observe the results.

Chapter 4

Implementation Details

In this chapter the procedure described to implement a novel full reference metric for assessing the quality of an image. This implementation is done in MATLAB (Appendix 1)

4.1 Procedure adopted for evaluating the quality of an image

1. Segmented the original image.

2. Finding the contrast of region in the image.

3. Compute the size of region in the image.

4. Compute the shape of region in the image

5. Finnd the location of region in the image.

6. Find the foreground/background regions in the image

7. Determine the quality index of the image

4.2 Step by Step Implementation of New Metric on Boat Image

4.2.1 Choose any original image from LIVE Image Quality Assessment Database

Fig 4.1 Original image of boat

4.2.2 Segmented image

Fig 4.2 Segmented Boat Image

4.2.3 Contrast of region

The contrast importance Icontrast of a region Ri is calculated as:

Icontrast(Ri) = gl(Ri) – gl(Ri-neighbours)... …............... (4.1)

where gl(Ri) is the mean grey level of region Ri, and gl(Ri- neighbours) is the mean grey-level of all of the neighbouring regions of Ri. Subtraction is used rather than division, since it is assumed that the grey-levels approximate a perceptually linear space. The resulting Icontrast is then scaled so that the maximum contrast importance of a region in an image is 1:0, and the range of Icontrast is [0.0-1.0].

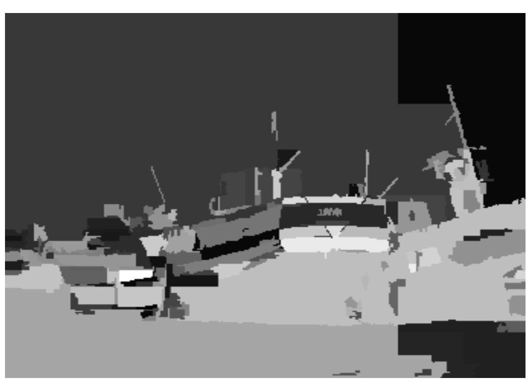


Fig 4.3 Contrast of boat image

4.2.4 Size of region

Size of region is calculated as

Isize(Ri) = min(A(Ri)/Amax, 1.0)..(4.2)

where A(Ri) is the area of region Ri in pixels, and Amax is a constant used to prevent excessive weighting being given to very large regions. Amax is equal to 1% of the total image area in the current implementation.

Fig 4.4 Size of region in boat

4.2.5 Shape of region

Region shape is calculated as:

Ishape(Ri) = bp(Ri)sp/A(Ri)..(4.3)

where bp(Ri) is the number of pixels in the region Ri which border with other regions, and sp is a constant. A value of sp = 1:75 is used to minimise the dependency of this calculation on region size. Long and thin regions and regions with a large number of edges, will have a high shape importance while for rounder or regular-shaped regions it will be lower. The range of Ishape is [0.0-1.0].

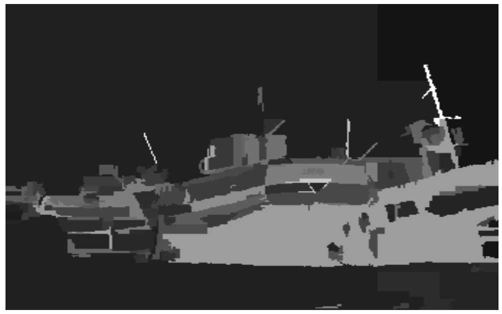


Fig 4.5 Shape of region in boat

4.2.6 Location of region

.Importance due to the location of a region is calculated as

Ilocation(Ri) = centre(Ri) A(Ri)...(4.4)

where centre(Ri) is the number of pixels in region Ri which are also in the central 25% of the image (calculated in terms of area). Thus regions contained entirely within the central quarter of an image will have a location importance of 1.0, and regions with no central pixels will have a location importance of 0.0.

Fig 4.6 Location of region in boat

4.2.7 Foreground / Background region

It is difficult to detect foreground and background objects in a still scene, since no motion information is present. However, a general assumption that can be made is that foreground objects will not be located on the border of the scene. Background regions can then be detected by determining the number of pixels which lie on the image border that are contained in each region.

Background/Foreground importance as given by

Ibackground(Ri) = 1.0 - min(borderpix(Ri)/0:5 total borderpix , 1.0)..................(4.5)

where borderpix(Ri) is the number of pixels in region Ri which also border on the image, and total borderpix is the total number of image border pixels. Any region which contains greater than half of the total number of image border pixels is likely to be part of the background, and is assigned Background/Foreground importance of 0.0.

Fig 4.7 foreground/ background of boat

4.2.8 Determine the quality index

Step 1: Compute Locations at which the Distortions are Visible.

Step 2: Combine the Visibility Map with Local Errors. After the map of visible locations is created, we use visibility-weighted. After the map of visible locations is created. Determine the quality index using Pdiff method.

Step 3: Compute SSIM of the image.

Step4: Quality Index of an image found as shown in fig 4.8.

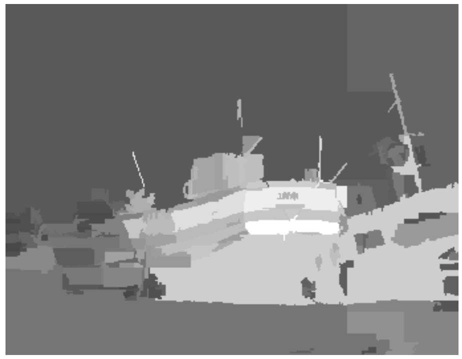


Fig 4.8 Quality Index of boat

The quality index of an image for new metric is 0.913 which is better than previous HVS metrics .

Chapter 5

Testing and Results

In this section, the performance of novel full reference metric is analyzed in terms of its ability to predict quality index of the images. To assess the quality index of the images, a novel full reference was applied to LIVE Image Quality Assessment Database.

5.1 LIVE Image Quality Assessment Database

The LIVE database contains 29 original images, 26 to 29 distorted versions of each original image, and subjective ratings of quality for each distorted image (differential mean opinion scores, DMOS values). The distortions used in LIVE are: Gaussian blurring, additive white noise, JPEG compression, JPEG-2000 compression, and simulated data packet loss of transmitted JPEG-2000-compressed images [20].

Fig 5.1 LIVE Image Quality Assessment Database

All images in LIVE Image Quality Assessment Database are of size 512x384 pixels.

All images of fixed size have been obtained by cropping selected fragments from the original images .

5.2 Various Types of Distortions

Table 5.1 presents the distortions modeled in our image database.

Table 5.1

S.No	Types of Distortion	Accounted HVS Properties
1	Additive Guassian Noise	Adaptivity, Robustness
2	Masked Noise	Local Contrast Sensitivity
3	High Frequency Noise	Spatial Frequency Sensitivity
4	Gaussian Blur	Spatial Frequency Sensitivity
5	JPEG compression	Color, Spatial Frequency Sensitivity
6	Image denoising	Spatial Frequency Sensitivity, Local Contrast Senstivity
7	Contrast Change	Light Level, Local Frequency Sensitivity
8	Quantization noise	Color, Local Contrast, Spatial Frequency
9	Intensity Shift	Light Level Sensitivity
10	Impulse Noise	Robustness

Figure 5.2 Non Eccentricity Distortions

a) b) c)

a) distorted image

b) enlarged fragments of the reference image

c)the same fragment with introduced distortions (the corresponding places are marked by white circles)

Figure 5.3 Contrast change

a) to larger contrast

b) to smaller contrast

5.3 Performance Comparison of Novel Full Reference Metric with other HVS based metrics for different types of distortions

Distortion Type	SSIM	VIF	VDM	PDM	Novel Metric
JPEG	0.67	0.69	0.69	0.72	**0.74**
JPEG 2000	0.76	0.77	0.76	0.78	**0.79**
Gaussian Blur	0.83	0.85	0.79	0.77	**0.88**
Guassian Noise	0.81	0.82	0.82	0.77	**0.87**

Table 5.2 Quality Index of different full reference image quality metrics

Above table 5.2 shows that Novel Full Reference metric provides better results as compared to other HVS based metrics for assessing the quality of the image for different types of distortions in the image.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

The rapid increase in the use of digital images has created a strong need for the development of objective image quality metrics. This thesis has presented a new metric for assessing the quality of images based on the concepts of HVS. The early human vision model improves on previous techniques by explicitly taking into account human response to complex natural images in the development and tuning of the model. The combination of these models has provided a signicant increase in the accuracy of the quality metric, as varied by subjective quality experiments.

6.2 Future Scope

There are so many metrics available for assessing the quality of image. But some of metrics are not able to correlate well with human vision and do not perform well for luminance of images. So, new techniques can be developed for assessing the quality of an image. Also existing approaches can be extended to assess the quality of the images.

REFERENCES

[1] Ing.Martin Slanina,Prof. Ing.VR,Csc, “A Comparison of Full Reference Image Quality Assessment Methods”, Institute of Radio Electronics, Czech Republic.

[2] Z. Wang and A.C. Bovik, “Modern Image Quality Assessment”, San Rafael, CA: Morgan & Claypool, 2006.

[3] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment:From error visibility to structural similarity,” IEEE Trans. Image Processing, vol. 13, No. 4, Apr. 2004.

[4] Z. Wang and A.C. Bovik, “ Mean Squared Error: Love It or Leave It”, IEEE Signal Processing, Jan 2009 .

[5] Z. Wang and A. C. Bovik, “A universal image quality index,” IEEE Signal Processing Letters, Vol. XX No. Y, March 2002.

[6] Z. Wang and A. C. Bovik, “Why is Image Quality Assessment so difficult?”, in Proc. IEEE Int.Conf. Acoust. Speech, and Signal Processing.

[7] Nora A. Naguib, Ahmed E. Hussein, etal. “Contrast Error distribution Measurement for Full Reference Image Quality Assessment”, Faculty of Engineering, Helwan Univrsity, Egypt.

[8] H.R Sheikh, A.C Bovik and G.de Veciana, “ An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistic”, IEEE Transactions on Image Processing, Vol.14, No.12, pp. 2217-2128, December 2005.

[9] YueWang, Tingting Jiang, etal. “ Image Quality Assessment Based on Local Orientation Distributions”, 28th Picture Coding Symposium, PCS 2010, December 8-10, 2010, Nagoya, Japan.

[10] Engeldrum, “A short image quality model taxonomy”, Journal of Imaging Science and Technology

48, 2 (March/April 2004), 160–165.

[11] Janssen, T. J. W. M., and Blommaert, F. J. J, "Image quality semantics" , Journal of Imaging Science and Technology" , 5 (September/October 1997), 555–560.

[12] Keelan, B. W " Handbook of Image Quality: Characterization and Prediction" Marcel Dekker Inc, 2002.

[13] Z.Wang, A.C Bovik and E.P. Simoncelli, " Handbook of Image and Video Processing", Chapter 8.3, 2nd Edition, Academic Press, 2005.

[14] D.M Chandler and S.S. Hemami, " A wavelet based visual signal to noise ratio for natural images, " IEEE Trans. Image Processing, Vol.14, pp.1918-1927, Nov. 2005.

[15] H.R. Sheikh, M.F. Sabir, etal. "A statistical evaluation of recent full reference image quality assessment algorithms", IEEE Trans. IP, vol. 15, pp. 3440-3451, 2006.

[16] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola and V. Lukin, "On between-coefficient contrast masking of DCT basis functions", CD-ROM Proceedings of the Third International Workshop on VPQM, 2007.

[17] William, K. Pratt, "Digital Image Processing", Fourth Edition, A John Wiley & Sons Inc. Publication, pp.465-529, 2007.

[18] De Ridder, H. Naturalness and image quality: Saturation and lightness variation in color images of natural scenes. Journal of Imaging Science and Technology, 6 (November/December 1996), 487–493.

[19] Silverstein, D. A., and Farrell, J. E. The relationship between image fidelity and image quality. In Proceedings of the International Conference on Image Processing (Lausanne, Switzerland, 1996), pp. 881–884.

[20] H.R. Sheikh Z.Wang, A.C Bovik and L.Cormack, "LIVE Image Quality Assessment

Database, Release. 2, 2005 http:// live.ee.utexas.edu/research/quality.

[21] D.-O. Kim and R.-H. Park, “Joint feature-based visual quality assessment,” Electron. Lett., vol. 43, no. 21, pp. 1134–1135, Oct. 2007.

[22] Dong-O Kim and Rae-Hong Park, “New Image Quality Metric Using the Harris Response”, IEEE Signal Processing Letters, Vol. 16, No. 7, July 2009 .

PUBLICATIONS

Sumit Mittal, Dr. Seema Bawa "Objective Full Reference Image Quality Assessment Metrics" International Journal of Advanced Research in Computer Science ISSN No. 0976-5697 pp. 214-215 Vol 2 , No.3, May-June 2011.

APPENDIX I

MATLAB

The name MATLAB stands for MATrix LABoratory. Dr. Cleve Moler, chief scientist at Math Works Inc. originally wrote MATLAB to provide easy access to matrix software developed in the LINPACK and EISPACK projects. The first version was written in the late 1970s for the use in courses in matrix theory, linear algebra, and numerical analysis. MATLAB is therefore built upon a foundation of sophisticated matrix software, in which the basic data element is a matrix that does not require pre-dimensioning. MATLAB is a product of the Math Works, Inc. and is an advanced software package specially designed for scientific and engineering computation. MATLAB is a high-performance language for technical computing [31]. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar non-interactive language such as C or FORTRAN. MATLAB has evolved over a period of years with input from many users. In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering, and science. In industry, MATLAB is the tool of choice for high-productivity research, development, and analysis.

MATLAB provides various types of toolboxes like Neural Network Toolbox, Image Processing Toolbox and Signal Processing Toolbox etc. Toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems.