OBJECTIVE ASSESSMENT OF QUALITY OF EXPERIENCE FOR DIVERSE IMAGE PROCESSING APPLICATIONS

BY

MUHAMMAD ALI QURESHI

A Dissertation Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In

ELECTRICAL ENGINEERING

NOVEMBER 2016

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MUHAMMAD ALI QURESHI** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING**.

Dissertation Committee

_____
Dr. Mohamed Deriche (Adviser)

_____
Dr. Mohamed Mohandes (Member)

_____
Dr. Abdelmalek Zidouri (Member)

_____
Dr. Azzedine Zerguine (Member)

_____
Dr. Khurram K. Qureshi (Member)

_____
Dr. Ali Ahmad, Al-Shaikhi
Department Chairman

_____
Dr. Salam A. Zummo
Dean of Graduate Studies

_____
Date   19| 12 | 16

The dissertation is dedicated to my beloved parents, wife, sisters, brothers, & friends for their endless love, support and encouragement in all stages of my life.

ACKNOWLEDGMENTS

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

First and foremost, I would like to thank Almighty Allah, for giving me the opportunity to pursue my Ph.D. studies at KFUPM and the courage to achieve research goals gracefully.

I gratefully acknowledge the Deanship of Graduate Studies, King Fahd University of Petroleum & Minerals (KFUPM), for awarding me a fully funded P.hD scholarship and providing me with excellent research environment and facilities.

I would like to deeply thank my adviser, Dr. Mohamed Deriche, for his continuous guidance and support in the research work throughout my Ph.D. studies. I learned a lot from his valuable knowledge and research experience. I will never forget his confidence towards me and patience in case of not meeting the deadlines. I would also like to say special thanks and acknowledgements to Prof. Azeddine Beghdadi, the Director of the L2TI Research Laboratory at University Paris 13, who invited me to visit his Lab and to do part of research work related to my Ph.D. studies under his guidance. I would also thankful to my committee members Dr. Mohamed Mohandes, Dr. Azzedine Zerguine, Dr. Abdelmalek Zidouri, Dr. Khurram Karim Qureshi for providing their valuable feedback.

I would also like to thank a number of people from the Electrical Engineering Department at KFUPM for their sincere support. These include Asjad Amin, Anum Ali, Hussain Ali, Dr. Naveed Iqbal, Waqas Afzal, Dr. Alam Zaib, Dr. Adeel Sabir, Dr. Rifaqat Hussain, among others. Among my colleagues and friends from Pakistan, I would like to say special thanks to Abdul Aziz, Ashiq Hussain, and Khan Muhammad for their continuous motivation and support in the time of need and difficulties.

Finally, I would like to thank some people with whom I have a much deeper relationship, not directly linked to this thesis. Among them are my parents for their unconditional love and always remembering me in their prayers. I would also like to thank my siblings for their sincere love and affection. Finally, gratitude goes to my wife for her understanding, patience, and tolerating my difficult schedule.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

$\oplus$ Exclusive OR

$*$ Convolution operation

α, β, γ parameters

τ Kendalls Rank Order Correlation Coefficient

ρ Spearmans Rank Order Correlation Coefficient

ζ Coefficient of Consistency or Transitivity

I_r An original image

I_d A distorted image

I_e An enhanced image

I_{inp} An inpainted image

$S(.)$ Saliency map of the original image

$S'(.)$ Saliency map of the inpainted image

Ω An inpainted region

Φ Remaining region $I - \Omega$

H Image height (rows)

W Image width (columns)

K_W Kendalls Coefficient of Concordance

p Pixel value under consideration

(x, y) Pixel index

b Block size

u Coefficient of Agreement

L Number of gray levels

LIST OF ABBREVIATIONS

ACR Absolute Category Rating.

AEBCE Adaptive Edge Based Contrast Enhancement.

AMBE Absolute Mean Brightness Error.

AME Absolute Measure of Enhancement.

AMEE Absolute Measure of Enhancement by Entropy.

ASVS Average Squared Visual Salience.

BorSal Border Saliency.

BPNN Back Propagation Neural Network.

CE Contrast Enhancement.

CEE Contrast Enhancement Evaluation.

CEED Contrast Enhancement Evaluation Database.

CF Colorfulness.

CLAHE Contrast Limited Adaptive Histogram Equalization.

CPBD Cumulative Probability of Blur Detection.

CS Compressed Sensing.

CSF Contrast Sensitivity Function.

DCR Double Category Rating.

DCT Discrete Cosine Transform.

DE Discrete Entropy.

DMOS Difference Mean Opinion Score.

DN Degree of Noticeability.

DR Dynamic Reference.

DRIQ Digitally Retouched Image Quality.

DS Double Stimulus.

DSCQS Double-Stimulus Continuous Quality Scale.

DSIS Double-Stimulus Impairment Scale.

EC Edge Content.

EME Measure of Enhancement.

EMEE Measure of Enhancement by Entropy.

EUM Extended Unsharpening Method.

FR Full Reference.

GCF Global Contrast Factor.

GD Gaze Density.

GHE Global Histogram Equalization.

GMSD Gradient Magnitude Similarity Deviation.

GRNN Generalized Regression Neural Network.

HOSVD Higher Order Singular Value Decomposition.

HVS Human Visual System.

IEM Image Enhancement Metric.

IIQA Image Inpainting Quality Assessment.

IQA Image Quality Assessment.

ITU International Telecommunication Union.

JNB Just Noticeable Blur.

JND Just Noticeable Difference.

KROCC Kendall Rank Order Correlation Coefficient.

LBP Local Binary Pattern.

LOE Lightness Order Error.

MICM Mutual Information based Contrast Measure.

MOS Mean Opinion Score.

MRETINEX Multi-scale Retinex.

MSE Mean Squared Error.

NR No Reference.

NSS Natural Scene Statistics.

PC Pairwise Comparison.

PDE Partial Differential Equation.

PLCC Pearson Linear Correlation Coefficient.

PSNR Peak Signal to Noise Ratio.

PWIIQ Parameter Weight Image Inpainting Quality.

QoE Quality of Experience.

RIQMC Reduced-reference Image Quality Metric for Contrast change.

RME Root Mean Enhancement.

RMSC Root Mean Square Contrast.

RMSE Root Mean Squared Error.

RR Reduced Reference.

RSE Radial Spectral Energy.

SDME Second Derivative like MEasurement.

SI Spatial Information.

SROCC Spearman Rank Order Correlation Coefficient.

SS Single Stimulus.

SSCQE Single-Stimulus Continuous Quality Evaluation.

SSIM Structural SIMilarity.

StructBorSal Structural Border Saliency.

SVD Singular Value Decomposition.

SVR Support Vector Regression.

TOPHAT Top Hat Transformation.

TUM-IID Technische Universitt Mnchen Image Inpainting Database.

VIF Visual Information Fidelity.

VisCoM Visual Coherence Metric.

VQEG Video Quality Expert Group.

THESIS ABSTRACT

NAME: Muhammad Ali Qureshi

TITLE OF STUDY: Objective Assessment of Quality of Experience for Diverse

Image Processing Applications

MAJOR FIELD: Electrical Engineering

DATE OF DEGREE: November 2016

Digital images and videos are becoming a vital source of information for Quality of Experience (QoE) and different multimedia related applications during recent years. Unfortunately, images and videos are likely to undergo various types of manipulations during different image processing operations. These manipulations affect the visual quality of images and videos. Image quality is evaluated either subjectively (done by humans) or objectively (using mathematical techniques). Objective measures are used for automatic monitoring of image/video quality, optimizing the control parameters in different image processing systems and algorithms (e.g., enhancement, restoration, inpainting, etc.). The formulation of objective Image Quality Assessment (IQA) problems is, however, very challenging. The task for No Reference (NR) IQA becomes even more challenging due to the un-

availability of the original image. Here, we present new frameworks for objective assessment of QoE for digital image degradations due to different types of manipulations including enhancement, distortions, and tampering.

In this work, we present a fast blind IQA metric for the images subjected to different degradations and a blind metric to quantify Blur distortion in color images with excellent results. Furthermore, we also develop a quantitative measure that can be used to detect some unpredictable side effects of Contrast Enhancement (CE) process. We also introduce new measures based on multi-metric fusions for Contrast Enhancement Evaluation (CEE). The study is the first of its kind, as performance of CE algorithms is extensively discussed in the literature but the performance of the CEE measures themselves is not well explained to date. Finally, we conduct a critical study of existing measures developed for IQA of inpainting applications, which can help researchers to benchmark new inpainting algorithms. In summary, the thesis provides a suite of new algorithms for evaluating different types of manipulations including distortions, enhancement, and image inpainting.

ملخص الرسالة

الاسم: محمد علي قريشي

عنوان الدراسة: تقييم موضوعي لجودة الصور في التطبيقات المختلفة

التخصص: هندسة كهربائية

تاريخ المناقشة: نوفمبر 2016

معالجة الصورة والفيديو اكتسبت أهمية كبيرة في تطبيقات مختلفة للوسائط المتعددة وخصوصا في السنوات الأخيرة. معالجة الصور والفيديو قد تتم بطرق مختلفة، وهذه المعالجات تؤثر على جوده الصور والفيديو. جودة الصور يتم تقييمها عموما عن طريق الانسان او عن طريق استخدام خوارزميات وتقنيات متقدمة في الحاسب.

لكن قياس جودة الصورة بطريقه موضوعية امر ليس بالسهل. حيث أن تقييم الجودة لها ارتباط كبير بالإدراك البشري ولكن معرفتنا لهذا الادراك وطبيعته معقدة. وعلى وجه الخصوص، تقييم جودة الصور يصبح صعب جدا عند عدم توفر الصور الأصلية.

هذه الأطروحة تحتوي على خوارزميات جديده للتقييم الموضوعي لجودة الصور الرقمية المعالجة. وتشمل كل من الصور المعالجة للتحسين والمعدلة والمشوهة وغيرها. هذا العمل، يشتمل على خوارزمية جديدة وسريعة ولا تحتاج الى مساعد جانبي ومعرفه مسبقة عن الصور الأصلية لتقييم جودة الصور. وخوارزمية أخرى لتقييم التموج الحاصل للصور الملونة للحصول على نتائج أفضل. وأيضا يشمل طريقه جديده لقياس بعض المؤثرات الجانبية الغير المتوقعة وذلك بصدد تحسين نوعية الصور. وأيضا طرحنا طريقه جديده مهجنة لتقييم كمية التحسين في الصور.

وهذه الأطروحة تتناول بشكل حصري المقارنة بين الطرق المستخدمة مسبقا في البحث العلمي والمهتمة بقياس التحسن النوعي للصور. حيث أن هذه الطرق استخدمت وبشكل كبير في القياس مع عدم وجود اي ورقة علمية منشورة للمقارنة بينهم بشكل شامل.

وفي النهاية، شملت الرسالة على دراسة تفصيلية للطرق المستخدمة لتقييم التحسن في الصور وبخاصة في الصور المحتوية على أجزاء مبهمة وناقصه. وهذه الدراسة تشكل نقطة مهمه للباحثين لتقييم الخوارزميات الجديدة. وفي الخلاصة نقترح في هذه الأطروحة مجموعة شاملة للخوارزميات لتقييم المعالجات المحتملة التي تمر بها الصور من تحسين وتشويه وتغير.

CHAPTER 1

INTRODUCTION

1.1   Background

With the rapid growth of Internet and advent of advanced image/video acquisition devices, digital images and videos are becoming an indispensable source of information for our daily living, our quality of experience, and the different social and economic aspects of society. Unfortunately, in practical applications, most images/videos data are subject to visual quality distortions during data acquisition, lossy compression, pre-processing (e.g. tone mapping, gamma correction, etc.), transmission errors due to a noisy communication channel, and diverse post-processing operations. Some of the distortions may even appear while securing the information in images/videos (e.g., watermarking, stegnography, etc.) [3].

During image acquisition from digital cameras, blurring artifacts may be introduced due to incorrect lens focus and noise may be added due to the shutter opening for a long time. Other blurring artifacts are caused by the movement of the camera and/or the object during image acquisition. Similarly, the space required to store raw images and videos is large, these image and videos undergo a compression process such as JPEG or JPEG2000, hence compression artifacts, i.e., blocking due to JPEG, blurring, and ringing due to JPEG2000, are intro-

duced. The different types of distortions affect the quality of images in a variety of ways. These distortions severely degrade image quality, which results in inaccurate perceptual judgment. As such, the provision of excellent quality of service to end users continues to be a major challenge to network service providers.

The question arises, how to assess consistently, the quality of a given image? The classical pixel-based measures like Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR), mostly used as quality indices, are unable to correlate well with the perceptual visual quality of images/videos. International Telecommunication Union (ITU) defines QoE as: "Overall acceptability of an application or service as perceived subjectively by the end users" [14]. Because of the above, the focus has now shifted from measuring accuracy in multimedia delivery to the provision of best-perceived multimedia quality to the end users. This prompted a major interest among researchers in developing robust IQA techniques (subjective and objective).

Subjective methods, as the name implies, involve humans in assessing image quality. Human observers provide the most realistic opinion on image quality, and the ratings are considered most reliable and accurate for perceived quality in a well-controlled environment and for a large number of observers. The specifications related to the methodology for subjective experiments are provided in ITU-R-BT.500-13 recommendations [15]. However, these methods are difficult to conduct, they are environment dependent, expensive, time-consuming, and not applicable for real-time applications [16]. For this reason, among others, we have witnessed during the last decade the growing interest, among academics, consumer electronics, IT industry, and in digital cinema etc., in developing automated or objective assessment techniques of image quality.

Objective IQA methods use mathematical techniques to extract and use characteristics features from the original and the distorted images and to use these

features in quantifying image quality. The objective methods, depending upon the availability of the original image, are grouped into Full Reference (FR), Reduced Reference (RR), and NR methods. In FR methods, the original image is required in addition to the processed image (e.g., distorted, enhanced, compressed, inpainted, etc.) and hence are not suitable for real-time applications, where the original image is usually not available. Whereas, there is no need for the original image in NR quality prediction methods. However, the original image information is partially required in the form of some extracted features in RR methods, which are seen as a compromise between FR and NR methods. Therefore, both RR and NR methods are the representative candidates for quality assessment of digital images for real-time applications.

The aim of objective quality assessment techniques is to predict perceived image quality, the way a human observer perceives it. While this field is still evolving, novel and better methods continue to emerge. It is also important to make the best use of these tools in real-world applications. In recent years, IQA has grown into a very active research sub-discipline under image processing.

The primary applications of image/video quality assessment techniques include:

- The monitoring and adjustment of image/video quality in real-time broadcast for delivering best quality image and video transmission.

- Benchmarking different sensing and acquisition technologies as well as in optimizing different image processing systems and algorithms for a particular task (e.g., image denoising, restoration, enhancement, inpainting, etc.).

- Parameter settings and optimization of different image processing systems and algorithms, etc..

The notion of visual information fidelity or image quality is highly related to

3

the way humans perceive distortions that may affect the quality of the observed image. Therefore, the IQA dilemma, in its traditional sense, has been long considered as a distortion estimation problem [17]. This, of course, is an important problem as it is desirable to have ready to use techniques to evaluate quality of the images subject to distortions or artifacts that may result from processing, lossy compression, or transmission. On the other hand, very few studies have been carried on the performance evaluation of image enhancement methods (better quality images rather than distorted images). Indeed, performing a quantitative evaluation of image quality improvement methods is a very challenging task. This is due to the absence of any objective measures able to account for some high-level vision tasks and their interaction with low-level image analysis when assessing perceptual quality in image enhancement [18]. This is also due to the difficulty in determining the most appropriate visual features to be used in the design of an overall image enhancement quality measure. Therefore, subjective evaluation is still the most reliable approach to assess the quality of enhanced images.

Enhancing image contrast is of major interest in many applications ranging from medical imaging [19], remote sensing [20], underwater imaging [21], defogging [22], etc. A plethora of CE methods has been proposed in the literature, and very few CE evaluations measures exist in this area. Moreover, there is no study to test the reliability of these measures themselves. Given the importance of CE in different applications, there is a need to investigate the performance of these measures in terms of robustness and consistency with human judgment. Moreover, there is no dedicated database for contrast-enhanced images so far. The existing contrast-processed databases contain images where the quality of processed images degraded due to the contrast manipulation. The conventional IQA metrics originally designed to quantify distortions are not well adapted for CE evaluation.

Similarly, image inpainting which is considered as a type of image tampering has also received considerable attention in different areas like the restoration of old and damaged documents, computational photography, etc. [3]. Image inpainting is a particular type of copy-move forgery used to restore missing or removing missing pixels or pixel regions in an image/video to make it as close as possible to the original or a given target image. Although a lot of research has been carried in introducing robust inpainting algorithms, limited efforts have been put in developing metrics for image inpainting quality evaluation. Among different types of distortions, most commonly observed distortions are the blur around edges and the contours in the restoration of large regions with missing pixels. The curved edges are also not restored. The quality evaluation of inpainted images is also important, with a limited research work has been carried in this regards. The ultimate choice is the subjective evaluation by human observers, which is time-consuming, complex, and challenging. In inpainting applications, usually the reference image is not available. Therefore, it is becoming even hard to develop NR inpainting quality evaluation metrics.

The objective IQA for different applications is the main focus of this thesis. In particular, we provide a comprehensive study of different IQA metrics for distortions, enhancement, and inpainting applications, complementing the limitations of existing work in the literature. The work covers IQA for distortions, enhancement, and a newly field inpainting. A number of contributions have been made under each of these categories.

1.2   Research Objectives

NR-IQA continues to be a very challenging problem due to the unavailability of original image information in various applications. The overall robustness of a

given NR-IQA algorithm depends primarily on the selected set of features and the way these are exploited for a particular IQA application. Currently, most of the metrics are based on quality of distortions. Here, we focus on the metrics mostly used to cover distortions, enhancement, and tampering (inpainting). More specifically, the main objectives of the thesis are:

- To propose a fast and simple NR-IQA metric to quantify different distortions.

- To propose an NR metric for blur prediction using color information for natural images.

- To develop a quantitative measure that could be used to detect and control some unpredictable side effects of image enhancement processes such as overshooting or halo effects.

- To develop a new database dedicated to CE images for performance analysis of CEE measures rather than CE methods.

- To provide a comprehensive statistical analysis of the data collected from subjective experiments on the above mentioned database.

- To perform a detailed analysis of the state-of-the-art CEE measures in terms of correlation with the subjective evaluation over the above-mentioned database as well as other publicly available CE databases.

- To propose a new metric for CEE and test it on the newly introduced CEE database.

- To perform a critical analysis of existing state-of-the-art image inpainting quality assessment measures.

1.3 Major Contributions

The main contributions of the thesis are:

- Development of a fast NR-IQA metric using texture moments and a machine learning approach [23].

- Development of an NR-IQA metric to quantify blur in natural images using tensors [24].

- Development of an NR metric based on mutual information to quantify global image contrast and to detect and control unpredictable side effects of CE [25, 26].

- Development and testing of a new database dedicated to contrast evaluation techniques. The quality rankings of enhanced images processed by different state-of-the-art CE algorithms were obtained from a psychophysical experimental setup [27].

- A detailed performance analysis of existing state-of-the-art CEE measures correlated to the human perception.

- Development of a new measure based on the multi-metric fusion of CEE measures strongly correlated with subjective human evaluation.

- A critical review of Image Inpainting Quality Assessment (IIQA) techniques and the introduction of a new metric for IIQA.

1.4 Thesis Organization

The rest of the thesis is structured as follows:

Chapter 2 provides details of objective quality assessment of digital images degraded with different artifacts, discussion on subjective experiments, summary of public image quality databases, evaluation measures and detailed descriptions of the proposed NR quality metrics.

Chapter 3 covers the quality evaluation of CE methods in general. It starts with the literature review of existing state-of-the-art CE evaluation measures, discussion on existing contrast manipulated databases. The proposed dedicated CE database and CE evaluation metric are discussed in detail.

Chapter 4 includes a critical review of the existing state-of-the-art IIQA metrics. It covers the literature review of image inpainting methods and quality assessment metrics. The strengths and shortcomings of existing IIQA metrics are highlighted and new guidelines for the development of new IIQA metric are provided.

Finally, we conclude the thesis in Chapter 5 and provide a discussion on future research perspectives.

CHAPTER 2

IMAGE QUALITY ASSESSMENT (IQA) FOR DEGRADATION EVALUATION

2.1 Introduction

Digital images and videos are becoming an essential part of our quality of experience and provide a source of information for various social and economic aspects of society. Over a simple one minute internet time, more than 3 million videos are viewed on Youtube, over 500,000 photos are posted online, and more than 20 million messages are exchanged on WhatsApp, many of which containing images. Unfortunately, with this substantial amount of internet traffic and the lack of control of content, the quality of images and videos posted suffers the most. Human vision is considered the best apparatus for perceiving and assessing image/video quality. Current research efforts aim at developing algorithms that try objectively to mimic the Human Visual System (HVS). Traditionally, image/video quality assessment methods have been grouped under two broad classes: Subjective and Objective methods (see Figure 2.1). Under each of the classes, a number
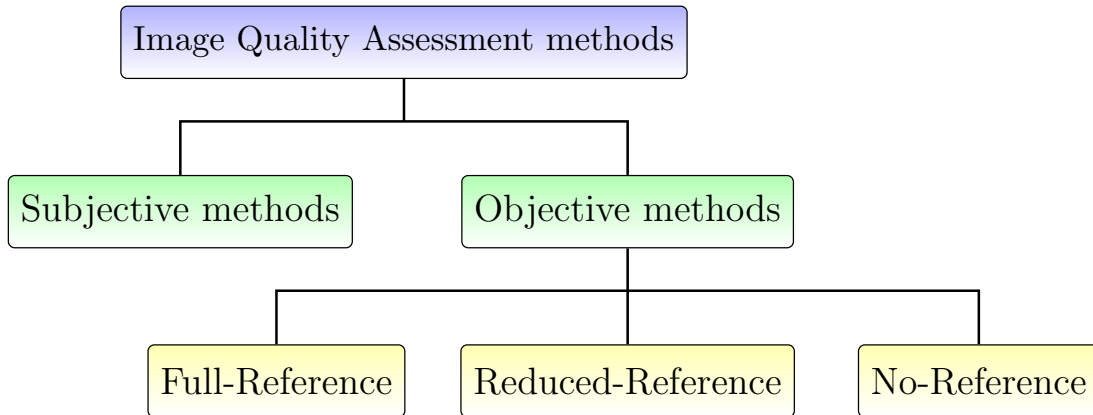
Figure 2.1: Classification of IQA methods

of approaches have been developed. These are discussed in more details further.

2.2 Subjective IQA Methods

Subjective methods involve human judgment of perceived quality, hence are considered as more reliable methods for real life applications. The most widely used recommendations are the ITU-R-BT.500.13 [28], which provide detailed explanations of materials, methods, and environment used in experimental testing. The ITU-R-BT.500.13 also provides discussion on statistical analysis of raw data collected from the subjective experiments. The subjective experiments are direct methods in which different subjects (observers) rate the quality of a given image. These methods need careful design considerations, well-controlled environment, and involve at least 15 observers to be meaningful [28]. However, these are time-consuming and cannot be used for real-time applications. They are generally used in the benchmarking of different objective image quality evaluation measures. Subjective methods can be further grouped based on the basis of rating and ranking [16]. In rating-based methods, participants assign a score to each stimulus presented to them either on an interval scale $(0 - 100)$ or categorical scale (Excellent, Good, Fair, Bad, and Very Bad). Based on the display of the

original image along with the test image, these methods are further classified into Single Stimulus (SS) and Double Stimulus (DS). So we have Absolute Category Rating (ACR) where only a single image is rated on a scale of discrete levels (e.g., 5−level scale), Double Category Rating (DCR) where both original and test images are displayed simultaneously and observers rate the test image on a discrete degradation scale based on the perceived quality as compared to the original image. The DCR methods are also called Double-Stimulus Impairment Scale (DSIS). Similarly, interval-based rating methods are also grouped into Single-Stimulus Continuous Quality Evaluation (SSCQE) and Double-Stimulus Continuous Quality Scale (DSCQS). In SSCQE, only a test image is shown to the observers and they are asked to rate the image quality using a slider over a continuous scale. Whereas in DSCQS, both original and test images are displayed in random order and the observers are unaware of the original image and they are asked to rate the quality of both images on a continuous scale. Between the category and interval rating based methods, the main problem in the interval (continuous) rating scales, is that people have their own perceptual judgment scales in their mind, hence it is very difficult to obtain an unbiased rating if the number of points on the scale is very large.

The ranking-based methods can be grouped into rank order-based methods and Pairwise Comparison (PC)-based methods. In rank order-based methods, different stimuli are displayed at once and the observers are asked to rank those according to their perceived quality judgement. This protocol is time-efficient; however, it is sometimes difficult to differentiate among the stimuli, particularly when the number of stimuli is more than three or four and the differences among the stimuli are also very small. Whereas, in PC-based methods, the stimuli are presented to the observers in pairs, and the observers choose whether stimuli A is better than stimuli B or vice versa, or both stimuli are alike. In this case, each

stimulus is compared with the other. The PC-based methods are simple because only one stimulus is compared with the other, and they are effective when the differences between the stimuli are very small compared to the rank-order or rating based methods.

The pairwise ranking raw data can be statistically analyzed in terms of coefficients of transitivity and consistency to sort out the bad participants as well as pathological stimuli [29]. Moreover, the pairwise ranking data can also be easily converted to rating scores. The PC ranking data can further be extended for more stimulus and images. However, since each stimulus is compared in pairwise manner with the others, the number of comparisons increases with the stimulus. For M stimuli (or methods in our case), the maximum number of pairwise comparisons becomes $\binom{M}{2} = \frac{M(M-1)}{2}$. An overview of different subjective methodologies used in IQA applications is shown in Figure 2.2. Note that in both rating and ranking based subjective experiments, the aggregated scores from all observers are considered as the overall ratings or ranking scores for each image.

Subjective methods are directly based on human visual perception, which makes them the most appropriate choice for image/video quality assessment; however, they exhibit a number of limitations. Human perception is dependent upon observers' mood, viewing distances, fatigue, and lighting conditions. These methods are also difficult to design, expensive, time consuming, and not recommended for real-time applications [30]. For this reason, significant research efforts have been made to develop objective IQA, metrics which correlate well with the rating obtained from human observers. In Section 2.5, we will discuss in details different objectives IQA metrics. Before that, we will give an overview of different subjective image quality databases and various performance evaluation measures used to validate the objective metrics.
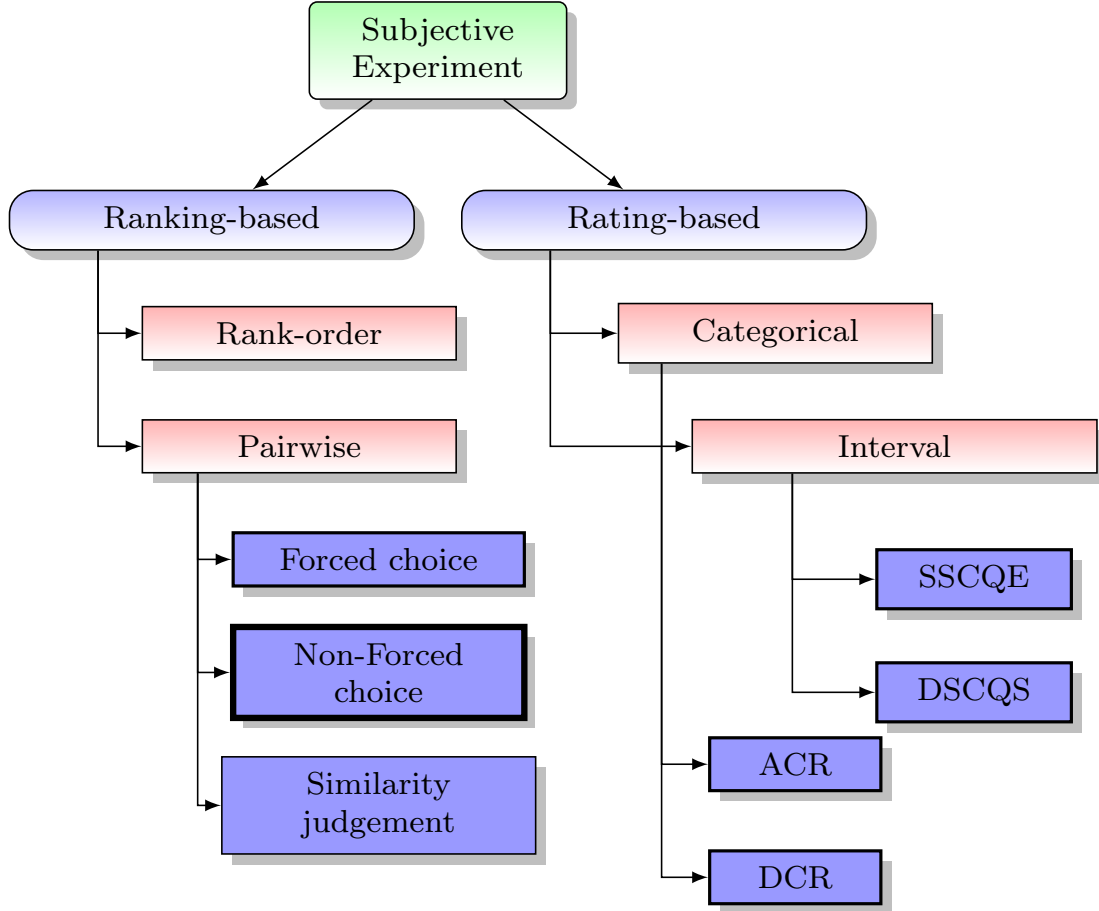
Figure 2.2: Classes of subjective methodologies used for quality assessment

2.3   IQA Databases

With the increased growth of research activities in IQA, it was important to introduce benchmarking IQA ground truth databases to be used for testing the different IQA metrics. They are also useful in the development of new metrics. Currently, many publicly available IQA databases along with the subject ratings are available. However, most of these are dedicated to IQA for degradation in image quality [31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41]. A brief description of these databases is given below.

13

Laboratory for Image and Video Engineering (LIVE) Database [31]: The database contains 29 original images, 779 processed images which are distorted with six distortions up to five levels. The distortions include compression artifacts (JPEG and JPEG2000), blurring, white Gaussian noise, and fast fading. The quality ratings are provided as Difference Mean Opinion Score (DMOS) scores in the range $(0 - 100)$.

Categorical Subjective Image Quality (CSIQ) Database [33]: The database was developed at Oklahoma State University, USA, and consists of 30 reference images and 866 distorted. Each original image is distorted using six different types of distortions with four to five distortion levels. The distortions are JPEG and JPEG2000 compression, Gaussian blurring, global contrast decrements, additive white and pink Gaussian noises. The ratings are given as DMOS scores on the scale $(0 - 1)$.

Tampere Image Database 2008 (TID2008) [32]: The database was developed at Tampere University of Technology, Finland. It contains 25 reference images, 1700 distorted images, and 17 distortions with four levels. The different distortions include noise distortions of various types (additive Gaussian, masked, spatially correlated, high frequency, impulse, and non-eccentricity pattern noise), Gaussian blur, image denoising, compression artifacts (JPEG, JPEG2000), transmission errors in compression (JPEG, JPEG2000), contrast artifacts, intensity shift, and variable intensity distortion at block level. The subjective experiments were carried out in three different countries both in the lab and using the internet. The quality ratings are given as Mean Opinion Score (MOS) scores in the range $(0 - 9)$.

Tampere Image Database 2013 (TID2013) [38]: The database is an extension of the TID2008 and is the largest image quality database consisting of totally 3000 distorted images, 25 reference images, 24 different types of distortions and five distortion levels. The quality ratings are given as MOS scores in the range $(0.2 - 7.3)$.

Image and Video Communication (IVC) Database [34]: The database consists of 10 reference images and 185 distorted images. The distorted images are created with four different types of degradations at five levels. The distortions include blurring, compression (JPEG and JPEG2000), and Local Adaptive Resolution (LAR) coding. The subjective ratings are given as MOS scores in the range $(1 - 5)$.

Media Information and Communication Technology (MICT) Database [37]: The database was developed by the University of Toyama, Japan. It consists of 14 original images and 168 distorted images. The database is limited to compression and communication artifacts and contains images distorted with JPEG and JPEG2000 compression. The ratings are provided as of MOS scores in the range $(1 - 5)$.

Colourlab Image Database: Image Quality (CID:IQ) [39]: This database is a recently developed one dedicated for IQA for distortions. The subjective experiments were performed in five different countries both in the laboratory environment and through the internet. The images in the database are distorted by compression artifacts (JPEG and JPEG2000), blurring, Poisson noise, and two gamut mapping methods.

Wireless Imaging Quality (WIQ) Database [35, 36]: This database consists of seven reference images and 80 distorted images, in grayscale and JPEG format. The artifacts are due to the simulated wireless channel. The ratings are provided as DMOS in the range $(0-100)$.

Cornell-A57 Database: The database consists of only three gray-scale original images and 54 distorted images. The six different types of artifacts are due to: additive white Gaussian noise, Gaussian blurring, quantization of 5-level wavelet high-frequency coefficients, additive white Gaussian noise, JPEG compression, and two JPEG2000 compression artifacts due to compression without visual frequency weighting and quantization based on dynamic contrast. The subjective ratings are provided as DMOS scores in the range $(0-1)$.

Camera Image Database (CID2013) [40]: This is a new database consisting of 480 real images captured by 79 cameras in six image groups. The database was evaluated by 188 observers.

A comprehensive summary of the above-mentioned databases including the number of original and distorted images, distortion types, levels of distortions, description of subjective experiments, and rating scores, is presented in Table 2.1.

2.4 Performance Evaluation Measures

With the diversity in the approaches used in measuring quality of images or image distortions come across the challenge that of normalizing the scores to a certain standard range. An important phenomenon to be considered in this mapping is the non-linearlity characteristics of human subjective scores. To account for these non-linearities in the subjective scores due to human opinions, the predicted objective scores need to be scaled using a nonlinear mapping function. A 5-

Table 2.1: Summary of subjective experiments used with different public IQA databases

Database	IVC[34]	LIVE[31]	A57[11]	MICT[37]	TID2008[32]	WIQ[36]	CSIQ[33]	TID2013[38]	CID2013[40]
Year	2005	2006	2007	2008	2008	2009	2010	2013	2013
Original Images	10	29	3	14	25	7	30	25	8
Processed images	185	779	54	168	1700	80	866	3000	480
Image resolution	512x512	*768 x 512	512 x 512	768 x 512	512 x 384	512 x 512	512 x 512	512 x 384	1600 x 1200
Degradation types	4	5	6	2	17	5	6	24	12 - 14
Distortion levels	5	5	-	6	4	-	4-5	5	-
Subjective method	DSIS	ACR	DMOS	SS	PC	DSCQS	Own	PC	ACR-DR
Score type	MOS	DMOS	DMOS	RAW	MOS	MOS	DMOS	MOS	-
Score scale	1-5	0-100	-	1-5	0-9	-	0-1	0-9	-
Screen resolution	CRT	CRT 21"	-	17"CRT	19"CRT + web	17"CRT	24" LCD	lab + web	24" LCD
Observers	15	20-29	7	16	838	60	35	971	188
Viewing distance	6HS	2-2.5 HS	-	4HI	varying	4HI	70cm	varying	80 cm
Image format	BMP	BMP	BMP	BMP	BMP	BMP	PNG	BMP	JPG
color/gray	color	color	gray	color	color	gray	color	color	color

parameter non-linear logistic fitting function is generally used as was discussed in details in [42]. The fitted objective scores, after nonlinear mapping, along with the subjective ratings (MOS/DMOS), are utilized for the estimation of few performance evaluation measures. The 5-parameter non-linear logistic function generally used is given as follows:

$$q = \beta_1 \left[\frac{1}{2} - \frac{1}{\exp^{\beta_2(x-\beta_3)}} \right] + \beta_4 x + \beta_5 \tag{2.1}$$

where q represents the fitted objective score after non-linear mapping, x is the calculated objective quality score, and β_k for $k = 1, 2, 3, 4, 5$ are the fitting parameters. These parameters are computed by minimizing the MSE between the subjective scores (MOS/DMOS) and the fitted values.

The Video Quality Expert Group (VQEG) [43] recommended some measures to validate the performance of objective image quality metrics. A brief description of commonly used performance evaluation measures is now given:

a. Pearson Linear Correlation Coefficient (PLCC) is used to measure the prediction accuracy (i.e. the ability to predict the subjective score with low error). It determines the strength of linear regression between the subjective scores (MOS/DMOS) and the objective scores after performing nonlinear regression analysis on the subjective scores. Its value ranges from -1 to $+1$ and it is calculated as:

$$\text{PLCC} = \frac{\sum_{i=1}^{N}(s_i - \bar{s})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^{N}(s_i - \bar{s})^2 \sum_{i=1}^{N}(q_i - \bar{q})^2}} \tag{2.2}$$

where s_i is the ith subjective score, q_i is the predicted objective score (after non-linear regression analysis) for the ith image/video, $\bar{s}$ and $\bar{q}$ are the averages of subjective and objective scores, respectively, for the whole database, and N is the

total number of images (or videos).

b. Spearman Rank Order Correlation Coefficient (SROCC) is used to measure the prediction monotonicity between metric scores and subjective scores (i.e., the degree to which the predicted scores agree with the relative magnitudes of subjective scores). The subjective scores, $\mathbf{s}$, and the objective scores, $\mathbf{x}$, are sorted and converted into their ranks and SROCC is calculated using Eq. (2.3). Since ranks are used in calculating the SROCC and the relative distance between the data points is ignored, the resulting score is independent of the non-linear mapping. Its value ranges from -1 to $+1$. The SROCC values close to $+1$ indicate that objective scores are in strong agreement with human perception, -1 means perfect disagreement and 0 means no correlation exists. The SROCC is computed as follows:

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2 - 1)} \tag{2.3}$$

where d_i is the difference between the ranks of ith image subjective and objective scores.

c. Kendall Rank Order Correlation Coefficient (KROCC) is a non-parametric rank correlation metric and is calculated as:

$$\text{KROCC} = \frac{2(N_c - N_d)}{N(N - 1)} \tag{2.4}$$

where N_c and N_d are the numbers of concordant and discordant pairs, respectively, in the list.

d. Root Mean Squared Error (RMSE) is used to measure the overall performance of the IQA metric and is calculated as .

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (s_i - q_i)^2} \qquad (2.5)$$

The objective scores after non-linear regression analysis are used in the calculation of the RMSE and PLCC.

In short, the evaluation metrics SROCC and KROCC are used to measure the prediction monotonicity whereas PLCC and RMSE assess the prediction accuracy of the objective quality assessment methods. Large values of SROCC, KROCC, and PLCC, while small values RMSE correspond to close relationship between the objective scores and the subjective ratings indicating the power of different objective IQA metrics.

2.5   Objective IQA Methods

With the above definitions and frameworks, we are moving to discuss the different approaches that have been proposed in the literature to measure image quality. As outlined earlier, objective methods, overcome the drawbacks of subjective methods by using mathematical techniques for extracting and using characteristic features from the reference and/or distorted images/videos. These features are then used to quantify quality. Objective quality assessment methods aim to predict perceived image/video quality score with high level of correlation with the subjective scores given by human subjects, which are the ultimate users in most image processing applications. Depending upon the availability of a reference image/video (i.e. an image/video with perfect quality), the objective methods can be grouped into FR, RR, and NR methods (see Figure 2.1). In FR methods, the reference image/video

is required in addition to the distorted image/video while there is no need for the reference image/video to predict the quality of a distorted image/video in NR methods. The reference image/video information is partially needed in the form of some extracted features in RR methods, which is seen as a compromise between FR and NR methods [44].

In recent years, more efforts have been put in developing IQA metrics that consider the different properties of the HVS. FR algorithms have attracted the most attention over the last decade. The reader can check the following references for FR-IQA algorithms [45, 46, 18]. Figure 2.4 shows the block diagram for a typical FR-IQA system model. FR-IQA methods are applicable in different off-line (stored) multimedia applications and IQA of this type is considered as almost a solved problem. We will see later that the biggest challenge resides in assessing quality when only the distorted image is available.

2.5.1 Full-Reference IQA Methods

In early research, the most widely used IQA methods compute visual quality by measuring pixel distortion, e.g., the MSE and the PSNR [47]. In pixel-based methods, the reference and the distorted images are compared on a pixel-by-pixel basis. The MSE is estimated as follows:

$$\text{MSE} = \frac{1}{\text{MN}} \sum_{i=1}^{M} \sum_{j=1}^{N} [\text{I}_\text{r}(i,j) - \text{I}_\text{d}(i,j)]^2 \tag{2.6}$$

where I_r, I_d are the reference and the distorted images respectively.

$$\text{PSNR (dB)} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) = 20 \log_{10} \left(\frac{255}{\sqrt{\text{MSE}}} \right) \tag{2.7}$$

The main advantages of PSNR and MSE are simplicity, ease in implementation, and clear physical meanings. However, such pixel-based methods were widely

criticized, for not correlating well with human visual perception. In Figure 2.3, we show a number of images affected by different types/levels of distortions but exhibiting the same value of the MSE.

Figure 2.3: Example of distorted images with similar MSE

FR-IQA methods have also been developed using image structure information. The most widely used FR metric is the Structural SIMilarity (SSIM). The SSIM assumes that natural images are highly structured, and the HVS perception is sensitive to structural distortions. Structural methods are based on comparing the structures of reference and distorted images and the structural degradation is considered as the quality score for the distorted image. In the SSIM [48], 3 types of similarities i.e. contrast similarity, structural similarity, and luminance similarity, are calculated in the spatial domain for overlapping blocks with one pixel overlap and using a sliding window approach. The product of these three similarities gives a local similarity map. The three components i.e., luminance

similarity, contrast similarity, and structural similarity are calculated using the following expressions:

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \qquad (2.8)$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \qquad (2.9)$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \qquad (2.10)$$

and finally, the SSIM is computed as:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha [c(\mathbf{x}, \mathbf{y})]^\beta [s(\mathbf{x}, \mathbf{y})]^\gamma \qquad (2.11)$$

where μ_x, μ_y, σ_x,σ_y, σ_{xy} are the mean, standard deviation and covariance between overlapping blocks of images $\mathbf{x}$ and $\mathbf{y}$. The constants C_1, C_2, and C_3 are used to avoid instability (for zero denominator). The overall quality score is obtained by averaging the local similarity scores over N blocks:

$$\text{SSIM} = \frac{1}{N} \sum_{i=1}^{N} \text{SSIM}(i) \qquad (2.12)$$

The SSIM popularity is mainly due to its simplicity and computational efficiency. Other variations of the SSIM include the MS-SSIM [49] using a multi-scale approach, the CW-SSIM [50] using complex wavelets, and the DW-SSIM [51] using discrete wavelets, etc. The major problem with the SSIM is that it does not perform well on blurred images. To solve this problem, a number of gradient-based approaches were introduced considering the fact that edges contain most of the image structure information.

Among the most popular gradient based structural FR-IQA methods is the Feature Similarity Index Metric (FSIM) [52] is the popular one. It uses the Phase Congruency (PC) and the Gradient Magnitude (GM) as low-level features for full

reference objective assessment of image quality. The motivation was that perceptual image quality depends on salient low-level features (image phase information) and these low level features change with distortions. Phase congruency is shown to be invariant to contrast changes, however HVS perception is highly dependent upon image contrast. To overcome this problem, image gradient (using the Scharr gradient operator) is used to capture the contrast as well as structure changes due to distortion. The phase congruency similarity map and gradient similarity map are computed as:

$$S_{PC}(i) = \frac{2.PC_r(i).PC_d(i) + C_1}{PC_r^2(i) + PC_d^2(i) + C1} \tag{2.13}$$

$$S_G(i) = \frac{2.G_r(i).G_d(i) + C_2}{G_r^2(i) + G_d^2(i) + C2} \tag{2.14}$$

where i represents pixel location and G_r, G_d are the gradients of the reference and the distorted image.

The local similarity map due to both phase congruency and gradient is obtained by multiplying both terms:

$$S(i) = S_{PC}(i).S_G(i) \tag{2.15}$$

The phase congruency value at each image location represents the importance of visual perceivable difference and used as weight in computing the overall quality score. The weights are calculated as:

$$PC_w(i) = \max(PC_r(i), PC_d(i)) \tag{2.16}$$

and finally the FSIM is given as:

$$FSIM = \frac{\sum_{i=1}^{N} S(i).PC_w(i)}{\sum_{i=1}^{N} PC_w(i)} \tag{2.17}$$

Another structure-based FR-IQA method is the Visual Saliency Index (VSI) [53] based on features calculated from visual saliency information of the reference and distorted images. Features based on image gradient map (using Scharr gradient operator) are also calculated, as visual saliency map is invariant to contrast. The local similarity map is obtained by multiplying both visual saliency map and image gradient map.

Recently, Xue et al. proposed the Gradient Magnitude Similarity Deviation (GMSD) [54] for FR-IQA using only the image gradient information (using the Prewitt gradient operator) to capture contrast and structural changes in image pixels occurring due to different distortions.

From the above mentioned methods, it is obvious that most of the FR-IQA assessment metrics are calculated in two steps. First, a local image similarity/dissimilarity quality map is obtained. The local quality map reflects the local quality of each image block in the distorted image. Then, an overall quality score for a given image is computed from these local maps in an all-important pooling stage. Among different pooling strategies, average pooling is the most widely used to obtain the overall score [48]. The main disadvantage of average pooling is that equal weights are assigned to all pixels. However, different distortions can affect different areas in an image based on the salient information and can give varying annoyance level (produced due to distortion). Moreover, edge pixels can give more visual information than pixels in smooth areas. Hence, HVS has different response in different areas/locations of an image while this phenomenon is totally ignored in average pooling. To overcome this problem, weighted averaging is used. In FSIM [52], the phase congruency value at each location in an image is used to weight the local quality score, while in VSI [53] local saliency map is used as a weighting factor for the local quality map to get overall score.

A new pooling strategy was discussed in the GMSD [54]. The standard de-

viation of local quality map is used as overall quality score for an image. The idea is that natural image contain different local structures. The degradations occur in these local structures due to different distortions, are not the same. In JPEG and JPEG2000, for example, the most prominent distortions are blocking, blurring and ringing. Blurring is prominently visible in textured areas rather than flat areas in an image. Blocking causes high quality degradations in smooth areas compared to textured areas. The global variation of image local quality (i.e. standard deviation of local quality map) is the best candidate for overall quality score for an image degraded due to multiple distortions. The different types of pooling strategies proposed for estimating the overall quality score from local quality scores are summarized in Table 2.2, where s represents local quality score at each pixel location, and N is total number of image pixels.

Before leaving the subject of FR-IQA techniques, it is worth noting that literature in this topic is very extensive. For this reason, we focused here on relevant and most common approaches. The reader is encouraged to refer to the following review/survey papers for more details [45, 46, 18].

In summary, FR-IQA methods are applicable in different off-line (stored) multimedia applications and IQA of this type is considered as almost a solved problem. We will see later that the biggest challenge resides in assessing quality when only the distorted image is available.

2.5.2   Reduced-Reference IQA Methods

In most practical applications, the reference image is not available. However, some of the reference image/video information may be available in the form of some extracted features [44]. Such scenario is called Reduced Reference IQA. Applications of RR-IQA techniques include real-time broadcast, tracking degradations in image quality to control the streaming resources, etc. Figure 2.5 shows the block

Table 2.2: Different pooling strategies used to compute the
overall quality score

Pooling strategy	Quality score (Q)	Examples	Remarks
averaging	$\frac{1}{N}\sum_{i=1}^{N} q_i$	SSIM [48]	Easy to compute. HVS perception at different areas on an image is not well considered.
weighting	$\dfrac{\sum_{i=1}^{N} q_i w_i}{\sum_{i=1}^{N} w_i}$	FSIM [52], VSI [53]	HVS perception is based on visual salient information, results are more correlated to subjective scores
standard deviation	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(q_i - \overline{q})^2}$	GMSD [54]	Effective for image degraded due to multiple distortions

$^{-}$ q_i represents local quality score at index i in local quality map.
$^{-}$ w_i represents weight for the local quality score at index i
$^{-}$ N represents total number of points in local quality map.

diagram for a typical RR-IQA system model.

In this figure, a reference image is sent to the receiver via a communication channel. The features are extracted from the reference image at the sender side and transmitted to the receiver through an ancillary channel. The reference image experiences distortion in the encoding stage as well as in the transmission channel. At the receiver, features are extracted from the distorted image similar to the
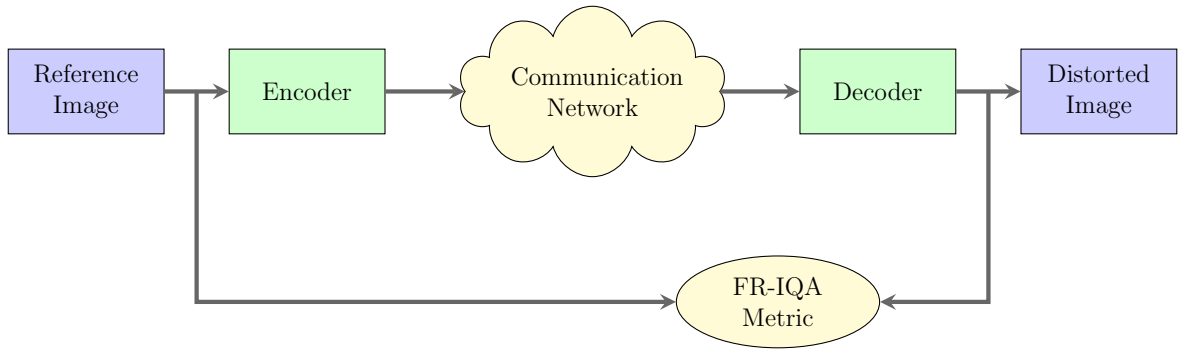
Figure 2.4: Block diagram of a typical FR-IQA system model
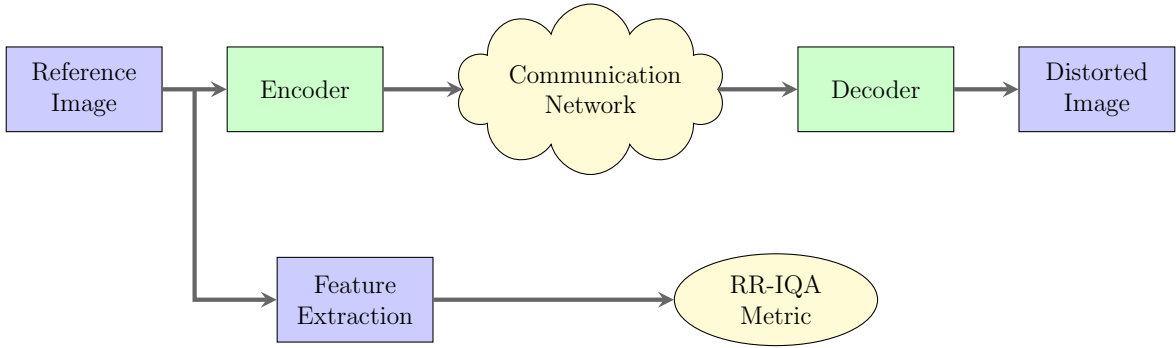
Figure 2.5: Block diagram of a typical RR-IQA system model
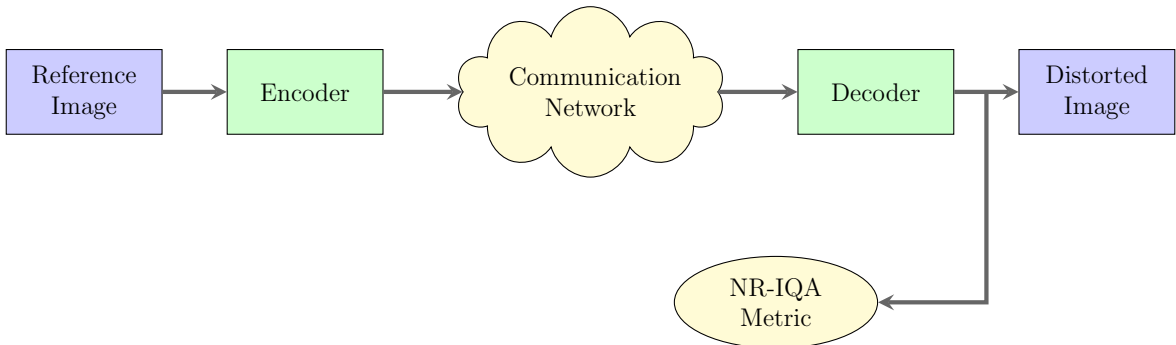
Figure 2.6: Block diagram of a typical NR-IQA system model

sender side. The features from both reference image and distorted image are used to estimate the overall quality score for the distorted image.

Wang et al. in [55] mentioned that for RR-IQA, the features should be perceptually relevant and sensitive to various distortions. They proposed a RR-IQA method based on the steerable pyramid wavelet transform and the natural scene statistics model. The image quality is estimated using Kullback-Leibler divergence between the marginal probability distributions of the wavelet coefficients from the reference and the distorted images.

In [56], Gao et al. proposed a framework for RR-IQA to mimic the HVS using multi-scale geometric analysis (MGA), Contrast Sensitivity Function (CSF), and the Webers law of Just Noticeable Difference (JND). In [57], Xue et al. proposed a RR-IQA algorithm based on modeling the subband coefficients using steerable pyramid transformations. The strongest coefficient edge-map corresponding to

the image gradient is built at each scale. The histogram of the SCM (Strongest Coefficient Map) is modeled with a Weibull distribution. The shape parameters of the Weibull distribution at different scales are used as RR features. The final quality score is calculated as a summation of geometric mean of absolute deviations and relative deviations of the reference and distorted image features.

In [58], Chetouani et al. proposed neural network based RR-IQA method. The statistical features are extracted in the wavelet domain from both reference and distorted images. The reference and distorted images are transformed using a 3-level wavelet decomposition. The edge-map for both the reference and the degraded image at each decomposition level is created as follows:

$$\text{Edge-Map}(i) = \sqrt{\text{LH}(i)^2 + \text{HL}(i)^2 + \text{HH}(i)^2} \qquad (2.18)$$

where LH, HL and HH are the horizontal, vertical and diagonal details subbands of an image at each decomposition level. For a 3-level wavelet decomposition, three edge-maps are created for each reference and distorted image. The mean and standard deviations are calculated from each edge-map for both the reference and the distorted image giving a total of 12 features (6 each for reference and distorted image). These features along with DMOS, are used for training a neural network. The trained model is then used as quality prediction of distorted images. The RR-IQA methods are the most appropriate choice for predicting quality closest to human subjective score compared to NR-IQA but they are limited for only those applications which require the reference image (in some form).

2.5.3 No-Reference IQA Methods

Even though both FR- and RR-IQA algorithms correlate well with subjective scores, NR-IQA is a more practical and challenging due to the unavailability of

reference image. Note that, often reference images are not available, e.g., television transmission; and hence NR-IQA is desirable, despite its computationally challenging nature. Figure 2.6 shows the block diagram for a typical NR-IQA system model. NR-IQA algorithms can be grouped into a) Distortion specific, b) Machine learning based, and c) Natural Scene Statistics (NSS) based. These categories are display in Figure 2.7.

A. Distortion-Specific based NR-IQA Techniques:

Distortion specific NR-IQA can only predict image quality with a certain type for distortion by extracting distortion aware features and hence have a limited scope [44].

The methods for blurriness detection, for example, are classified as spatial, transform, and hybrid methods. The spatial methods are further divided into edge based and non-edge based methods. Marziliano et al. [59] estimated blurriness effects based on average edge widths. Ong et al. [60] estimated blurriness effects based on edge widths in both the edge direction and its gradient direction. Among non-edge based spatial domain methods, Wee et al. in [61] estimated sharpness based on the largest eigenvalues of the covariance matrix of the image pixels. Zhu et al. in [62] estimated sharpness based on the Singular Value Decomposition (SVD) of the local image-gradient matrix. These metrics were also shown to perform well in the presence of noise.

Among the transform domain methods, Marichal et al. [63] estimated blurring effects based on the histogram of nonzero Discrete Cosine Transform (DCT) coefficients for 8×8 blocks. Similarly, Caviedes et al. [64] estimated sharpness based on the kurtosis of the DCT coefficients computed for 8 x 8 block centered at edge pixels. The overall blur estimate is computed as average of local kurtosis. The problem with edge-based blur estimation methods is that they fail in case
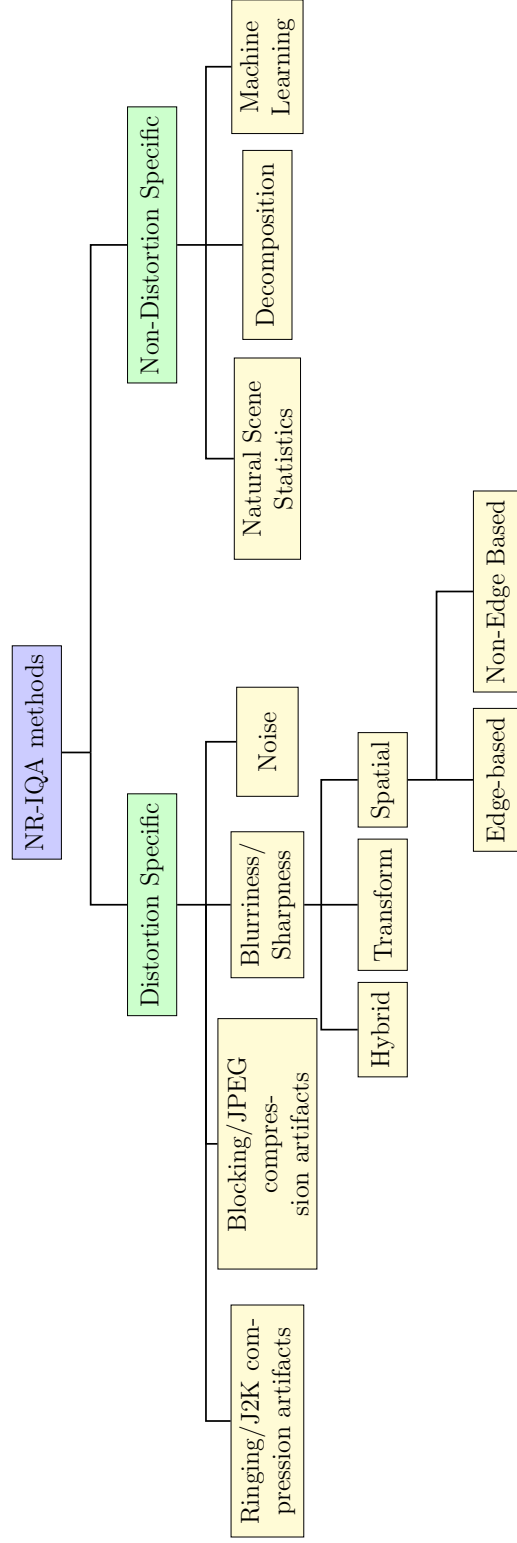
Figure 2.7: Classification of NR-IQA methods

of highly blurred images. To overcome this problem, Chetouani et al. in [65] proposed a metric for blur estimation without using the image edge information. The blur index is computed from the frequency domain radial analysis applied to both distorted image and its filtered version. While Vu and Chandler [66] estimated sharpness based on the weighted average of the log energies of the DWT subbands.

Among the hybrid methods, Vu and Chandler [66] estimated sharpness effects based on a combination of spectral and spatial measures. The spectral measure uses the slope of the local magnitude spectrum, and the spatial measure uses the local total variation of pixel values; these two measures were then combined using a weighted geometric mean to generate an image sharpness map, which averaged into a scalar indicating overall perceived sharpness.

In NR-IQA for JPEG compression artifacts, the general approach involves measuring edge strength at block boundaries, and then using this measure to estimate the visibility of the blocking, often based on masking. The image quality is then determined based on this estimate of perceived blockiness. These algorithms consider blocking as the most significant artifact originating from the compression process, so they first extract features to characterize the relative magnitudes of blocking artifacts.

In [67], Wang et al. proposed an efficient metric for blockiness distortion due to JPEG compression. They used the average absolute differences across the block boundaries and within block boundaries to estimate the blocking in JPEG images.

For JPEG2000 compression artifacts, the general approach involves measuring the amount of blurring or edge-spread by using edge-detection techniques. Other methods have also been developed based on natural-scene statistics. In [68], Sheikh et al. proposed to use the wavelet subband probabilities. The features extracted from these probabilities are used to estimate quality score using a

nonlinear fitting function. Zhang et al. [69] used kurtosis of the DCT coefficients as features for no-reference quality assessment of JPEG 2000 compressed images. The authors demonstrated the effectiveness of the metric in terms of parameter-free operations and computational efficiency. Other approaches can be found in [45, 46, 18].

B. Non-Distortion specific NR-IQA methods:

To handle all types of distortions, these algorithms are divided into machine learning based and natural scene statistics based.

B1. Machine Learning based NR-IQA Techniques: The algorithms belonging to this class, use machine learning based approaches for no-reference image quality assessment. Machine learning methods are powerful mathematical tools for solving prediction problems and provide good approximations of functional relationship between known sets of input and output data. This helps in predicting image quality scores that are close to that of the HVS. In [70], Tong et al. used a machine learning approach for NR-IQA based on neural network. The network was trained for both high-quality and low-quality image classes and binary classification was used to predict the quality of the distorted image by estimating the probability, the distorted image belongs to these two classes.

In [71], Tang et al. used low-level texture and natural scene statistics features based on complex wavelet transform for NR image quality assessment (LBIQ). These features are used for training three different regression networks and a final quality score is calculated as a weighted combination of quality estimation from the three regression models.

In [72], Li et al. presented a NR image quality assessment algorithm using a Generalized Regression Neural Network (GRNN). The mean and entropy of the

phase congruency map, and the entropy and gradient of the distorted image are used as features. The final quality score is calculated by approximating the functional relationship between these features and subjective scores using the GRNN. In [73], Ye et al. presented the CBIQ-I and CBIQ-II algorithms based on visual codebooks. The Gabor features are extracted from local image blocks to form the codebooks. The quantized features are then used to estimate image quality via either an example-based regression or support-vector regression. Further improvement in CBIQ-II was discussed by Ye et al. in [74], by using features from unsupervised learning instead of Gabor features which are shown to be effective across different distortions.

B2. Natural Scene Statistics based NR-IQA Techniques: NSS-based approaches assume that natural scenes possess certain statistical properties and that the presence of distortion will affect these properties. In this category, the perceptual relevant features are used to estimate the quality of the distorted image. In these methods, a two stage classification/regression network is generally used for quality prediction. In the classification stage, the distortions are characterized, and in the regression stage, the features along with the DMOS scores of training images are used to train the regression network.

In [75], Moorthy et al. presented the BIQI algorithm to estimate image quality using statistical features from a 9/7 Discrete Wavelet Transformation. The wavelet subband coefficients are modeled by Generalized Gaussian distribution, and mean and variance of the distribution are used as features. A feature vector of 18×1, is created using (3 scales $\times$ 3 orientations $\times$ 2 parameters). The same features are then used in the classification stage to characterize the distortion, and then in regression stage to estimate image quality.

To improve the performance of BIQI [75], Moorthy and Bovik in [76], presented

the DIIVINE algorithm for NR-IQA using statistical features extracted from different subband coefficients using a steerable pyramid transformation across 2 scales and 6 orientations. A feature vector of dimensions 88×1 was used for quality prediction using classification followed by a regression stage.

In [77, 78], Saad et al. proposed a no-reference IQA technique called BLind Image Integrity Notator using DCT Statistics (BLIINDS). Its two variants BLIINDS-I and BLIINDS-II use DCT statistics. In both BLIINDS-I and BLIINDS-II, the DCT is calculated for image blocks of 17×17. The DCT contrast and DCT-based structural features are extracted for each DCT image block. The DCT contrast is the average of the ratio of the non-DC DCT coefficient magnitudes in the image block normalized by the DC coefficient of that block. The DCT-based structure features are based on the kurtosis and anisotropy of each DCT block. The performance of BLIINDS-I is improved in BLIINDS-II using a generalized statistical model of local DCT coefficients and the model parameters are used as features to estimate image quality. The problem with these metrics is that they do not perform well for JPEG and FF (fast fading) distortions in the LIVE2 dataset.

In [79], Mittal et al. presented the BRISQUE algorithm, for real-time NR image quality assessment using spatial domain image statistics. 18 features are extracted for each of 2 image scales and total 36 features are used for distortion classification followed by regression to predict image quality.

In [80], He et al. proposed a blind IQA metric based on sparse representation of NSS features calculated from 4-level details subband coefficients. The NSS features are mean, variance and entropy of the wavelet subbands at different scales.

Instead of using transform domain features, Xue et al. in [81], proposed a NR image quality assessment method based on joint statistics of contrast features i.e. Gradient Magnitude (GM) and Laplacian Of Gaussian (LOG). The results are

comparable to other state of art NR-IQA methods.

NSS based NR-IQA algorithms work only with natural images; as under such distortions, images appear to be unnatural. However, modeling of natural images is a difficult task and does not apply to a good percent of images commonly used in practice.

Another approach for image quality assessment using the NR-IQA metrics was discussed by Chetouani et al. in [82]. The distortions were classified using Linear Discriminant Analysis (LDA) and FR-IQA metrics were used as features. Finally the quality is estimated using the most appropriate IQA metric.

A number of surveys have been carried by different researchers, some insight into the advantages and disadvantages of different techniques can be found in [45, 46, 18]. We have made two contributions towards objective quality assessment of image degradations. These are discussed in more details in the following sections.

2.6   A   Fast   No-Reference   IQA   Metric   using Law's Texture Moments

In this work, a computationally efficient NR-IQA algorithm is proposed that uses basic filtering operations in spatial domain. The features are calculated using Laws' filters proven to be efficient in texture analysis followed by range filtering. The overall quality score of an image is predicted using a simple GRNN (GRNN was shown in earlier work to provide better results than the traditional neural network). Laws filtered images are created by separable masks, which are easy to implement. Range filtering is an example of local filtering requiring few computations. The GRNN provides fast learning and smooth prediction. The proposed algorithm has low computational complexity, making it suitable for real-time applications. The performance of the proposed technique is confirmed, using the

LIVE 2 IQA database [31]. The proposed approach is shown to provide excellent results that are robust across different distortions, and is computationally less expensive than most existing techniques.

2.6.1   The Proposed Technique

Different types of distortions disturb the edges and contrast of images in different ways. To capture this fact, we propose to start with a feature extraction stage involving features that are perceptually-motivated. This stage is followed by a machine learning stage to predict an image quality score. The overall structure of the proposed technique is displayed in Fig. 2.8, and the details, of each individual block, are discussed below.
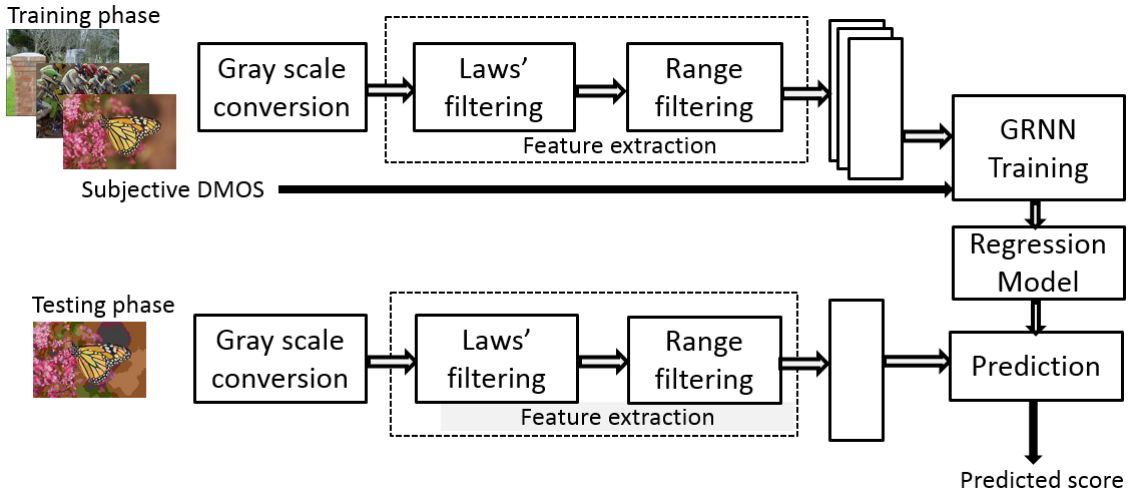
Figure 2.8: Experimental setup block diagram for the proposed method

(A) Feature Extraction The extraction of representative features plays an important role in texture analysis and pattern recognition problems. Texture plays an important role in human visual perception, as such the features need to be perceptually motivated for robust IQA. In [83], it was suggested that the preprocessing stage in HVS resembles band-pass filters. Furthermore, human visual perception is based on contrast rather than absolute intensity levels, so the

Table 2.3: Twenty-Five possible 2D Laws' masks

L5L5	L5E5	L5S5	L5W5	L5R5
E5L5	E5E5	E5S5	E5W5	E5R5
S5L5	S5E5	S5S5	S5W5	S5R5
W5L5	W5E5	W5S5	W5W5	W5R5
R5L5	R5E5	R5S5	R5W5	R5R5

$$
\begin{bmatrix}
1 & 4 & 6 & 4 & 1 \\
4 & 16 & 24 & 16 & 4 \\
6 & 24 & 36 & 24 & 6 \\
4 & 16 & 24 & 16 & 4 \\
1 & 4 & 6 & 4 & 1
\end{bmatrix}
\quad
\begin{bmatrix}
-1 & -4 & -6 & -4 & -1 \\
-2 & -8 & -12 & -8 & -2 \\
0 & 0 & 0 & 0 & 0 \\
+2 & +8 & +12 & +8 & +2 \\
+1 & +4 & +6 & +4 & +1
\end{bmatrix}
$$

(a)   (b)

Figure 2.9: Example of Laws' filter masks (a) L5L5 (b) E5L5

change in the minimum and maximum pixel values is also important. Under these observations, we opted to extract our perceptual features from the Laws' filtered images, followed by a range filtering stage.

(A1) Laws' Texture Moments: Laws, in [83], proposed 5 zero-summing (except $L5$) 1-D filter masks corresponding to the different local spatial textural features of images for the level, edge, spot, ripple and wave texture information. The five 1-D masks are $L5 = [1, 4, 6, 4, 1]$, $E5 = [-1, -2, 0, 2, 1]$, $S5 = [-1, 0, 2, 0, -1]$, $R5 = [1, -4, 6, -4, 1]$ and $W5 = [-1, 2, 0, -2, 1]$. From these 1-D masks, 25 2-D filter masks can be generated by convolving vertical 1-D filter masks with horizontal 1-D filter masks (see Table 2.3). Each filter mask is capable of extracting particular geometric information from the texture image (edges, lines, and spots). For example the $L5E5$ filter mask is used to capture vertical edges while $E5L5$ is used for extracting horizontal edge information.

(A2) Range Filtering: The original range filter was proposed, by Bailey et al. [84], to calculate the spatial intensity changes and to highlight edges in images. In range filtering, each pixel in an image is replaced by the difference of

(a) original image
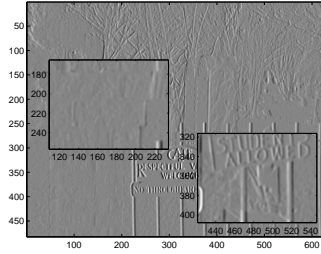
(b) V5L5
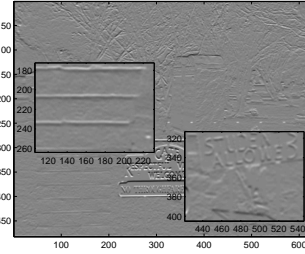
(c) E5L5

Figure 2.10: Examples of Laws filtering process

(a) V5L5
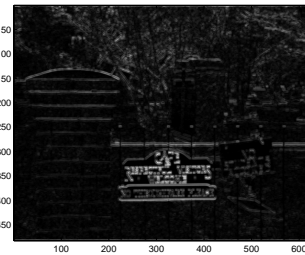
(b) E5L5

Figure 2.11: Example of range filtering on Laws filtered images

the maximum and minimum value of pixel intensities within a local neighborhood (window). Fig. 2.11 shows some examples of range filtering performed on some Law's filtered images. A significant enhancement of image edges is obtained.

The idea is to apply a sequence of filters on the image and for each filtered image, features are calculated. For the kth filter mask of size 5×5, the filtered image L_k is obtained by convolving the original image I with the filter mask h_k. Fig. 2.10 shows an example of applying some Laws' filter masks on an image from the dataset. Based on the above, for each image in the dataset, 25 filtered
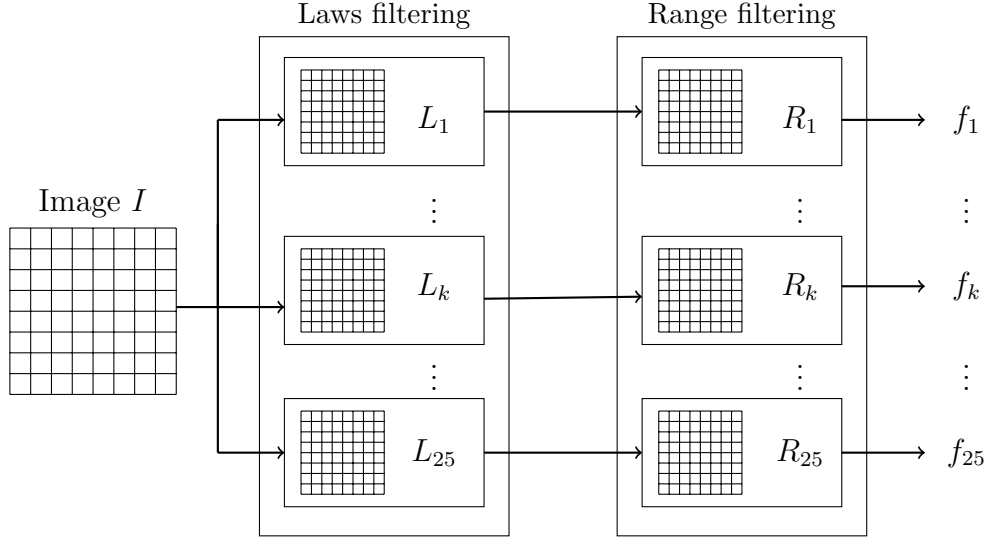
Figure 2.12: Feature extraction stage for the proposed NR-IQA metric

images are created by convolving it with 25 different Laws' masks. These masks are used to capture the occurrence of lines, points and edges within the image texture. Since masks are separable, their realtime implementation is simple and fast. The kth feature, f_k, is calculated by taking the average of the local range filtered images R_k. Fig. 2.12 shows the block diagram of the feature extraction stage.

(B) Generalized Regression Neural Network: In addition to the extraction of perceptually motivated features, an efficient feature fusion algorithm is also important in predicting a single quality score. Among different existing approaches, we have adopted, here, the GRNN, given its robust performance, its dynamic network structure, and simplicity. The GRNN was proposed by Spetcht [85] and is a type of probabilistic neural network that requires few training samples for effective learning, in comparison with the conventional Back Propagation Neural Network (BPNN). The probability density function used in GRNN is the Gaussian Distribution. Each training sample, $\mathbf{X}_i$, is used as the mean of a Normal

Distribution:

$$\hat{Y}(\mathbf{X}) = \frac{\sum_{i=1}^{n} Y_i \exp(-(\mathbf{X} - \mathbf{X}_i)^T(\mathbf{X} - \mathbf{X}_i)/2\sigma^2}{\sum_{i=1}^{n} \exp(-(\mathbf{X} - \mathbf{X}_i)^T(\mathbf{X} - \mathbf{X}_i)/2\sigma^2} \qquad (2.19)$$

where $\mathbf{X}$ represents a given input, $\mathbf{X}_i$ is a training vector, Y is the predicted value, n is the number of training samples (observations), and σ is a smoothing parameter.

A schematic diagram of the GRNN-IQA architecture is shown in Fig. 2.13. It consists of 4 layers, i.e input, hidden (pattern), summation, and output. The feature vector is applied to the input layer. The number of inputs is equal to the number of features. The number of neurons in the hidden (pattern) layer is equal to the input training vectors (samples). The input features are used with a Gaussian pdf (probability density function) in each pattern unit, then, the relationship between the input and the response of the pattern layer is stored in the unit. The summation layer has two units. Both units compute the weighted sum of the output of pattern layer units. For the first unit, the weight is set to Y_i and constitutes the numerator of Equation (2.19), while in the second unit, the weight is unity and represents the denominator of Equation (2.19). The output layer computes the quotient of the two outputs from the summation layer, and results in $\hat{Y}(\mathbf{X})$, the prediction, Y, of conditioned upon $\mathbf{X}$.

Compared to the conventional neural network model, the GRNN is non-iterative and can learn from training data in one-pass. Its advantages are: simple architecture, few training parameters, fast training and, excellent stability.

2.6.2 Performance Evaluation

The experiments are performed on a well known publicly available database [31]. The experimental procedure for the proposed technique is summarized in Fig.
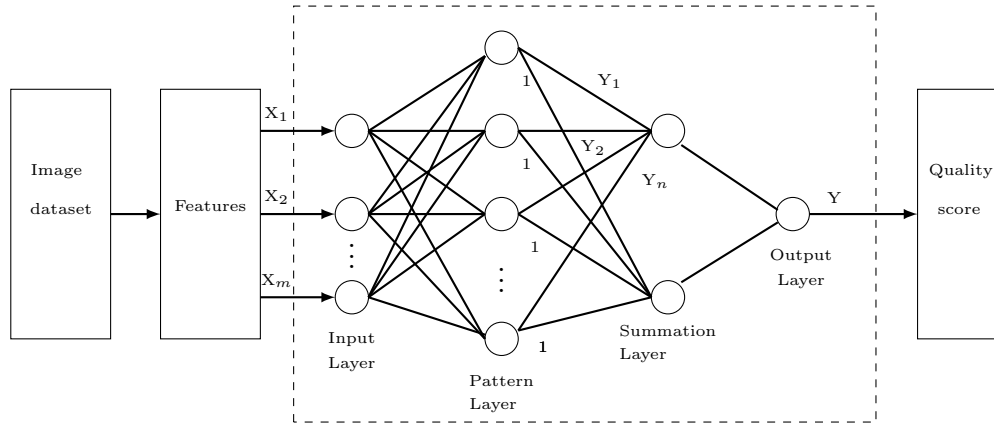
Figure 2.13: Schematic diagram of GRNN for image quality assessment

2.8. First, every image, I, in the database, is convolved with the 25 different Laws' filter masks to yield 25 filtered images. Then, for each filtered image, range filtering is applied to make the image details more prominent. Finally, the mean value of each range filtered image is calculated and a feature vector of dimension 25×1 is created for each distorted image. In order to have consistent results, the set of images belonging to 29 subjects in the LIVE 2 database are divided into five groups. (see Table 2.4). Then, 5-Fold cross validation is used to avoid bias in the regression process. The training is performed on four groups, and tested on the remaining group. The process is repeated such that each time a unique group is selected for testing, and the remaining four groups are used in training.

Table 2.4: Database Groups used for the Experiments

Dataset	Image Categories
G1	Sailing1, bikes, dancers, house, paintedhouse, statue
G2	Lighthouse2, rapids, womanhat, churchandcapitol, building2
G3	Monarch, parrots, sailing2, ocean, studentsculpt, carnivaldols
G4	Cemetry, manfishing, coinsinfo, sailing4, lighthouse, caps
G5	Plane, stream, buildings, woman, flowersonih35, sailing3

2.6.3 Experimental Results

The performance of the proposed NR-IQA method is tested for each distortion as well as a combination of all distortions in the LIVE 2 database. The performance evaluation metric selected is the SROCC which is used to measure the monotonicity between the subjective and objective scores. The results are summarized in Table 2.5. From the results, it is clear that the proposed technique is efficient for the images distorted with Gaussian blur, JPEG 2000, fast fading and white noise. The algorithm fails with JPEG, as our algorithm is not based on transform domain analysis. The regression plot of the predicted objective score and subjective DMOS for each distortion is also shown in Fig. 2.14. In order to benchmark the proposed method, the experimental results are compared with the traditional GRNN algorithm [72], which is also an NR-IQA using the same database. From the comparison, it is clear that the proposed method using Laws texture features, outperforms the GRNN [72] on all distortions except for JPEG distortion. The results were also compared with another NR-IQA algorithm; the BIQI [75], and again the proposed method outperforms on all distortions except for JPEG distortion. The main advantage of the algorithm is its simplicity, as only basic masking operations are required for feature extraction and can be efficiently implemented in hardware for real-time applications.

2.6.4 Discussions

To show the low complexity of the proposed algorithm, the overall computational efficiency of the proposed algorithm is compared with that of the NR BIQI[75], GRNN[72], BRISQUE[79], DIIVINE[] and BLIINDS-II[78] algorithms. The runtime of each algorithm to compute the quality score for a single image of resolution 768 x 512 in the LIVE database is calculated. The procedure is repeated 100 times and the average is shown in Table 2.6. The tests are performed on an Intel Core-i5

43

Table 2.5: SROCC for different algorithms and DMOS

Method		GBLUR	JPEG	J2K	FF	WN
Proposed	G1	0.942	0.737	0.707	0.595	0.979
	G2	0.915	0.884	0.739	0.872	0.964
	G3	0.972	0.898	0.889	0.899	0.971
	G4	0.887	0.809	0.854	0.806	0.987
	G5	0.911	0.957	0.930	0.545	0.971
	Avg.	**0.942**	0.857	**0.864**	**0.763**	**0.974**
GRNN [72]		0.833	0.872	0.816	0.735	0.979
BIQI[75]		0.846	0.891	0.799	0.707	0.951
ICA		0.9061	0.6465	0.8200	0.8164	0.9768
BLIINDS[78]		0.957	0.839	0.922	0.750	0.973

processor at 2.50 GHz, 4GB RAM, Windows 7 (64-bit). From Table 2.6, it is quite evident that the proposed algorithm outperforms all other algorithms. Thus the proposed algorithm is well suited for real-time blind IQA applications.

Table 2.6: Computational Complexity of different NR-IQA algorithms

Method	Time (seconds) per image
GRNN [72]	0.23
BIQI[75]	18.4
BRISQUE[79]	0.45
BLIINDS-II[78]	45.53
DIIVINE[76]	27.35
Proposed Algorithm	**0.11**

In this work, a fast NR image quality assessment method is proposed. Since human visual perception corresponds to bandpass filters, the Laws' filter masks are used to extract different bandpass filtered images. The results are highly correlated with human perceptions for the images distorted with white noise and Gaussian blur. Compared to the traditional approaches, the proposed method is recommended for NR-IQA due to its fast learning and low computational load. The proposed technique is simple to implement as the feature extraction stage is based only on basic convolution operations, and can be efficiently implemented in
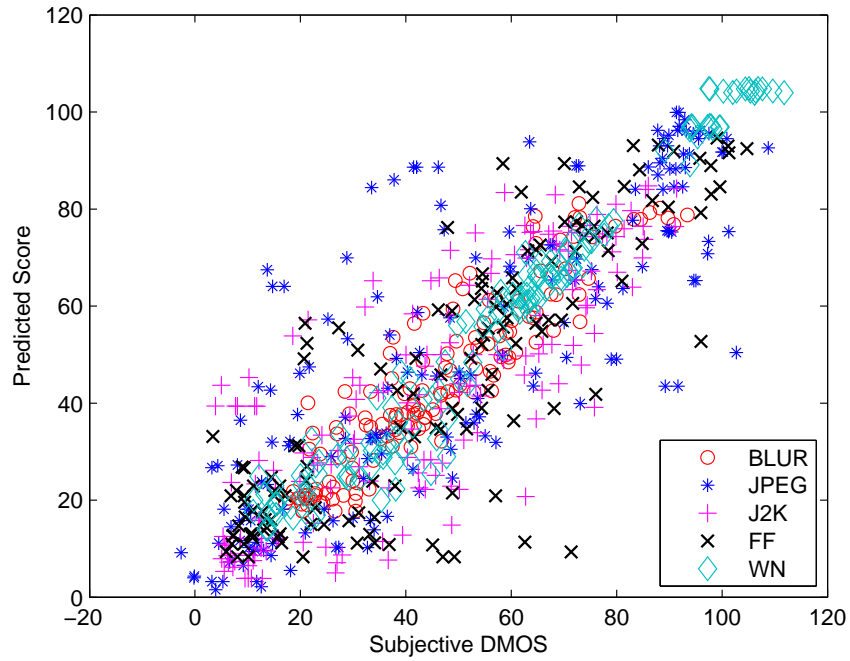
Figure 2.14: Regression plots between DMOS and predicted score for different distortions in LIVE 2 dataset.

hardware making it very suitable for real-time multimedia applications.

Another NR-IQA metric is also proposed to quantify Blur in color images. The reason for selecting blur distortions for our work is that it is considered as an important component in the spectrum of distortions. In the following section, we will discuss this metric in more details.

2.7 A No-Reference Blur Metric for Color Images using Higher Order Singular Values

During different processing stages, various artifacts are introduced in digital images. The blur is most commonly observed distortion among these artifacts, which is due to the limitations of acquisition equipment (i.e., out-of focus camera lens, low-lighting conditions, relative movement etc.) and different processing stages

before final viewing by the user. More importantly, blur also affects the edge information which is considered as a key factor to human perception of quality.

In FR scenario, the blur assessment is a simple task. However, in no-reference case, the task of distortion assessment becomes a more challenging one. In the literature, different methods have been developed for blind blur assessment of digital images. In [86], Ferzli et al. discussed an image sharpness/blur assessment technique based on the concept of Just Noticeable Blur (JNB). They determine the probability of blurriness required around the edge before it can be perceived or noticeable by the HVS. By using the same concept of JNB [86], Narvekar et al. [87] proposed a NR image sharpness metric based on the Cumulative Probability of Blur Detection (CPBD) at an edge. In [65], Chetouani et al. proposed an approach for no reference blur estimation based on Radial Spectral Energy (RSE) analysis. The NR blur metric was computed by adding the blur to an image and measuring its impact using the radial energy analysis in the frequency domain. There also exist some methods for quality assessment of digital images based on signal energy analysis using transform domain such as SVD. Among these, Shnayderman et al. [88] were the first to use SVD for FR image quality assessment. The reference and distorted images were divided into non-overlapping blocks and singular values for the corresponding blocks were calculated. The mean distance between the singular values of the reference and the distorted image blocks were used for quantifying quality. For NR-IQA using SVD, Sang et al. [89] demonstrated that the singular values of an image when plotted against their indices follow an exponentially decreasing curve with the degree of the exponent varying with the amount of blur. The same authors in [90] also proposed a blind blur assessment metric based on blur similarity by following the same idea in [65]. The blurred images were re-blurred with a Gaussian kernel of size 11×11. The similarity between the singular values of the distorted image and re-blurred image was used as a blur

index.

The methods discussed earlier as well as other state-of-the-art existing IQA techniques are mostly based on either the luminance component of a color image, or use separate color channels followed by pooling of the results to get the final quality score. For an RGB color image, there exists a strong correlation among the Red, Green, and Blue color components. Different distortions may influence different color components and disturb the correlation among them as well. The loss of color due to the different types of degradations substantially affects human perception. Moreover, the perception of blur is also different for Red, Green, and Blue color components. The reason is that the blur is directly related to the focus and in turn it depends on the wavelength of the incoming light (i.e., color). Indeed the focal length of the lens is related to the refractive index which varies with the wavelength of the color.

Therefore, it is inappropriate to completely ignore the correlation among the color components in IQA. The idea is to search for representations where the inter-channel correlation can be exploited in order to capture the effect of blur on the three channels. Wang et. al. [91] proposed the use of three color channels for FR quality assessment using SVD of Quaternion matrix. The Quaternion matrix was generated by taking the local variance of Red, Green, and Blue channels as imaginary part and the luminance component as the real part. The overall quality score was derived by computing the distance in singular values of image blocks in the reference and the distorted images.

Motivated by the superior performance of SVD for the luminance component of images in NR-IQA [90, 89], we introduce here a new framework for IQA of color images using the so-called higher order singular values. We propose to use tensor analysis to fully represent the correlation among different color components. Tensors are used to represent high dimensional data and to extract useful information

47

from high dimensions rather than from the 2D matrices [92, 1]. The Higher Order Singular Value Decomposition (HOSVD) is an efficient tensor decomposition technique [92]. It has been widely used in different image processing applications including color restoration [93] and denoising [94]. Cheng et al. [95] used tensors for the first time for FR quality assessment of color images. Here in this work, we extend for the first time, introduce a new framework for using tensors for blind blur assessment from color images using HOSVD. We consider a given RGB color image as a tensor and compute the higher order singular values from its unfoldings. We provide, some mathematical background of SVD, and tensors in the next section, followed by our proposed algorithm.

2.7.1 Mathematical Background

SVD decomposition of 2D Images A 2D gray scale image, $\mathbf{A} \in \mathbb{R}^{M \times N}$, satisfying some regularity conditions, can be decomposed using SVD as:

$$\mathbf{A} = \mathbf{U\Sigma V^{T}} \tag{2.20}$$

where $\mathbf{U} \in \mathbb{R}^{M \times M}$ is the matrix of left singular vectors, $\mathbf{V} \in \mathbb{R}^{N \times N}$ is the right singular matrix, and $\mathbf{\Sigma} \in \mathbb{R}^{M \times N}$ is the rectangular diagonal matrix of singular values arranged in descending order. The singular values vector can be extracted as $\mathbf{d} = \mathrm{diag}(\mathbf{\Sigma}) = [\sigma_1, \sigma_2, \cdots, \sigma_r]$ for $i = 1, 2, \cdots, r$, r being the rank of $\mathbf{A}$. The $\mathbf{U}$ and $\mathbf{V}$ give structural information along the rows and columns of $\mathbf{A}$ while the $\mathbf{d}$ represents the luminance or energy information of $\mathbf{A}$.

In [89], Sang et al. showed that the singular values computed from the luminance component of natural images when plotted against their indices, follow an exponential decay. The exponent coefficient of the decay varies with the amount or degree of blur present in the image. In [89], the authors provide the derivations

to quantify the quality scores from the singular values. For the sake of complete-ness, we briefly outline these steps again to show how the exponent coefficient is calculated which is then used as a quality score.

As mentioned previously, the SVD values of a given image decrease exponen-tially with increasing blur. The inverse power function or exponential can be expressed as:

$$\mathbf{y} = \mathbf{x}^{-\alpha} \tag{2.21}$$

where α is the exponent coefficient.

By taking the natural logarithm of both sides of (2.21), we get:

$$\ln\left(\frac{1}{\mathbf{y}}\right) = \alpha \ln(\mathbf{x}) \tag{2.22}$$

Let $\mathbf{w} = \ln(\mathbf{x})$ and $\mathbf{z} = \ln\left(\frac{1}{\mathbf{z}}\right)$, we get

$$\mathbf{z} = \alpha \mathbf{w} \tag{2.23}$$

which is a linear equation in terms of the coefficient α. We can solve for α using a Least Squares approach, i.e., minimizing mean square error:

$$\min \sum_{k=1}^{r} e_k^2 = \min_{\alpha} \sum_{k=1}^{r} (z_k - \alpha w_k)^2 \tag{2.24}$$

The optimal value of α, is obtained by taking the derivative of (2.24) w.r.t α, and setting it to zero:

$$\alpha = \frac{\sum\limits_{k=1}^{r} z_k w_k}{\sum\limits_{k=1}^{r} w_k w_k} = \frac{\sum\limits_{k=1}^{r} \ln(d_k)\ln(k)}{\sum\limits_{k=1}^{r} \ln(k)\ln(k)} \tag{2.25}$$

where d_k represents the singular value at index k for an image.

Preliminaries on Tensors and HOSVD While SVD is well suited to analyze 2D matrices, it cannot be used directly with higher dimension arrays. For such arrays, the concept of SVD has been extended using tensor theory. To formally define a tensor, let $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ be an Nth order tensor having N indices where $I_1, I_2, \cdots, I_N$ are the upper limits of each dimension. Therefore, we can say that a scalar is a zero-order tensor denoted by lowercase italic letters (e.g., a), a vector is a 1st-order tensor denoted by lowercase bold letters (e.g., $\mathbf{a}$), a matrix is a 2nd-order tensor denoted by capital bold letters (e.g., $\mathbf{A}$), and matrices with more than two dimensions are higher-order tensors represented by calligraphy letter ($\mathcal{A}$). Here, we propose to represent a given RGB color image as a 3rd-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, having 3 indices, where $I_1, I_2,$ and I_3 represent height, width, and number of color channels (i.e., 3 for Red, Green and Blue).

The tensors can be decomposed into fibers (modes) and slices by fixing all indices except for one and two respectively. Each index in a tensor is called mode, and upper limit of indices in each mode is called mode dimension. For 2D matrix, columns and rows are mode-1 and mode-2 fibers respectively. The 3rd-order tensor (an RGB image) has 3 modes (column, row and tube fibers) and dimension of each mode is I_1 (height), I_2 (width), and I_3 (3 for color channels) corresponding to $a_{:i_2 i_3}$, $a_{i_1 : i_3}$, and $a_{i_1 i_2 :}$ respectively. The mode-1 (columns) and mode-2 (rows) represent spatial information while mode-3 is used to represent color channels. The sub-tensors can be defined by fixing one or two indices of a 3rd-order tensor.

For different applications, tensors are often transformed into 2D matrices. The rearranging of elements of a tensor into a 2D matrix is known as unfolding or matricization. For a 3rd-order tensor (an RGB image) $\mathcal{A}$, a 2D matrix or slice is obtained by fixing one of the three indices. A frontal slice is obtained by fixing the 3rd index i_3, and denoted as $\mathbf{A}_{::i_3}$. Fixing the 2nd index, we get the lateral (or

(a) A 3rd-order tensor   (b) Mode-1 (column) fibers

(c) Mode-2 (row) fibers   (d) Mode-3 (tube) fibers
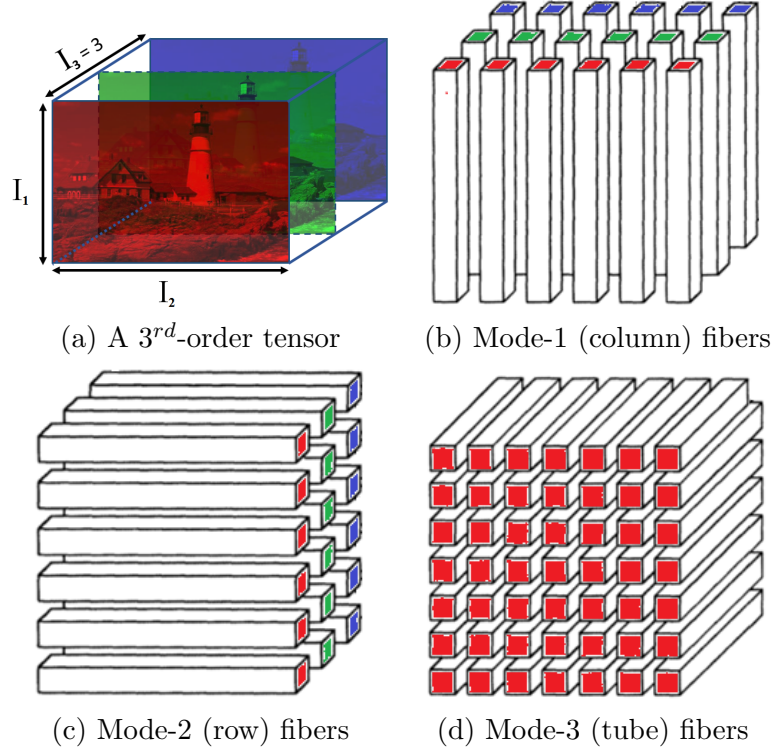
Figure 2.15: An example of 3rd-order tensor and Mode-n fibers [1]

vertical) slice $\mathbf{A}_{:i_2:}$, and fixing the 1st index, we get the horizontal slice $\mathbf{A}_{i_1::}$. In this way, the mode$_{(n)}$ unfolding of an RGB image is a matrix $A_{(n)} \in \mathbb{R}^{I_n \times \prod_{k,k \neq n} I_k}$. For illustration, mode$_{(n)}$ unfolding or slices of a 3rd-order tensor are shown in Fig. 2.16. The HOSVD is the SVD of each of the tensor modal unfoldings [92]. For a tensor $\mathcal{A}$, it is defined as:

$$\mathbf{U}_n^T \mathbf{A}_{(n)} = \mathbf{\Sigma}_n \mathbf{V}_n^T \quad \text{for} \quad 1 \leq n \leq d \tag{2.26}$$

where $\mathbf{U}_n^T$ and $\mathbf{U}_n^T$ are unitary matrices, the matrix $\mathbf{\Sigma}_n$ contains the singular values of $\mathbf{A}_{(n)}$ on the diagonal, and d represents the size of a tensor.

2.7.2   The Proposed Technique

We first decompose the color image into three unfoldings and compute the higher order singular values for each unfolding. Confirming previous results, we observed
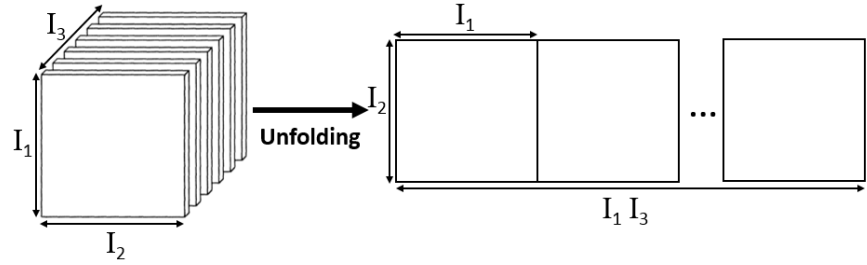
(a) Frontal slices $\mathbf{A}_{(1)} \in \mathbb{R}^{I_2 \times I_1 I_3}$
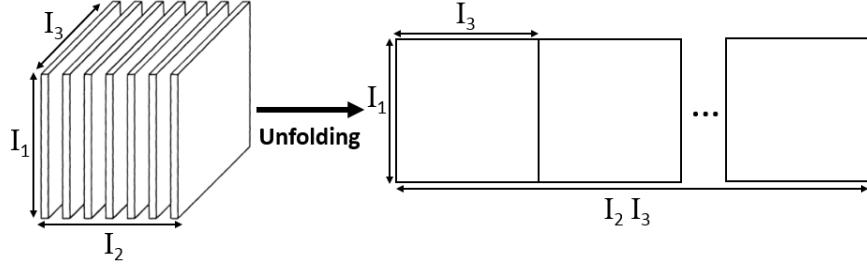
(b) Lateral slices $\mathbf{A}_{(2)} \in \mathbb{R}^{I_1 \times I_2 I_3}$
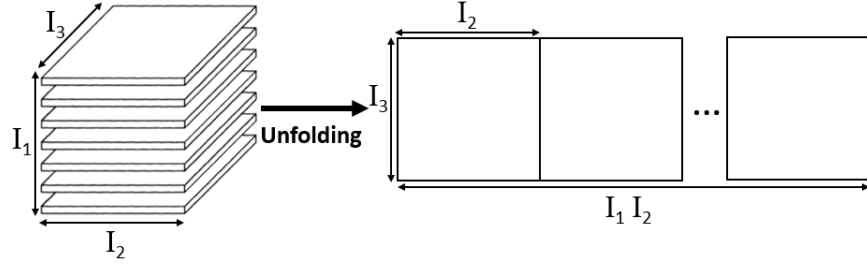
(c) Horizontal slices $\mathbf{A}_{(3)} \in \mathbb{R}^{I_3 \times I_1 I_2}$
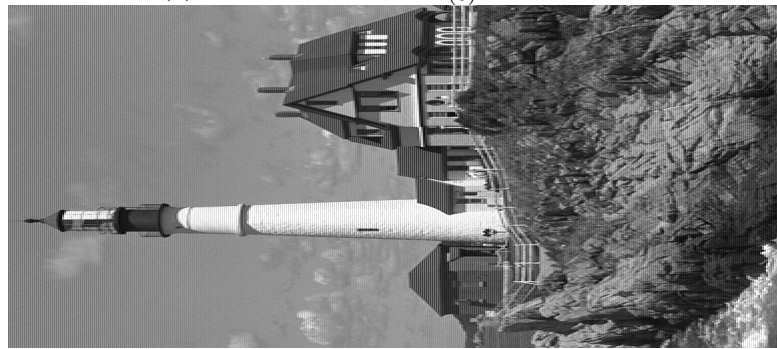
(d) Mode$_{(2)}$ unfolding of a color

Figure 2.16: An example of different unfoldings of a 3rd-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$

the exponential decay of the singular values. Fig. 2.17 shows two images with different degrees of blur and the plots of higher order singular values for the original image and its five blurred versions, from the TID2013 database. From the plots, we clearly notice that the degree of the exponent varies with the degree of blur. The correlation between the degree of blur and the decay parameter is even stronger than the case of the luminance alone as considered in previous work and will be shown in our experiments. Throughout all the experiments, we observed a higher correlation between the subjective score and the blur metric computed using HOSVD of mode$_{(2)}$ unfolding. The reason is that mode$_{(2)}$ unfolding contains more spatial and inter-channel correlations than other modes. Therefore, we perform HOSVD only on mode$_{(2)}$ unfolding. The proposed technique is summarized here:

Step 1: For an RGB image, perform matrix unfoldings, $\mathbf{A}_{(2)} \in \mathbb{R}^{I_1 \times I_2 I_3}$

Step 2: Take HOSVD of unfolding $\mathbf{A}_{(2)}$. $\mathbf{A}_{(2)} = \mathbf{U}^{(2)}\mathbf{\Sigma}^{(2)}\mathbf{V}^{(2)T}$, where $\mathbf{\Sigma}^{(2)}$ is a diagonal matrix with higher order singular values corresponding to mode$_{(2)}$ unfolding of the color image.

Step 3: Compute the blur metric using (2.25).

2.7.3 Experimental Results

We carried our experiments on four publicly available and commonly used image quality databases i.e., CSIQ [33], LIVE2 [31], TID2013 [38], and the newly available database CID:IQ [39]. Since the method deals with blur assessment in color images, we only used the blur distorted images from these databases for the experiments. The performance was evaluated using the SROCC, the PLCC, and the RMSE. High values of SROCC and PLCC and low values of RMSE correspond to close relationship of the objective scores to the subjective ratings.

A 5-parameter logistic fitting function [42] is used for the calculation of PLCC

and RMSE to account for the non-linearity in the subjective scores due to human opinions. The fitting function we used here is:

$$Q(q) = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(q - \beta_3))} \right] + \beta_4(q) + \beta_5 \qquad (2.27)$$

where Q represents the fitted objective score after non-linear mapping, q is the calculated objective quality score, and β_k for $k = 1, 2, 3, 4, 5$ are fitting parameters. These parameters are calculated by minimizing the mean-squared error between the subjective scores and the fitted values. The predicted objective scores (after non-linear mapping) and subjective scores (MOS/DMOS) for the blurred images as well as the original images from the CSIQ, LIVE, TID2013 and CIDIQ databases are shown in scatter plots in Figure 2.18. In these plots, the objective scores are kept on x-axis while y-axis represents MOS values for the blurred images represented as plus symbols $(+)$.

We have compared our results to different state-of-the-art NR-IQA techniques that are related to blind blur assessment. The results of SROCC, PLCC, and RMSE for different databases are shown in Table 2.7. The proposed metric gives better performance for three of the four databases. For the LIVE2 database, the proposed metric has also consistent performance except for [90], where the results are very much comparable. From the weighted average, it is also evident that the proposed metric achieves the best overall performance in terms of prediction accuracy and monotonicity at the cost of slight increase in computational complexity.

(a)   (b)
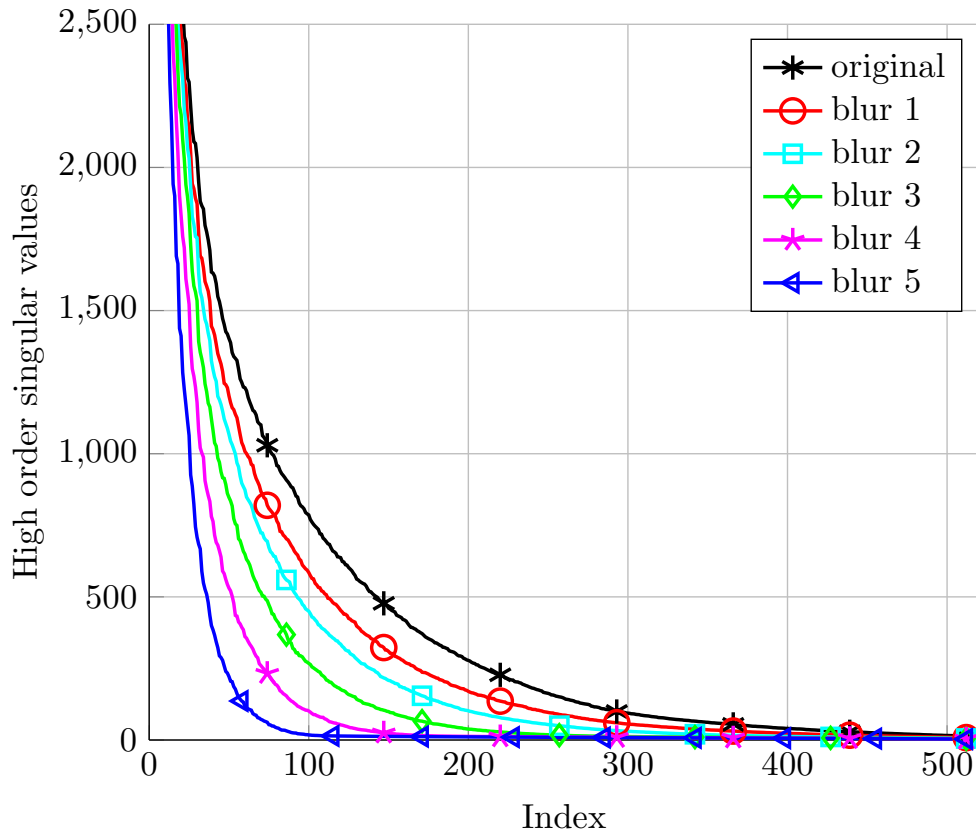
(c)

Figure 2.17: Sample images with different degrees of blur from the TID2013 database [2]: blur 4, MOS = 3.23, $\alpha = 1.48$ (a), blur 5, MOS = 2.16, $\alpha = 1.70$ (b), singular value curves for blurred images using HOSVD (c)

Table 2.7: Summary of different evaluation metrics on four databases, the first two best results are highlighted

Database	Evaluation Criteria	JNB [86]	CPBD [87]	SRE [65]	NR-SV [89]	NSVD [90]	HOSVD
CSIQ [33]	SROCC	0.7884	0.8904	0.8452	0.9167	**0.9196**	**0.9334**
	PLCC	0.8316	0.9188	0.8826	**0.9433**	0.9417	**0.9543**
	RMSE	0.1662	0.1181	0.1407	**0.0994**	0.1007	**0.0894**
LIVE2 [31]	SROCC	0.8416	0.9425	0.9030	0.9454	**0.9502**	**0.9497**
	PLCC	0.8480	0.9154	0.8912	0.9370	**0.9484**	**0.9375**
	RMSE	12.7048	9.6520	10.8732	8.6241	**8.0151**	**7.5703**
TID2008 [32]	SROCC	0.6667	0.8412	0.7511	**0.9089**	0.8933	**0.9163**
	PLCC	0.6939	0.8325	0.7689	0.9190	**0.9218**	**0.9461**
	RMSE	0.8458	**0.6502**	0.7503	0.7246	0.7495	**0.6591**
TID2013 [38]	SROCC	0.7837	0.8929	0.8152	**0.8970**	0.8833	**0.9086**
	PLCC	0.8014	0.8774	0.8023	**0.8839**	0.8833	**0.8928**
	RMSE	1.0925	0.8826	1.0979	**0.8611**	0.8623	**0.8287**
CID:IQ [39]	SROCC	0.5057	0.7829	0.7120	**0.8386**	0.8188	**0.8730**
	PLCC	0.5153	0.7846	0.7542	**0.8672**	0.8499	**0.9008**
	RMSE	1.4988	1.0842	1.1484	**0.8708**	0.9226	**0.8868**
Weighted Average	SROCC	0.7409	0.8820	0.8252	**0.9031**	0.8977	**0.9190**
	PLCC	0.7609	0.8793	0.8385	**0.9113**	0.9101	**0.9239**
	RMSE	4.0673	3.0883	3.4897	2.7536	**2.6003**	**2.4611**

Table 2.8: Results for different color spaces on CSIQ database

Color Space	SROCC	PLCC	RMSE
RGB	0.9334	0.9543	0.0894
CIELab	0.9321	0.9537	0.0899
YCbCr	0.9321	0.9554	0.0884

To further assess the performance of our metric using tensors, we also used two other color spaces shown to rely on weakly correlated components, namely the CIELab and YCbCr color spaces. The results we obtained were consistent across different color spaces whether these rely on strongly or weakly correlated components. To save space, we report in Table 2.8, our results for the CSIQ database comparing RGB, CIELab, and YCbCr color spaces. We note again that the results are consistent across different color spaces.

2.7.4   Discussions

A novel methodology for quantifying blurring effects in color images using the concept of higher order singular values is proposed. The spatial and inter-channel correlations, in the color image, are exploited using tensors to quantify the amount of blur more efficiently and consistently rather than using the traditional luminance component only or the individual color channels in existing techniques. A color image is considered as a 3rd-order tensor and decomposed into a superset of 2D matrices or so-called unfoldings. The higher order singular values are calculated for these unfoldings using conventional SVD. SVD is a mathematical concept used to decompose 2D matrices (or images) into a sequence of three transformations (two rotations and one scaling). The scaling is used to quantify the amount of variability of the given 2D data in the two main directions. We show in the proposed work that the extracted higher order singular values consistently follow an exponentially decreasing curve. Moreover, we show that the degree of such
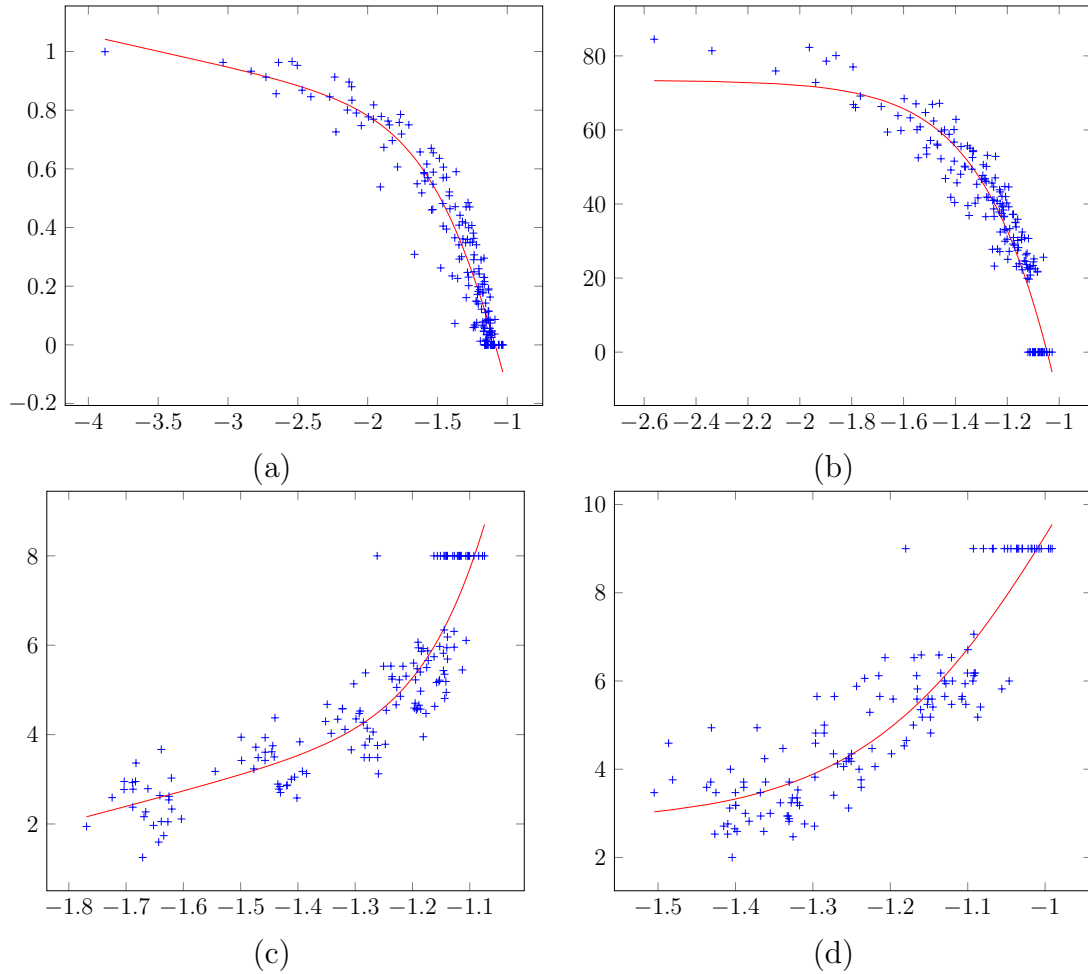
Figure 2.18: Regression plots between subjective and predicted scores for blur distortion in databases (a) CSIQ (b) LIVE2 (c) TID2013 (d) CID:IQ

exponential decay varies closely with the amount of blur a given image is subject to. Our experimental results, performed on different public IQA databases, validated the power and consistency of the proposed metric across different color spaces compared to state-of-the-art no-reference blur assessment metrics. The proposed technique could be embedded in camera sensors to provide photo after blur removal and could be use used in multimedia applications for best quality of experience delivery of videos to the end users. It is expected to create new research opportunities for researchers in the field of IQA and multimedia industry.

2.8 Summary

The notion of visual image quality is highly related to the way humans perceive distortions that may affect the quality of the observed image. The IQA dilemma, has been long considered as a distortion estimation problem. On the other hand, very few studies have been carried on the performance evaluation of image enhancement methods (better quality images rather than distorted images). Looking at the challenging nature of CEE in different applications, we have made some contributions under this category and these are discussed in details in the next Chapter.

CHAPTER 3

IQA FOR ENHANCEMENT EVALUATION

3.1 Introduction

Enhancing image contrast is of major interest in many applications ranging from medical imaging [19], remote sensing [20], underwater imaging [21], defogging [22], etc. A plethora of CE methods has been proposed in the literature, and it becomes rather difficult to provide a comprehensive and complete survey of published work in this area. On the other hand, very few studies have been done on the performance evaluation of image enhancement methods (better quality images rather than distorted images). Indeed, performing a quantitative evaluation of image quality enhancement methods is a very challenging task. This is due to the absence of any objective measures able to account for some high-level vision tasks and their interaction with low-level image analysis when assessing the perceptual quality of image enhancement [18]. This is also due to the difficulty in determining the most appropriate visual features to be used in the design of an overall image enhancement quality measure. Therefore, subjective evaluation is still the most reliable approach to assess the quality of enhanced images. Moreover, there is

no study to test the reliability of objective CEE measures themselves. Given the importance of CE in different applications, there is a need to investigate the performance of these measures in terms of robustness and consistency with human judgment.

3.2   Related Work

3.2.1   Existing Contrast Enhancement Databases

One of the first studies on CEE has been proposed in [96]. However, it was only restricted to images containing two classes of pixels (i.e., one object on a uniform background or many similar objects on a uniform background). The CE evaluation was based on the bimodality analysis of the gray-level distribution. Thereafter, some simple and interesting CEE measures have been proposed in [97, 98, 99, 100]. These measures are not inspired by the classical approaches of IQA. The proposed measures are based on the computation of a global index derived from some local measures related to contrast. These are inspired originally by Michelson and Weber-Fechner contrast measures. These measures are based on min-max operations that make them more noise sensitive. The authors proposed some improvements to overcome these limitations by using entropy of local contrast, or by introducing logarithmic arithmetic operations inspired by the non-linear HVS response. In the study conducted in [97, 98, 99, 100], no complete subjective experiments were performed, and the performance analysis was only based on the perceptual judgment of output images. Moreover, the tests were conducted on a limited set of images (very often grayscale images), and the measures were not evaluated on any dedicated database but only on few images from the TID2013 database that has been built for traditional IQA purpose [38]. Furthermore, the statistical analysis of these measures and comparison with some representative

CE methods were also missing.

In contrast, Damon et al. [101], proposed another study based on a database containing processed images obtained by changing color, saturation, brightness, sharpness, and their combinations. The subjective evaluation was performed to assess the quality of processed images. The use of classical IQA approaches in a reverse order was proposed, i.e., the given image (enhanced image) is considered as the reference and the original image as the distorted one. It has also been reported that the Visual Information Fidelity (VIF) [102] measure offers better performance as compared to many of the classical IQA measures. The authors in [101] improved the results by proposing a more efficient measure combining contrast, sharpness, and color in an empirical manner.

Following the approach of Damon et al. [101], another study of contrast change evaluation was discussed in [103] using a database consisting of 15 original and 633 enhanced images. The global contrast of images is modified using non-linear mapping functions. The conventional IQA measures designed for degradations assessment were then used to assess the quality of the processed images from the database. For this purpose, a RR metric was derived combining the entropy of phase congruency image and other higher-order statistics of local features computed from the histogram of the observed image. However, the enhancement evaluation methods based on conventional IQA approach are not appropriate for CE measure evaluation. Indeed, for example, the approach followed by Damon et al. [101] is not convincing and could produce contradicting results. Indeed, an image with good perceptual quality, in the sense of traditional FR-IQA context, is considered as very similar to the original one. For example, it would correspond to a SSIM value near one, which does not serve the objective of CE. Whereas, in the context of CE, the objective is not to process the image so as to make it as close as possible to the original image.

Another recent study, by Fang et al. [104] on quality assessment of contrast distorted images was carried using the natural scene statistics model. The contrast problem is considered only in terms of distortion.

Besides these works, predicting visual quality of enhanced images for different applications has also been investigated in some interesting studies [105, 40, 106, 107, 108, 39]. Ledda et al. [105] proposed a database for only subjective evaluation of six tone mapping methods. The PC was performed in a subjective experiment to rank these methods in accordance with the perceived quality. But the authors did not perform CEE performance analysis. Virtanen et al. [40] provided another database related to tone-mapping applications. It contains images degraded with different types of distortions and images with variation of contrast due to gamut mapping. The main objective of the database was to validate the performance of existing IQA metrics designed mainly for degraded images. Another similar database was also proposed in [39] to evaluate gamut mapping, blurring, and other distortions.

In addition to the above, Chen et al. [108] developed a database for CEE of images in bad visibility (i.e., haze, underwater, and low light environment). The images were enhanced through different dehazing methods and the performance of various enhancement algorithms was discussed. In this work, the original and pair of enhanced images were shown on the same screen to allow the observer to compare the enhanced images with respect to the original image.

Another less studied application, namely image retargeting quality assessment, has been addressed in [106, 107]. Here, subjective and objective quality evaluation of retargeted images was performed using dedicated databases. In [106], the authors provided a database containing images by different retargeting methods. The subjective quality of the retargeted images was measured in terms of rank in a pairwise subjective experiment, and the performance of different retarget-

ing evaluation measures were assessed in terms of correlation analysis. Similarly, Ma et al. [107] also carried out the same study except, instead of ranking, they provided the rating scores on a different proposed database. To summarize the related works carried to date, we provide, in Table 3.5 and 3.6, our own perspective on the main contributions made in this field of research.

3.2.2   Contrast Enhancement Evaluation Measures (CEE)

The improvement in image quality after CE can be evaluated using a multitude of objective measures. Although, we can see a lot of research efforts towards the development of CE algorithms, the objective CEE measures are limited and specific to different applications. The CE evaluation is different from conventional IQA. The reason is that in conventional IQA, the image which is similar to the original is considered as of good quality, and the similarity decreases with the increase in degradation. It is worth noticing that when using classical IQA, like SSIM, which is FR metric for quality assessment of degraded images, its value is close to one, when there is no distortion in an image and its value is less than one in the case of a degraded image. However, in the case of CE, we start from an input image and try to improve its quality. This processing is expected to produce more visible structures and the obtained images are rather different from the original one. If we use the SSIM for the contrast enhanced image, it will give value less than one, which does not correspond to an image of good quality. It has been observed that only the VIF measure [102], which is based on classical FR-IQA approach, yields interesting results. Indeed, the VIF produces a value less than one for the degraded images and greater than one for the case of enhancement.

Damon et al. [101], proposed that, to assess the quality of enhanced images, one can use the given image (enhanced image) as the reference and the original image as the distorted one and apply conventional IQAs. Whereas, in the case of

NR-IQA, the CEE measures are derived from the given image. Some measures like sharpness, blurriness, SVD based measures, details visibility map, after CE could be used to derive NR-CE quality measures. Recently, Fang et al. [104] used natural scene statistics to quantify the quality of contrast-enhanced images by looking the enhancement process as degradation process and apply the conventional IQA for the contrast distorted images in classical IQA databases. But in general, it is not applicable to CE applications.

In this section, we provide a brief overview of the measures used in our study. For the sake of completeness, we also provide the mathematical expressions for the measures as well. Based on the availability of the original image, we can group these measures into two broad classes, i.e., FR and NR measures (see Fig. 3.1). Moreover, based on the methodology used, we have also categorized these measures into Statistics-based, Gradient/Energy-based, and HVS-inspired CE evaluations (see Fig. 3.2). These measures are usually derived from grayscale images. For color images, the luminance component is used for contrast assessment. In this work, we adopt some state-of-the-art measures and our aim is to investigate how well these measures are consistent with the human judgment of quality. These measures are computed using only the luminance component of images. The mathematical expressions for the CEE measures are also provided in Tables 3.2, 3.3, and 3.4. To be consistent with the use of variables in the mathematical expressions of CEE measures, we list the description of each variable in Table 3.1. In the following, we start with a brief description of each category of CEE measures.
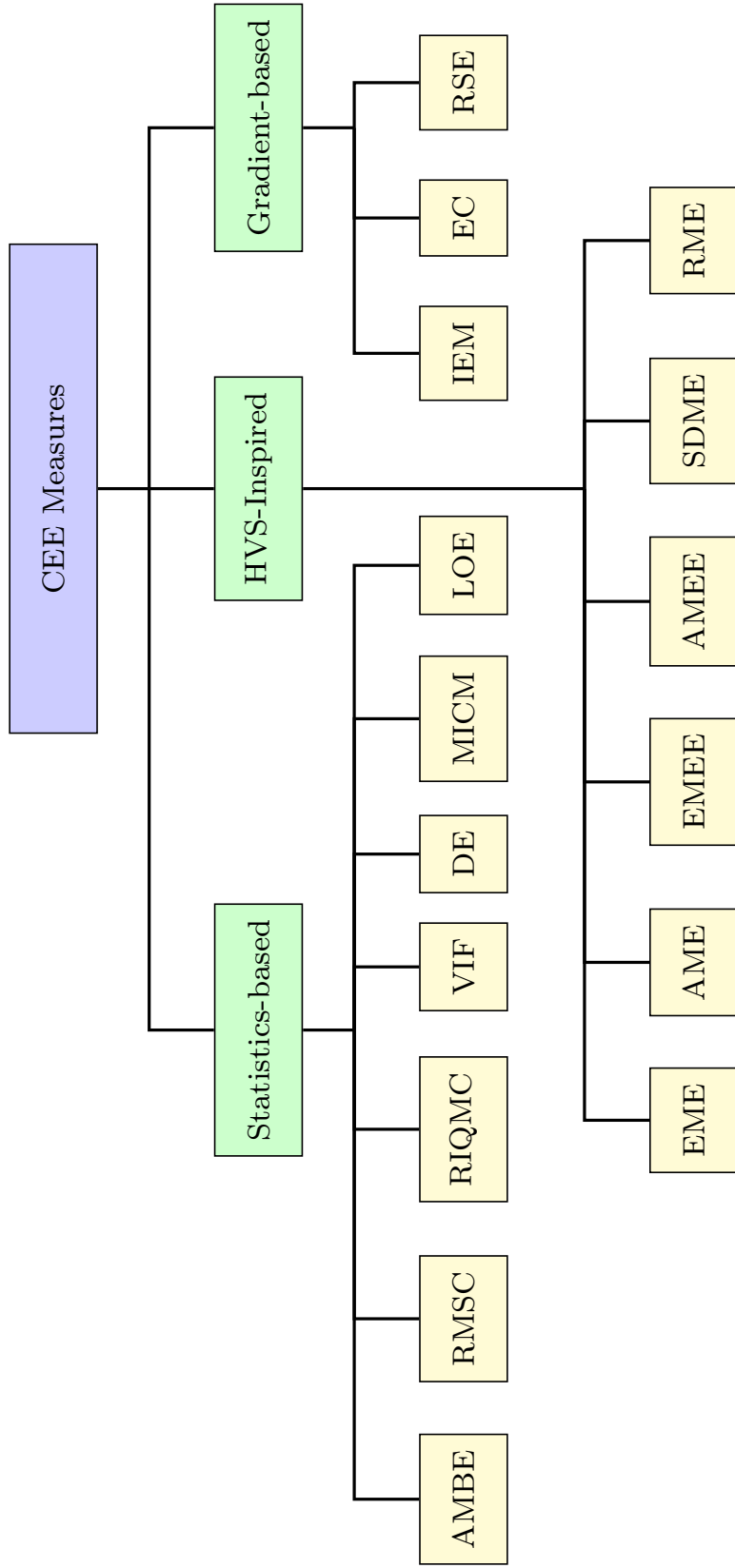
Figure 3.1: Proposed framework for categorization of different CEE measures based on the availability of original image
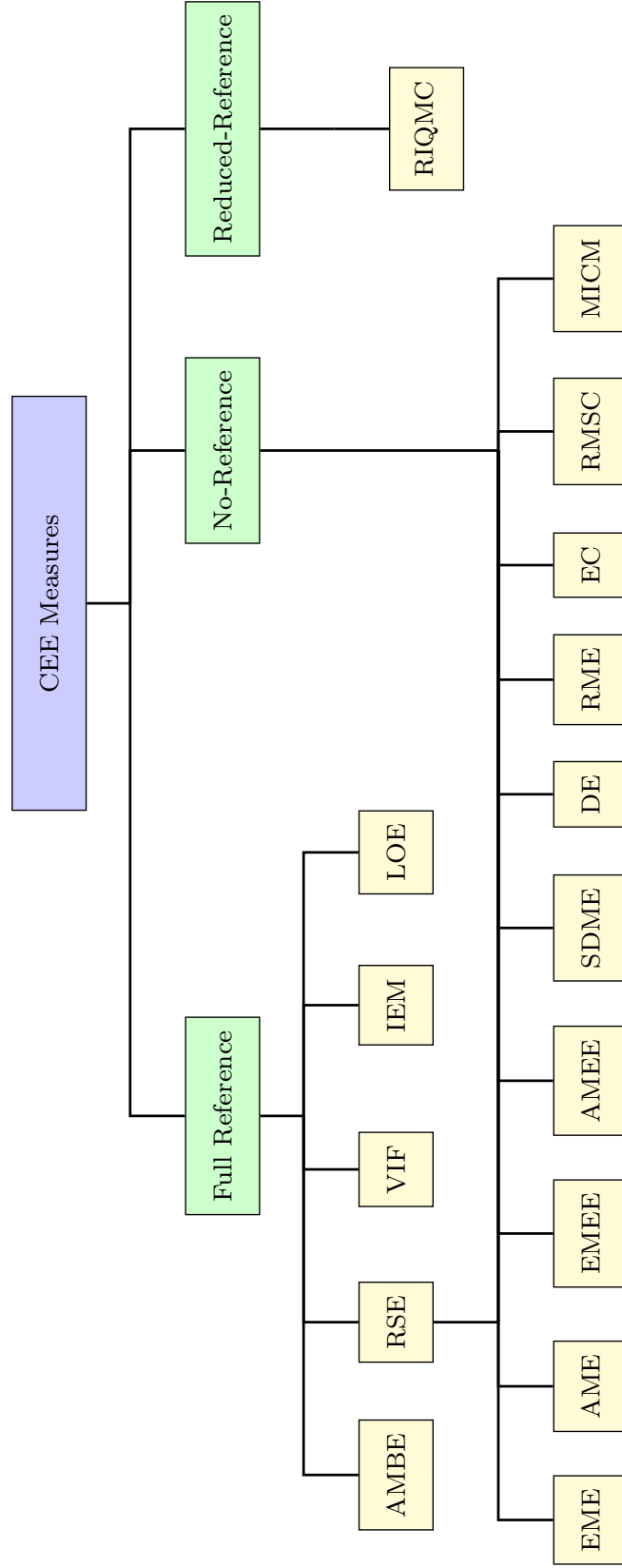
Figure 3.2: Proposed framework for categorization of different CEE measures based on the methodology used

Statistics based CEE Measures

Absolute Mean Brightness Error (AMBE): It is used to evaluate how much original brightness is preserved in the enhanced image [109]. It is calculated as the deviation of the mean intensity of the enhanced image from that of the original image. For CE, it is desirable that the original brightness of an image is to be preserved. The lower value of AMBE means that the enhanced image has good brightness preservation. Here, brightness preservation does not mean that the image natural look (quality) also preserves. Either a very low value or the highest value of AMBE also indicates poor performance in case of CE.

Root Mean Square Contrast (RMSC): It is a pixel-based NR metric [110]. It gives high values for the images containing a major bright portion (e.g., sky, sea, etc.). High values of RMSC correspond to image with better contrast. However, it is not considered as an effective measure of CE since its value also increases with the appearance of some undesirable artifacts and noise amplification.

Reduced-reference Image Quality Metric for Contrast change (RIQMC): It is a RR metric used to quantify image contrast and naturalness [103]. It combines the entropy of phase congruency image, and four statistical features computed from image histogram (i.e., mean, variance, skewness, and kurtosis). The first order statistical feature, F_1, is computed by penalizing very large and very small mean values using the Gaussian kernel and is defined as follows:

$$F_1 = \exp\left[- \left(\frac{E(I_e) - \mu}{\beta}\right)^2\right] \tag{3.1}$$

where μ and β determine the mean and shape of the Gaussian kernel.

The context-free contrast feature, F_2, defined as a function of variance com-

Table 3.1: Notations used for CEE measures.

Notation	Description
I_r	original image
I_e	enhanced image
H	image height (rows)
W	image width (columns)
b	block size
i, j	pixel indices
L	number of gray levels
I_{ij}	image pixel value at index (i, j)
$\bar{I}$	mean pixel value in an image
c	constant (c = 0.0001) to avoid division by zero
$E(.)$	statistical expectation
B_1, B_2	number of blocks along rows and columns

puted from the image histogram is expressed as follows:

$$F_2 = E[p(I_e)^2] - E[p(I_e)]^2 \qquad (3.2)$$

where $p(I_e)$ represents the histogram of an enhanced image.

Similarly, the higher-order statistical features, i.e., skewness, F_3 and kurtosis, F_4, are computed as follows:

$$F_3 = \frac{E[I_e - E(I_e)]^3}{\sigma^3(I_e)} \qquad (3.3)$$

$$F_4 = \frac{E[I_e - E(I_e)]^4}{\sigma^4(I_e)} - 3 \qquad (3.4)$$

where σ is the standard deviation of the gray-levels in the image.

The similarity feature, F_5, defined as the difference between the entropy of phase congruency of enhanced image and original image:

$$F_5 = H_{PC}(I_e) - H_{PC}(I_r) \qquad (3.5)$$

where $H_{PC}(.)$ represents the entropy of phase congruency image.

Finally, the RIQMC is computed as a linear combination of the five features using the following expression.

$$\text{RIQMC} = \sum_{i=1}^{5} w_i F_i \tag{3.6}$$

where the weights, w_i, for $i = 0, 1, \cdots, 4$, represent the contribution of each feature in the final contrast metric.

The RIQMC fusion depends on the weights, but they are not provided in the text. Through experiments, we have observed that with increasing the contrast (improvement in quality), the RIQMC value decreases.

Visual Information Fidelity (VIF): It is a FR quality metric used to quantify the loss of original image information due to processing or transmission of the given image. The original and test images are decomposed into different subbands and the mutual information to be perceived by HVS from these subbands is calculated for both images. The measure is expressed as the fraction of original image information that can be perceived by HVS from the test image. VIF measures can be used as a quality metric for both degraded and enhanced images [102]. Its values are equal to, less than, and greater than one for the original, degraded, and enhanced images respectively.

Discrete Entropy (DE): It measures the amount of information or randomness of gray-levels in an image [111]. It is well known that the increase in contrast highlights the subtle details in an image and results in an increase in entropy value. It is a global measure based on the overall histogram of an image and fails to consider the local details and spatial correlations among the pixels. The higher

values of DE correspond to image with more details visibility amplification and is considered as image with good quality.

Mutual Information based Contrast Measure (MICM): It is a NR metric used to quantify the global image contrast and to detect and control the side effects of CE in few neighborhood-based methods [25]. It is based on mutual information derived from the joint probability mass function of a gray level co-occurrence matrix and is given by:

$$\text{MICM} = \sum_{i=0}^{L} \sum_{j=1}^{L} p_{ij} \log_2 \left(\frac{p_{ij}}{p_x(i)p_y(i)} \right) \tag{3.7}$$

where p_{ij} is the joint probability mass function of the luminance channel, whereas p_x and p_y represent the marginal probabilities calculated along the rows and columns of co-occurrence matrix respectively. It is better than 1st order entropy, we take into consideration the spatial correlation among the pixels using the gray level co-occurrence matrix. It is simple to compute, however, it does not provide information about image unnaturalness.

Lightness Order Error (LOE): It measures the naturalness preservation in the enhanced imaged based on estimating the lightness order error between the original image and enhanced image [112]. The lightness of an RGB image, I^c, is obtained by taking the maximum of the three color components.

$$L_{ij} = \max_{c \in \{r,g,b\}} [I_{ij}^c] \tag{3.8}$$

where r, g, and b represent the red, green, and blue color components in an RGB image.

The relative order difference of the lightness between the original image and

its enhanced version is calculated as follows:

$$RD_{ij} = \sum_{x=1}^{M}\sum_{y=1}^{N}(U(L_{ij}^{r}, L_{xy}^{r}) \oplus U(L_{ij}^{e}, L_{xy}^{e})) \qquad (3.9)$$

$$U(a,b) = \begin{cases} 1 & \text{for } a \geq b \\ 0 & \text{otherwise} \end{cases} \qquad (3.10)$$

where $U(a,b)$ is a unit step function and $\oplus$ is exclusive-or operator.

The final LOE measure is calculated as:

$$LOE = \frac{1}{(W \times H)}\sum_{i=1}^{W}\sum_{j=1}^{H} RD_{ij} \qquad (3.11)$$

In the original implementation, the downsampled version of both images was used to reduce the computational complexity with the downsampling ratio of $r = 50/\min(H, W)$.

Since the relative order of lightness represents the light source directions and the brightness variations, the naturalness of an enhanced image is related to the relative order of lightness in different local areas. Small values of LOE indicate that the naturalness is well preserved in an enhanced image in comparison with the original image. The authors claimed that metric well provides naturalness information. However, naturalness is a complex property and is difficult to define.

Gradient/Energy-based CEE Measures

The following measures are based on either local signal activity or energy as measured through the gradient operator or the spectral energy distribution. Indeed, any increase/decrease of the contrast inevitably affects the pixel intensity gradient and the spectral energy distribution in the spatial-frequency domain.

Image Enhancement Metric (IEM): It is a FR metric proposed by Jaya et al.

[113] and is calculated by subdividing an image into non-overlapping blocks. The ratio of the sum of absolute values of differences of the center pixel from its eight neighbors in all blocks of the enhanced image and the corresponding blocks in the original image represents the IEM value. The absolute intensity differences between a pixel and its neighbors corresponding to the reference and enhanced images are used to account for the change in contrast and sharpness. Typical values for the image blocks are 3×3 or 5×5. For identical images, IEM is equal to one. The values of IEM greater than one means image contrast and sharpness are increased.

Edge Content (EC): It is a blind objective measure based on the local gradient of the image intensity [114]. In its expression, $\Delta I\left(.\right)$ represents the gradient magnitude of the pixel value computed from the Sobel edge operator. Higher values of EC correspond to images with more contrast. The overall EC value for a complete image is calculated by averaging the local EC values for each block.

Radial Spectral Energy (RSE): It is based on radial spectral energy analysis developed for blind image sharpness assessment [65]. It is based on the idea that the effect of adding a certain amount of blur to a given image depends on the original quality of this image. In other words, a contrasted image is more sensitive to blur effect than a less contrasted image. The enhancement measure is computed as the variation of the radial spectrum due to contrast enhancement. The radial energy on the original image and its contrasted version are calculated as follows:

$$E_{R_r}(\omega) = \frac{1}{K} \sum_k |\mathcal{I}_r(\omega, \theta_k)| \qquad (3.12)$$

$$E_{R_e}(\omega) = \frac{1}{K} \sum_k |\mathcal{I}_e(\omega, \theta_k)| \tag{3.13}$$

where $\mathcal{I}(.)$ is the Fourier transform of the image signal $I(.)$ at a particular radial frequency ω and in θ_k direction, $\theta_k = \dfrac{k\pi}{K}$ and $\omega = \sqrt{(u^2 + v^2)}$ where u, v are spatial frequencies, and K is total number of directions.

Then the blur index is computed as:

$$\text{RSE} = \log \left(\frac{1}{\omega_{max}} \sum_\omega |E_{R_r}(\omega) - E_{R_e}(\omega)| \right) \tag{3.14}$$

where ω_{max} is the maximum radial frequency within the image and can be calculated as $\omega_{max} = \sqrt{u_{max}^2 + v_{max}^2}$, where u_{max} and v_{max} are the maximum values of spatial frequencies u and v. The $\log(.)$ is used in the expression to make the measure non-linear in accordance with HVS response. An increase of RSE corresponds to increase in contrast.

HVS-Inspired CEE Measures

Some simple CEE measures have been proposed in [97, 98, 99, 100, 115]. These approaches are not inspired by the traditional IQA measures as suggested by [101]. The proposed measures are based on the computation of a global index derived from some local measures related to contrast and gradient. These CEE measures are mainly inspired by the Michelson and Weber-Fechner contrast measures which are not really adapted to natural scenes. These measures have been evaluated on a limited set of images processed by some CE methods. However, these studies do not provide a comprehensive analysis of the validity of these measures on various images and different CE methods. The main CEE measures of this class are now briefly discussed.

Measure of Enhancement (EME): The EME was proposed by Agaian et al. [97] and is a NR metric based on a contrast measure using the pixel value dynamic range (min-max values) within a block. The image is first divided into non-overlapping blocks of the same size (say 8×8). The EME value is computed based on the minimum and maximum pixel values in each block, respectively. The overall measure is computed by averaging the local EME values for image blocks. Since log of ratios of maximum and minimum intensities within each block can be written as difference, EME may represent signal dynamic range of the image. EME increases with the increase in image contrast.

Measure of Enhancement by Entropy (EMEE): The EMEE measures the entropy in the local contrast as defined in [98]. It also increases with the increase in image contrast. The use of entropy is motivated by the fact that any small variation in the contrast would convey additional amount of information on the spatial content of the image. This consequently would affect the entropy value.

Absolute Measure of Enhancement (AME): Similarly to EME, the AME [99], is also a block-based logarithmic Michelson contrast based measure. The AME decreases with the increase in image contrast.

Absolute Measure of Enhancement by Entropy (AMEE): The AMEE measures the entropy in the local Michelson contrast of an image as defined in [99]. It increases with the increase in contrast. The reason for using the entropy is also the same as for EMEE measure.

Second Derivative like MEasurement (SDME): This measure is based on the fact that the local contrast is highly related to the local variations of the signal

[100]. This could be captured by any derivative operator, here, a pseudo-second order derivative operator is used. It is also a block-based measure with default block size either 3×3 or 5×5. The authors claimed that this measure is less noise sensitive than the other similar measures based on only min-max operations. It decreases with an increase in image contrast.

Root Mean Enhancement (RME): It incorporates both RMS contrast and properties of HVS [98]. It measures the relative RMS contrast in the log domain. It is calculated by subdividing an image into non-overlapping blocks (say 3×3 or 5×5). For low contrast images, RME value is small, whereas it is large for high contrast images.

It is also worth noting that the goal of CEE measures should not be limited to quality assessment but also to provide a quantitative measure that could be used to control some unpredictable after-effects due to CE. The side effects due to CE are color mismatch, color bleeding, saturation, overshooting, halo effects, blocking/ringing artifacts amplification, and other undesirable effects. The existing CEE measures either increase or decrease with the increase in contrast and none of the measures could predict the side effects due to CE. In this regard, we proposed a NR metric based on mutual information computed from gray level co-occurrence matrix to show how this measure may help in evaluating the artifacts in CE processes. The details of this work are provided in the next section.

Table 3.2: The expressions of the Statistics-based CEE measures

CEE measures expressions	Type
$\text{AMBE}[109] = \lvert E\left(I_r\right) - E(I_e) \rvert$	FR
$\text{VIF}[102] = \dfrac{\displaystyle\sum_{k \in subbands} I_M(\vec{C}^{N,k} \vert \vec{F}^{N,k})}{\displaystyle\sum_{k \in subbands} I_M(\vec{C}^{N,k} \vert \vec{E}^{N,k})}$	FR
$\text{RMSC}[110] = \sqrt{\dfrac{1}{HW-1}\displaystyle\sum_{i=1}^{W}\sum_{j=1}^{H}\left(I_{ij} - \bar{I}\right)^2}$	NR
$\text{DE }[111] = -\displaystyle\sum_{x=0}^{255} p(x)\log_2 p(x)$	NR

$^-$ $I_M(\vec{C}^{N,k}\vert\vec{F}^{N,k})$ and $I_M(\vec{C}^{N,k}\vert\vec{E}^{N,k})$, represents the mutual information that can be extracted from a particular wavelet subband k in the original (F) and test (E) images respectively, C represents the wavelet coefficients.
$^-$ $p(x)$ represents the normalized image histogram.

3.3 A New No-Reference CEE Measure based on Mutual Information

Contrast enhancement in its broad sense is considered as a process by which some characteristics of the image signal are highlighted. The objectives of image enhancement differ and depend on the considered applications. From the point of view of purely signal processing perspective, enhancing contrast signal may produce interesting results but in the same time it may generate some undesirable effects from the perceptual image quality aspects. Indeed, for example global approaches, such as histogram based CE methods [116, 117], may produce saturation in some dark and bright zones and consequently reduce the visibility of some details. Neighborhood or local based methods have been developed to overcome the limitation of global methods [118, 119]. However, local analysis based methods may also produce noise amplification, overshooting, color mismatch, blocking effect accentuation, and other undesirable effects [120, 121]. Therefore, looking

Table 3.3: The expressions of the Gradient-based CEE measures

CEE measures expressions	Type
$\text{EC}[114] = \dfrac{1}{H \times W} \sum\limits_{i=1}^{W} \sum\limits_{j=1}^{H} \lvert \Delta I(i,j) \rvert$	NR
$\text{RSE}[65] = \log \left(\dfrac{1}{\omega_{max}} \sum\limits_{\omega} \lvert E_{R_r}(\omega) - E_{R_e}(\omega) \rvert \right)$	NR
$\text{IEM}[113] = \dfrac{\sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} \sum\limits_{n=1}^{8} \lvert I_{ij}^{e,c} - I_{ij}^{e,n} \rvert}{\sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} \sum\limits_{n=1}^{8} \lvert I_{ij}^{r,c} - I_{ij}^{r,n} \rvert}$	FR

$^-$ I_{ij}^{max}, I_{ij}^{min}, and I_{ij}^{cen} are the maximum, minimum, and center pixel intensity within the block (i,j), respectively.

$^-$ n represents pixel neighborhood index

$^-$ E_{R_r} and E_{R_e} represents Radial Spectral Energy of reference and enhanced image respectively, ω is the radial frequency.

for some strategies to control the image CE is really very useful in many applications. One of the most challenging problem is then not only to develop objective measure for CEE but also and more importantly to control the effects of CE on the perceptual quality of images. This control process should be consistent with the subjective appreciation of the treated images. Some simple CEE measures have been proposed in the literature [109, 114, 122, 97, 123, 98]. But to our best knowledge there are very few works dedicated to the development of measures that could be used to control the side effects of CE. The intent of our work is to propose a new framework for quantifying the side effects of some CE methods and especially local based methods which has been proven more efficient than global ones. Here we limit the study to two representative methods of neighborhood-based approaches, namely Adaptive Edge-based Contrast Enhancement (AEBCE) method [124] and a new unsharpening method introduced in this chapter for the first time which we termed Extended Unsharpening Method (EUM). The framework introduced here is based on mutual information concept [125]. The idea

Table 3.4: The expressions of the HVS-Inspired CEE measures

CEE measures expressions	Type
$\text{EME}[97] = \dfrac{1}{B_1 \times B_2} \sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} 20 \ln \left(\dfrac{I_{ij}^{max}}{I_{ij}^{min} + c} \right)$	NR
$\text{EMEE}[98] = \dfrac{1}{B_1 \times B_2} \sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} \alpha \left(\dfrac{I_{ij}^{max}}{I_{ij}^{min} + c} \right)^{\alpha} \ln \left(\dfrac{I_{ij}^{max}}{I_{ij}^{min} + c} \right)$	NR
$\text{AME}[99] = \dfrac{-1}{B_1 \times B_2} \sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} 20 \ln \left(\dfrac{I_{ij}^{max} - I_{ij}^{min}}{I_{ij}^{max} + I_{ij}^{min}} \right)$	NR
$\text{AMEE}[99] = -\dfrac{1}{B_1 \times B_2} \sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} \alpha \left(\dfrac{I_{ij}^{max} - I_{ij}^{min}}{I_{ij}^{max} + I_{ij}^{min}} \right)^{\alpha} \ln \left(\dfrac{I_{ij}^{max} - I_{ij}^{min}}{I_{ij}^{max} + I_{ij}^{min}} \right)$	NR
$\text{SDME}[100] = \dfrac{-1}{B_1 \times B_2} \sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} 20 \ln \left| \dfrac{I_{ij}^{max} - 2I_{ij}^{cen} + I_{ij}^{min}}{I_{ij}^{max} + 2I_{ij}^{cen} + I_{ij}^{min}} \right|$	NR
$\text{RME}[98] = \dfrac{1}{B_1 \times B_2} \sum\limits_{i=1}^{B_1} \sum\limits_{j=1}^{B_2} \left| \dfrac{\log |I_{ij}^{cen} - \bar{I}_b|}{\log |I_{ij}^{cen} + \bar{I}_b|} \right|$	NR

- I_{ij}^{max}, I_{ij}^{min}, and I_{ij}^{cen} are the maximum, minimum, and center pixel intensity within the block (i, j), respectively.
- $\bar{I}_b$ is the average pixel value within the block centered at index (i, j)

of using information-based measure is dictated by the fact that any process that tends to enhance image contrast would inherently change the spatial dependency relations between pixels and therefore global and local spatial information contained in the image. It is shown through this study that, this information based framework opens a new promising approach for quantifying and controlling the CE side effects such as over-shooting or halo effects, by blindly determining the critical point where over enhancement starts. The proposed framework is briefly introduced and illustrated on some real color images of various contents. The new mutual information measure is analyzed on the two mentioned CE methods AEBCE and EUM.

3.3.1 Preliminaries

The Co-occurrence matrix

The proposed metric is based on the grey-level co-occurrence matrix [126]. This is motivated by the fact that the co-occurrence matrix has been widely used in image analysis and especially for texture analysis and classification [127]. From the co-occurrence matrix, we compute the contrast, joint entropy and the mutual information for the Luminance component of color images in CIELAB space.

The proposed metric derived from information theory will be used for tracking image quality enhancement. We show that the variations of pixel intensity distribution are well visible on the co-occurrence matrix as shown in Figure 3.3 where we consider a typical case of local CE and blurring effect on natural images. We observe that the width of ellipse in the 2D histogram plot of the co-occurrence matrix changes with enhancement. For a blurred image, the ellipse width is thinner compared to the original image, and increases when the image contrast is increased as shown in Figure 3.3 (e) and (f) respectively.

For the sake of completeness, we recall in the following some basic notions used in this section. Let $[I]$ be an image of size $H \times W$, i.e. where each pixel (x, y) can take values $I(x, y)$ in the range $[0, K - 1]$, where K = 256, for 8-bit per pixel. The co-occurrence matrix $[C]$, is computed by examining all the pair of pixels in the image situated at a given distance and direction. The element of this matrix is defined as follows:

$$C_{ij}(r, \theta) = \sum_{x=1}^{H} \sum_{y=1}^{W} \delta(x, y) \tag{3.15}$$

where

$$\delta(x, y) = \begin{cases} 1 & \text{if } I(x, y) = i \text{ and } I(x + \Delta x, y + \Delta y) = j \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$
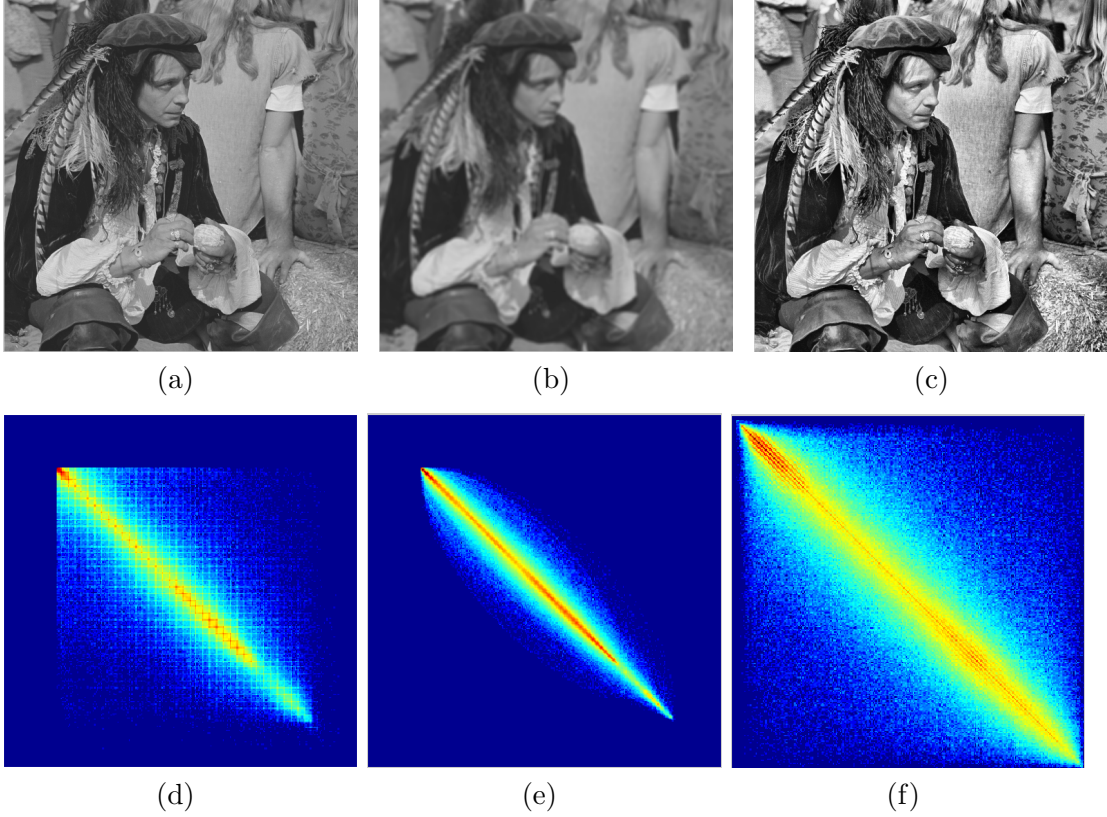
Figure 3.3: Effects of CE and blur on the co-occurrence matrix (a) Original image (b) Gaussian blurred image (c) Contrast enhanced image. (d)-(f) 2D histograms of co-occurrence matrices for images (a)-(c) respectively.

$r^2 = (\Delta x^2 + \Delta y^2)$, $\theta = tan^{-1}\left(\frac{\Delta x}{\Delta y}\right)$, and C_{ij} represents the frequency of occurrences of the gray level j adjacent to gray-level i in θ direction and at distance r.

In what follows, we will omit the parameters r and θ. So, for a given direction and distance, the second order probability mass function could be estimated as:

$$[p_{ij}] = \frac{C_{ij}}{\sum\limits_{k=0}^{K-1}\sum\limits_{l=0}^{K-1} C_{kl}} \tag{3.17}$$

Adaptive Edge-based Contrast Enhancement (AEBCE)

It is based on local CE using edge information [124]. For a given image, I, the local contrast in a window of size $w \times w$ centered at pixel location (k, l) is defined as follows:

$$C_{kl} = \frac{|I_{kl} - \overline{E}_{kl}|}{|I_{kl} + \overline{E}_{kl}|} \tag{3.18}$$

$$\overline{E}_{kl} = \frac{\sum_{(i,j)\in w_{kl}} I_{ij}\Phi_{ij}}{\sum_{(i,j)\in w_{kl}} \Phi_{ij}} \tag{3.19}$$

where $\overline{E}_{kl}$ is the mean edge-gray value computed over the local window, and Φ_{ij} represents edge value of pixel (i, j), computed from any gradient operator (e.g. Sobel, Prewitt etc.). The image contrast is then increased by applying any increasing and bounded to $[0, 1]$ function on the local contrast such as:

$$C'_{kl} = T(C_{kl}) \text{ such that } 0 \leq C'_{kl} \leq 1 \text{ and } C'_{kl} \geq C_{kl} \tag{3.20}$$

The new image pixel value is computed as follows:

$$I'_{kl} = \begin{cases} \overline{E}_{kl}\frac{1-C'_{kl}}{1+C'_{kl}} & \text{if } I_{kl} \leq \overline{E}_{kl} \\ \overline{E}_{kl}\frac{1+C'_{kl}}{1-C'_{kl}} & \text{otherwise} \end{cases} \tag{3.21}$$

For color images, the CE is applied to each color component. The edge based CE method closely resembles human visual perception. However, caution must be taken to avoid over enhancement or other unstable effects such us halo or noise amplification. For automatic CE, it is desirable to find an optimal point where the saturation starts to appear as halo effects near the edges. The proposed metric effectively determines this optimal point for best contrast.

Extended Unsharpening Method (EUM)

The unsharp masking method is used to increase the contrast of images. Here, we extend this method and use binomial smoothing filter instead of average filter. The reason for using the binomial filter is that it preserves the spatial coherence between pixels and the contours better than the box filter which tends to produce contrast inversion in high frequency components or edges. But the problem with this approach is that it is more sensitive to noise. For our experiments, we fixed the mask size to be 3×3 and generated different contrast images for $\gamma = 1, 2 \cdots , 20$. For color images, the unsharp masking is applied on the luminance component in CIELAB color space and then converted back to RGB color space. Given an image I, its smoothed version I_{smooth} is computed by using a lowpass filter as follows:

$$I_{smooth}(x, y) = (I * h)(x, y) \tag{3.22}$$

Here h is the impulse response associated with the binomial filter. A mask for 3×3 binomial filter is as follows:

$$h = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{3.23}$$

$$I_{usm}(x, y) = I_{smooth}(x, y) + \gamma(I(x, y) - I_{smooth}(x, y)) \tag{3.24}$$

The unsharp masking is applied only to Luminance component of color images in CIELAB color space.

3.3.2 The Proposed Technique

The proposed measure is based on the 2nd order statistics computed from the gray-level co-occurrence matrix. In the preprocessing stage, we convert the color image from RGB to the perceptual color space CIELAB. Only the luminance component is then processed and analyzed. In order to capture the spatial correlation between neighboring pixels, the joint probability mass function is calculated from the co-occurrence matrix of the luminance component. It could be noticed that a co-occurrence matrix with dispersed values reveals the richness of the image details, while a concentrated co-occurrence matrix along the diagonal corresponds to low contrasted image. Therefore, the use of the co-occurrence matrix for evaluating the overall contrast of an image is relevant.

For a given color image, the co-occurrence matrix associated with the luminance channel is computed for distances $d = 1, 2$ and two orthogonal directions. The second order probability function is derived from the co-occurrence as defined in (3.17).

We compute the joint entropy from the image co-occurrence matrix using (3.25):

$$H = - \sum_i \sum_j p_{ij} \log_2 (p_{ij}) \tag{3.25}$$

where p_{ij} is the joint probability mass function of the luminance channel.

The mutual information within the rows and the columns of the co-occurrence matrix is then calculated from the joint probability mass function and marginal probabilities using 3.26, 3.27, and 3.28:

$$p_x (i) = \sum_{j=1}^{K} p_{ij} \tag{3.26}$$

$$p_y (i) = \sum_{i=1}^{K} p_{ij} \tag{3.27}$$

84

$$MICM = \sum_i \sum_j p_{ij} \log_2 \left(\frac{p_{ij}}{p_x(i)p_y(i)} \right) \qquad (3.28)$$

This measure is used as a powerful tool for quantifying the CE of color images. It could be also used to detect and control the side effects, such as saturation, overshooting or noise amplification that may result from CE.

3.3.3 Experimental Results

To evaluate the performance of the proposed metric, we performed our experiments on different color images. The images were of dimension 512×384. First of all, we generate contrast images using edge-based and unsharp masking-based CE methods by changing their parameters. The gray-level co-occurrence matrix was calculated for neighboring pixel distances, $d = 1, 2$, and directions $\theta = 0°, 90°$ from the luminance component in CIELAB color space. From the co-occurrence matrix, we computed the entropy, contrast, and mutual information. The performance of the proposed metric is evaluated for AEBCE and EUM CE methods already discussed in Section 3.3.1. Figure 3.5 shows the halo effect produced due to saturation of pixel values at the edges for different window sizes in the edge based contrast algorithm, for a zoomed portion of the corresponding images in Fig. 3.4. The effects of CE on the entropy, contrast, AMBE, EC, EME, AME, and the proposed measures (i.e., mutual information) for edge-based method is shown in Fig. 3.7. From the curves in Figure 3.7, we observe that the AMBE does not provide information about the best contrast as we notice an increasing exponential curve with increasing parameter. Similarly, the other metric, AME has decreasing exponential curve and is unable to provide information about the optimal CE. The other metrics i.e, entropy, contrast, EME, and EC increase with window size and they start to saturate after a certain point. However, we observe a sharp decay in the proposed measure curve after reaching a certain point (the

point after that saturation starts). Similarly for EUM-based CE, our proposed measure provides an optimal point after that increase in contrast results in overshooting effects and clearly visible in Fig. 3.6 and 3.8. We also tested our proposed metric on other images and we observed consistency across different test images. All of our experiments showed the superiority of the proposed metric in addition to providing the point of best contrast not provided by other CEE measures.

Figure 3.4: Effects of increasing window size in AEBCE technique (a) original image [b-d] enhanced image with different window size (b) 3×3 (c) 5×5 (d) 7×7

3.3.4 Discussions

Through this study, we demonstrated the inefficiency of the existing objective CEE measures in capturing the undesirable effects that may results from CE such as overshooting or noise amplification. The proposed measure based on some information concepts offers an efficient solution for analyzing and detecting such side effects for some neighborhood based CE methods. The other interesting result is that the proposed measure offers an efficient index to localize the size of
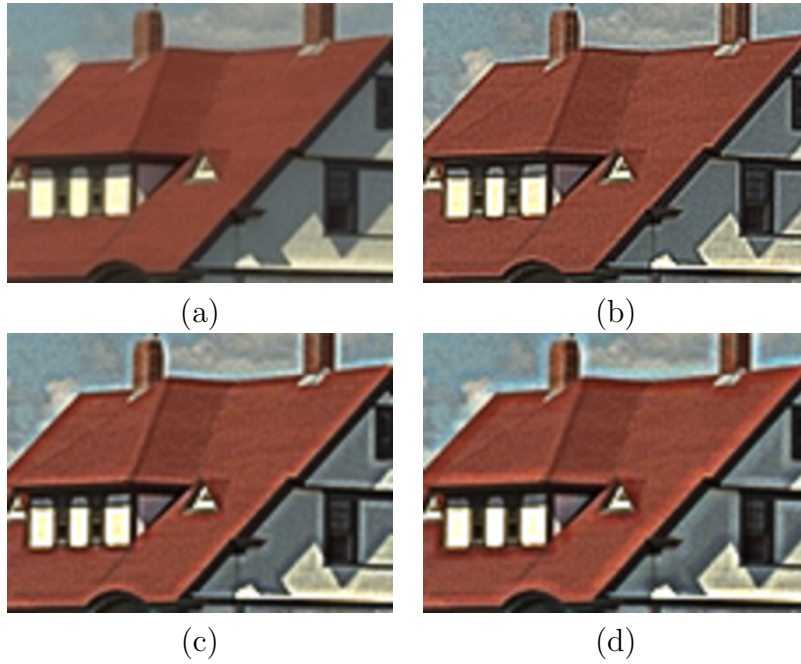
Figure 3.5: Effects of increasing window size in AEBCE technique (a) original image [b-d] zoomed regions with different window size (b) 3×3 (c) 5×5 (d) 7×7
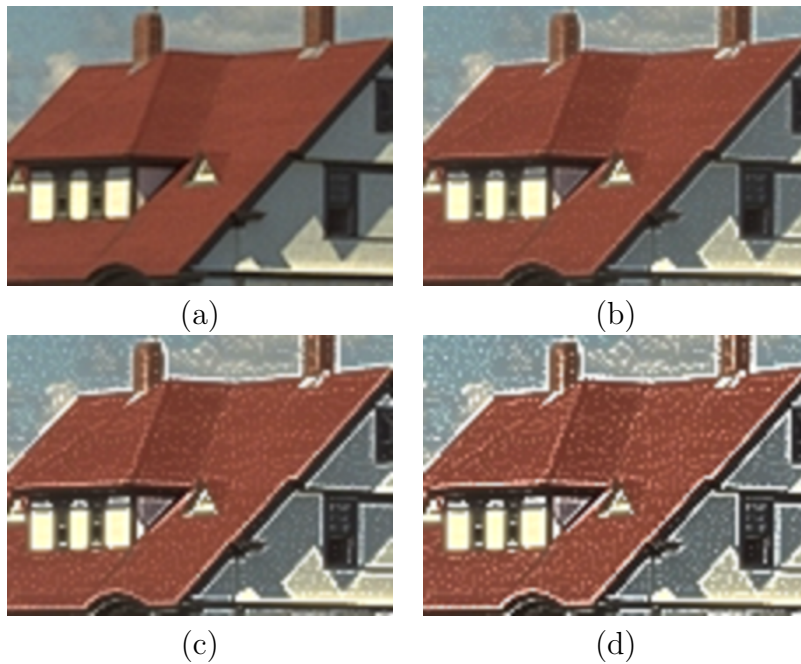
Figure 3.6: Effects of increasing window size in EUM-based CE technique (zoomed regions) (a) original image [b-d] enhanced images with different γ values (b) $\gamma = 3.5$ (c) $\gamma = 7.5$ (d) $\gamma = 10$
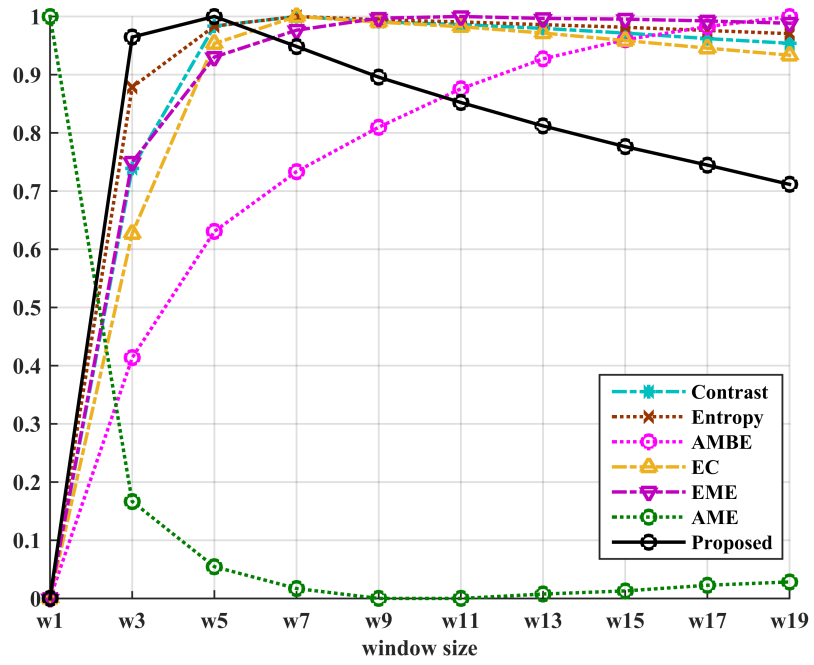
Figure 3.7: Comparison plots of different CEE measures for local AEBCE method
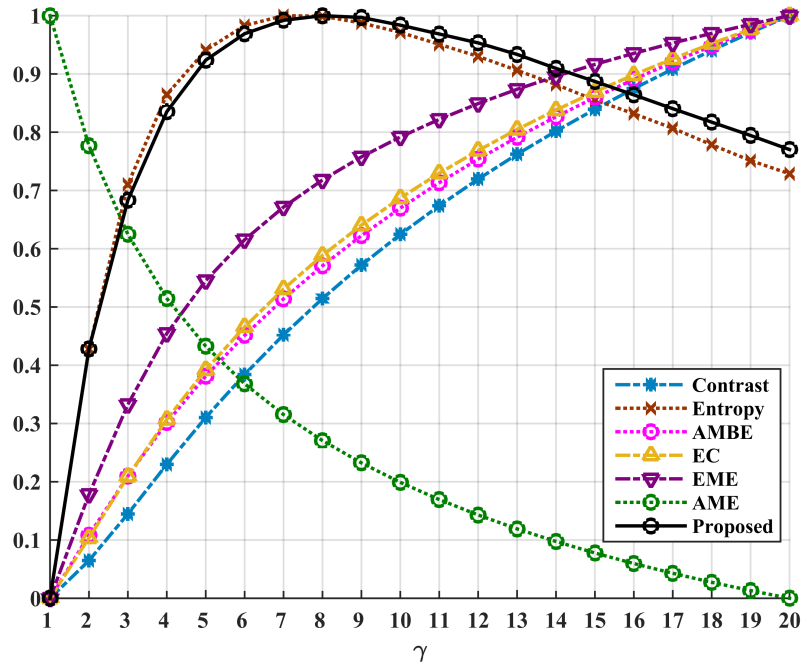
Figure 3.8: Comparison plots of different CEE measures for EUM-based CE method

the window at which the CE tends to decrease the quality of the image. The proposed metric is expected to open a new era for developing perceptually motivated algorithms for evaluating the image enhancement techniques as well as analyzing techniques for finding saturation point beyond which enhancement becomes unpleasant to the viewer. One idea to be explored in the future is the use of the proposed measure in order to develop adaptive local CE methods. The extension to video would be also an interesting perspective work.

In the following section, we provide a detailed description of our new database containing images obtained using six CE methods as well as a brief discussion on the testing methodologies, testing environment, and performance analysis of state-of-the-art CEE measures on the proposed database.

3.4 A New Database for Contrast Enhancement Evaluation

In this work, we develped and tested a new database dedicated to CE images for perfomance comparison of CEE measures following the main relevant ITU guidelines designed for the subjective experiments [28]. It is worth noting, that our methodology differs from previous works in many aspects; (1) The objectives are not the same. We aim here to analyze the performance of CEE measures in contrast to the work in [40, 101, 103] in which the performance analysis of classical IQA measures (i.e., IQA for distortions) was discussed, (2), The database is not the same compared with classical IQA databases like TID2013 [38], CSIQ [33] containing contrast images and few existing contrast databases [101, 103]. These databases contain simulated changes of global contrast using a simple pixel value mapping function so as to produce a decrease in contrast. The authors consider these transformations as contrast distortion. Whereas, in our framework, we deal

with the artifacts and distortion that may happen when applying CE operations. The distortion that might appear in the enhanced images after processing with CE methods are, for example, color saturation, color loss, blocking and ringing amplification in the case of compressed images, noise amplification, and halo effects and some others. The common databases did not contain any of these after effects due to CE. Moreover, in our case, we use different representative CE methods. (3), In contrast of all the databases, we do not want to estimate distortion in terms of decrease in quality like in classical IQA, rather our goal is to assess and quantify, subjectively and objectively, the increase in quality. (4), The application is entirely different compared to the CEE of tone mapped and retargeted images [106, 107, 105].

To the best of our knowledge, there are only two dedicated databases related to contrast manipulation [101, 103], where the processed images are obtained using simple artificial pixel-based transformations. Whereas, in our proposed database, some realistic CE artifacts are considered and provided with subjective ranking of different CE methods, which can be used to validate the performance of new CEE measures. The proposed database will help in preliminary validation of new image CEE measures without performing dedicated subjective experiments. The main contributions of this work are:

- To provide a comprehensive performance analysis of the state-of-the-art CEE measures in terms of correlation with the subjective evaluation provided in the developed database as well as on other existing contrast manipulated databases.

- To evaluate six representative CE methods on a set of images representing different kinds of visual content. Here, our objective is to analyze the performance of CEE measures rather than CE methods. We focused only on some representative CE methods.

- To provide a comprehensive statistical analysis of the data collected from subjective experiments on a new and unique CE dedicated database.

- To propose a multi-metric fusion to improve the correlation performance with the subjective ranking.

Here, we provide a brief discussion of the new database, selection and creation of images, testing environment, testing procedure, performance comparisons of different CEE measures on the new database.

3.4.1 Database Creation

We constructed a new database named as Contrast Enhancement Evaluation Database (CEED2016), containing 30 original color images and 180 enhanced images with a size of 512×512 pixels. The database is built with our own images and some common pictures used by the image processing community. The images in the database are shown in Fig. 3.10.

Selection of Images: It is well-understood that the human perception of image quality is highly dependent upon the scene content under observation. For this reason, we selected images with different textures, color distributions, and contrast variations. We have used three quantitative measures for the selection of images. These measures are Colorfulness (CF) [128], Spatial Information (SI) [128], and Global Contrast Factor (GCF) [129]. A brief description of each measure is given below:

Colorfulness (CF): It is a perceptual indicator of the variety and intensity of colors in the image [128]. The Red (R), Green (G), and Blue (B) color components are converted into opponent color space as follows:

$$r_g = R - G \tag{3.29}$$

Table 3.5: Summary of subjective experiments used with different image processing databases

	[105] (2005)	[106] (2010)	[107] (2012)	[101] (2012)	[130] (2013)	[108] (2014)	[39] (2014)	[40] (2015)	**Ours** (2016)
Original Images	23	37	57	26	15	30	23	8	30
Processed images	-	-	171	104	400	300	-	480	180
Image Resolution	-	-	-	512×512	720×576	-	800×800	1600×1200	512×512
Subjective Method[1]	PC	PC	ACR	-	SS	PC	-	ACR-DR	PC
Score type	Rank	Rank	MOS	DMOS	MOS	Rank	9 scales catg rating	-	Rank
Screen Resolution	-	-	1920×1280	1920×1200	-	-	1920×1080	LCD 24"	1920×1200
Observers	48	210	30	9	22	-	17	188	23
Viewing distance[2]	-	Web	-	-	$3H_I$	-	50cm, 100cm	80cm	$2H_S$
Processing Methods	6	8	10	-	-	-	-	-	6
Evaluation Measures	NIL	6	6	-	-	-	-	-	12
Application[3]	TM	RT	RT	CE	CE	DH	GM	GM	CE

[1] SS - Single Stimulus, ACR (Absolute Category Rating), DR (Dynamic Reference)
[2] H_I (Image height), H_S (Screen height), Web (through web based interface)
[3] TM (Tone Mapping), RT (Retargeting), CE (Contrast Enhancement), DH (Dehazing), GM (Gamut Mapping)

Table 3.6: Summary of subjective experiments used with different image enhancement databases

Database	TID2008	CSIQ	TID2013	CID2013	CCID2014	DRIQ	Ours
Year	2008	2009	2013	2013	2014	2012	2016
Original Images	25	30	25	15	15	26	30
Processed images	200	116	250	400	655	104	180
Image resolution	512 x 384	-	512 x 384	720 x 576	768 x 512	512 x 512	512 x 512
Contrast types	2	-	-	3	7	-	6
Subjective method	PC	Own	PC	SS	SS	-	PC
Score type	MOS	DMOS	MOS	MOS	MOS	DMOS	Rank
Score scale	0-9	0-1	0-9	0-5	0-5	-	-
Screen resolution	19"CRT + web	24" LCD	lab + web	-	-	1920 x 1200	1920 x 1200
Observers	838	25	-	22	22	9	23
Viewing distance	varying	70cm	varying	3HI	3HI	-	2 HS
Application							
Processing Operation	Distortion	Distortion	Distortion	Distortion	Distortion	Distortion	Enhancement

$$y_b = \frac{(R+G)}{2} - B \qquad (3.30)$$

The CF is then given by:

$$\mathrm{CF} = \sqrt{\sigma_{r_g}^2 + \sigma_{y_b}^2} + 0.3\sqrt{\mu_{r_g}^2 + \mu_{y_b}^2} \qquad (3.31)$$

where σ_i and μ_i for $i \in [r_g, y_b]$ represent the standard deviations and the mean of the pixel values in the opponent color space.

Spatial information (SI): It is an indicator of edge energy and is calculated as the root mean square of the edge magnitude over the entire image [41]:

$$\mathrm{SI} = \sqrt{\frac{L}{1080}}\sqrt{\sum_{k=1}^{N}\left(\frac{\Delta_k^2}{N}\right)} \qquad (3.32)$$

where Δ_k represents the gradient magnitude computed from the Sobel operator at the kth pixel, N is the total number of image pixels, and L is the vertical resolution of the image.

Global Contrast Factor (GCF): It is a global measure of the overall image contrast as perceived by the HVS. This contrast measure accounts for the multi-scale characteristics of the HVS. It is based on a multi-resolution decomposition scheme and a weighting process. The global contrast is then expressed as the weighted average of the local contrast computed at different resolution levels. The contrast weighting function is derived from a psychophysical experiment [129]. It is calculated as follows:

$$GCF = \sum_{i=k}^{N} w_k c_k \qquad (3.33)$$

where w_k and c_k represents weights and average local contrast of the image for a

given resolution and N is the number of resolution levels.

$$c_k = \frac{1}{W^k \times H^k} \sum_{i=1}^{W^k \times H^k} \frac{1}{l_{c_i^k}}, \text{ for } k = 1, \cdots, N \qquad (3.34)$$

where $l_{c_i^k}$ represents local contrast for i^{th} pixel at the kth resolution, W^k and H^k represents image width and height at k^{th} resolution. The local contrast is computed as the average of the differences of pixel values with its four nearest neighbors.

Since the database contains images enhanced by different CE methods, and to see the effect of improvement in quality in a clearer way, we have used this measure to select images with varying contrast from low to high.

Using the measures above, we provided a scatter plot for the images in our database (Fig. 3.9). Here, '$\star$' symbol is used, to represent images. From the plots, it could be observed that the database contains images with diverse spatial information, colorfulness, and global contrast features.

Creation of Enhanced Images: In the literature, we can find numerous CE methods. In our work, we selected six CE methods as a representative set of the most common approaches used in the literature. These methods are: AE-BCE [124], Contrast Limited Adaptive Histogram Equalization (CLAHE) [131], DCT [132], Global Histogram Equalization (GHE) [133], Top Hat Transformation based (TOPHAT) [134], and Multi-scale Retinex (MRETINEX) [135]. The above were selected to cover the different classes of CE including: histogram-based, edge-based, transform-based, morphological-based, and HVS-inspired methods [132, 134, 124, 133, 135, 131]. We have used the codes for some methods accessible from the original papers author's websites. For GHE and CLAHE, we have used the MATLAB built-in functions `histeq` and `adapthisteq` respectively. Since, the main goal of the study is the performance comparison of CEE mea-
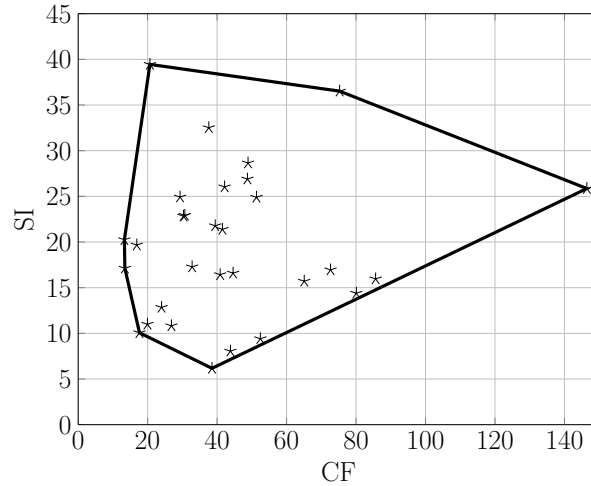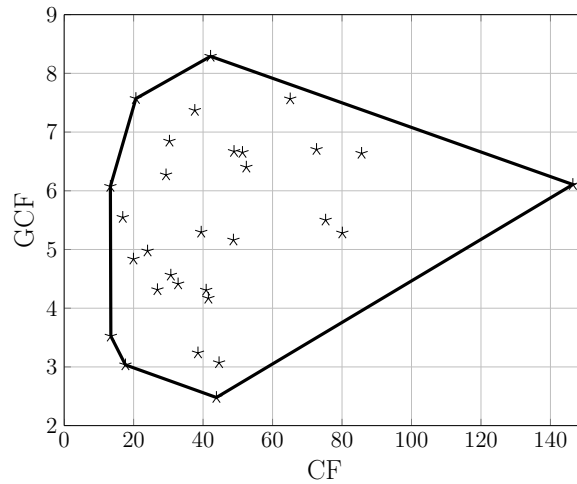
Figure 3.9: Scatter plots for all images in the new database between (a) Spatial information versus Colorfulness (b) Global contrast factor versus Colorfulness

sures instead of CE methods, therefore, for our experiments, we have used the CE algorithms with their default parameters without tuning the algorithms for performance optimization.

Among the original images, we have also included six compressed images (three for JPEG and three for JPEG2000) with moderate compression so as to observe the effect of contrast enhancement that may increase the visibility of these masked artifacts. In this way, we can also observe the capabilities of different CEE metrics in quantifying these particular CE after effects i.e., blur, ringing amplification. In

Fig. 3.11, we show some enhanced images with visible artifacts due to CE.

3.4.2 Testing Environment

The subjective experiments were performed at Université Paris 13 at Laboratoire de Traitement et Transport de l'Information (L2TI). The images were displayed on a calibrated LCD monitor in a dark room environment to avoid any problem with the illumination adaptation of background. The details of display parameters are shown in Table 3.7.

Table 3.7: Display setup used in the subjective experiments

Parameter	Description
Type	LCD
Model	EIZO Color Edge CG242W
Screen	24.1 inch
Resolution	1920×1200 pixels
Calibration device	Eye-One Match 3
Color space	sRGB
Color temperature	6500K
White point luminance	119 cd/m^2
Display frame rate	60Hz
Contrast	80
Room Environment	Dark
Gamma	2.2
Background color	Gray $(128, 128, 128)$

Twenty-three observers both experts and non-experts and coming from different age groups, gender, and background participated in the experiment. All the observers had either normal vision or corrected to normal vision and they were undergone a pre-screening procedure for color vision and visual acuity. The observers were forced to perform the experiments at a fixed distance of twice the screen height. The non-expert observers were not informed about the definition of contrast, and were asked to give their preference about the image they feel perceptually better than the other compared to the original image. They were

allowed to give the same rank for both images in case of the equivalent degree of quality.

The database contains 30 original images. Each original image is enhanced by six CE methods, resulting in 180 enhanced images in addition to the original images. For each original image, the six enhanced images were shown in pairs to the observers. The number of possible combinations to display for each original image are $\binom{6}{2} = \frac{(6)(5)}{2} = 15$. We allowed the observers to take their time for the subjective experiments and they were not forced to finish early. However, they were informed that the whole subjective tests take on average 30 minutes.

3.4.3 Testing Procedure

Considering the advantages of PC-based methods, we opted to use the non-forced choice balanced pairwise ranking protocol (Condorcet method) in our subjective experiments. The interface for the subjective experiments was developed in MATLAB, where, for each original image, we randomly displayed all possible pair combinations of enhanced images to the observers. We also showed the original image in the center of the screen (a pair of enhanced images are to its left and right), to facilitate the analysis of after effects of CE. The observers had the choice to rank equally similar stimuli. A screenshot of the graphical interface is shown in Fig. 3.12. In the PC ranking protocol, each enhanced image is compared with the others in pairs and ranking results are stored in a preference matrix. An aggregated preference matrix for the 23rd image in the database is shown in Table 3.8. From the Table 3.8, it can be observed that the CLAHE method is highly preferred, whereas GHE is least preferred by all the observers. The preference data was collected for all the images in our database for statistical and correlation analysis.

Table 3.8: A sample preference matrix for 23rd image (i.e., mosque) aggregated over preferences of 23 observers. In our experiment, $M_1 = \text{AEBCE}$, $M_2 = \text{CLAHE}$, $M_3 = \text{DCT}$, $M_4 = \text{GHE}$, $M_5 = \text{TOPHAT}$, $M_6 = \text{MRETINEX}$.

-	M_1	M_2	M_3	M_4	M_5	M_6	p_i
M_1	–	2.5	10	23	22	3	60.5
M_2	20.5	–	17.5	23	23	15.5	99.5
M_3	13	5.5	-	23	23	10	74.5
M_4	0	0	0	–	1	0	1
M_5	1	0	0	22	–	0	23
M_6	20	7.5	13	23	23	–	86.5

3.4.4 Statistical Analysis

The data gathered from the subjective experiment was processed to verify its reliability and validity. The reliability relates to the consistency and it is further related to the closeness of agreement in the preference ranking among different observers (also called inter-rater reliability). Whereas, validity relates to the accuracy of the data. However, it does not mean that the data with high reliability is also accurate. For the preference based pairwise rank data, the reliability was measured using Kendall's Coefficient of Concordance (W) [29] and Coefficient of Agreement. Other measures are Kendall's Tau (τ) and Spearman's Rank Order Correlation Coefficient (ρ).

(A) Coefficient of Agreement (u): The coefficient of agreement or inter-rater reliability is a measure of understanding among a group of observers in their judgments. It is measured on a continuous scale in the range $[0-1]$. Kendall and Babington et al. [136] proposed coefficient of agreement, u, among the observers and defined it as:

$$u = \frac{2 \sum\limits_{i,j=1,i\neq j}^{M} \binom{a_{ij}}{2}}{\binom{S}{2}\binom{M}{2}} - 1 \tag{3.35}$$

where M is the number of CE methods, S is the number of observers, and a_{ij} represents the number of times image enhanced by method M_i is preferred over the

image enhanced by method M_j. Its value is equal to one, when all the observers (or raters) agree on their preferences.

To test for the significance of coefficient of agreement (u), we have performed a chi-squared test (χ^2). The χ^2 values are calculated as follows:

$$\chi^2 = \frac{M(M-1)(1+u(S-1))}{2} \tag{3.36}$$

The degree of freedom for this χ^2 statistic is selected as $\frac{M(M-1)}{2}$. The minimum value of u is $\frac{-1}{(S-1)}$ and $\frac{-1}{S}$ for even and odd number of observers respectively. For our experiment, with 23 observers, the minimum value of the consistency coefficient $(u_{\min})$ is $\frac{-1}{23} = -0.0435$. The null hypothesis H_0 is rejected when the observed χ^2 is greater than its critical value.

(B) Coefficient of Consistency or Transitivity (ζ) The pairwise rank data is further assessed for inconsistency. It relates to the transitivity property in a paired comparison. It is determined from the number of intransitivity or circular triads in a set of ranking. It is also called intra-rater agreement and is calculated for each observer and image. The coefficient of consistency in a set of pairwise comparison can be calculated using the relation [136]:

$$\zeta = 1 - \frac{C}{C_{\max}} \tag{3.37}$$

where C represents the number of circular triads and $C_{\max}$ is the maximum possible circular triads in a pairwise comparison. C is calculated using the following relation:

$$C = \frac{M}{24}(M^2 - 1) - \frac{1}{2}M \tag{3.38}$$

where $M = \sum(a_i - (M-1)/2)^2$, a_i is the number of times stimulus i was preferred

over other stimuli. The maximum value of C is given by:

$$C_{\text{max}} = \begin{cases} \frac{(M^3 - 4M)}{24} & M \text{ is even} \\ \frac{(M^3 - M)}{24} & M \text{ is odd} \end{cases} \tag{3.39}$$

Note that, $\zeta = 1$ represents a perfect consistency in the pairwise comparisons.

The consistency coefficient for each observer is calculated by averaging consistency coefficients across all the images used in the experiment. Whereas, the consistency coefficient for each image is computed by averaging consistency coefficients for all observers participated in the experiments (see Table 3.9 and 3.10).

(C) Kendall's Coefficient of Concordance (W) is also used to measure the degree of agreement in the rankings among different observers. It is calculated as follows:

$$W = \frac{(12 \times S)}{(S^2(M^3 - M) - S \times T)} \tag{3.40}$$

where S and M are the number of observers and the number of methods respectively. T is the correction factor, when there are ties in the rank. T is zero when there is no tie within the rank. The correction factor T is calculated as follows:

$$T = \sum_{k=1}^{K} t_k^3 - t_k \tag{3.41}$$

where K is the total number of tie groups, and t_k is the total number of ties in a particular group.

To determine the significance of W, we calculated the χ^2 value given by:

$$\chi^2 = S(M - 1)W \tag{3.42}$$

Then, the probability of getting the results by chance (p-value) is also calculated using the χ^2 distribution. The p-values for the experimental data indicates that

(1) chalet	(2) mountain	(3) plane1	(4) aerial-view	(5) playground
(6) city river	(7) camel	(8) puppies	(9) elephant	(10) sculpture
(11) swan	(12) monkey	(13) deer	(14) tree	(15) flower1
(16) bush	(17) fruits	(18) flower2	(19) painting	(20) street
(21) building	(22) fort	(23) mosque	(24) plane2	(25) barbara
(26) mandrill	(27) clock	(28) zelda	(29) pepper	(30) clown
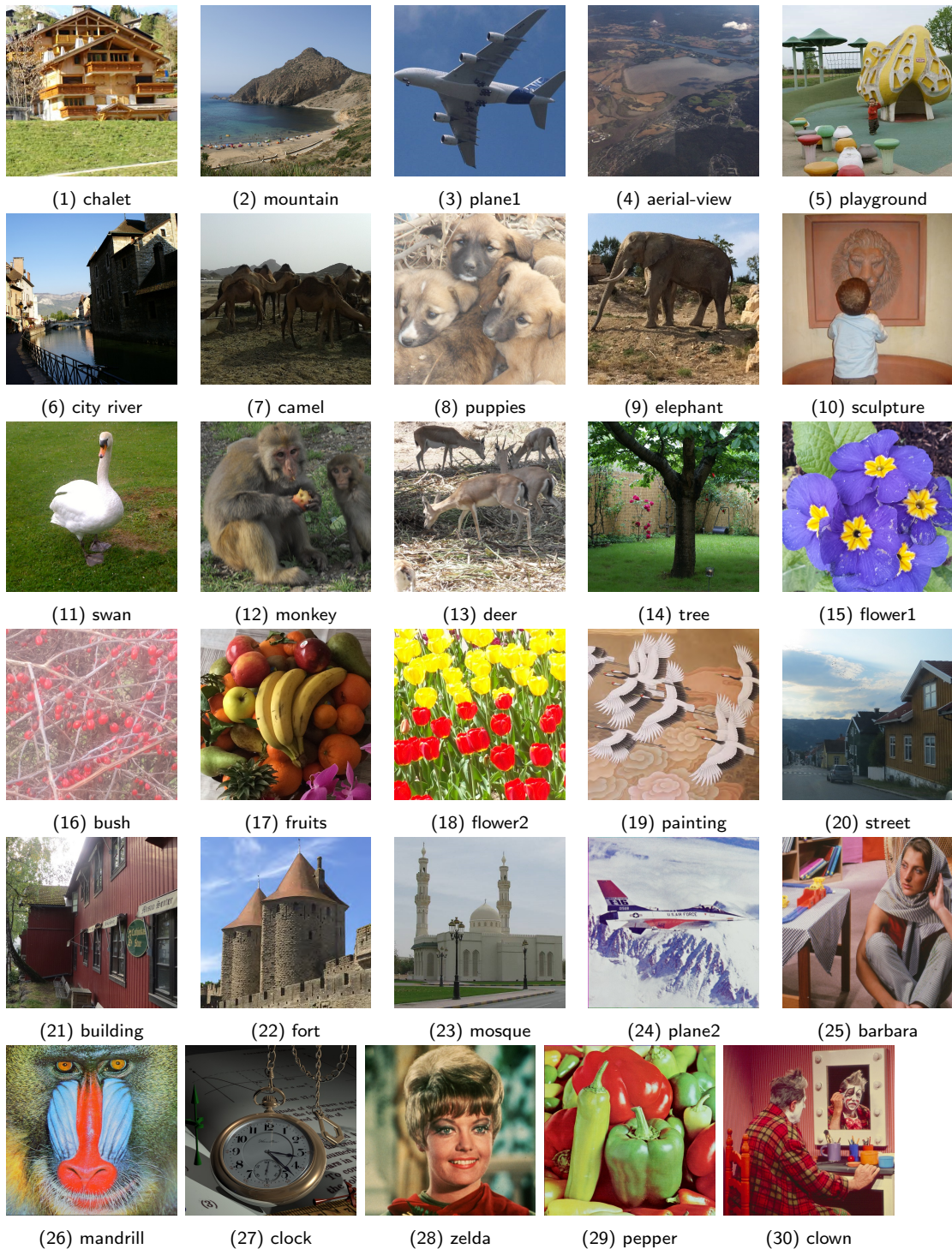
Figure 3.10: Images in the database (Images 1-23 are self captured while images 24-30 are standard test images

Figure 3.11: Illustration of some artifacts due to CE (a) color shift, (b) halo effects, (c) blocking, and (d) ringing
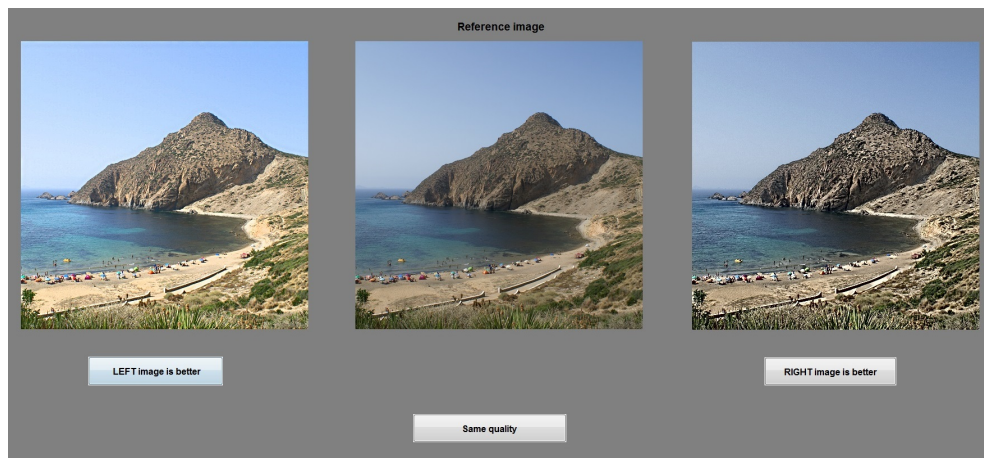
Figure 3.12: The display environment where the original and enhanced images are presented at the same time

Table 3.9: Summary of statistical coefficients for the subjective experiment data

Image	u	χ^2	p_{value}	$\bar{\zeta}_{\text{observer}}$	W	χ^2	p_{value}
1	0.67	235.83	< 0.05	0.94	0.82	94.44	< 0.05
2	0.38	141.30	< 0.05	0.84	0.55	63.47	< 0.05
3	0.55	196.91	< 0.05	0.86	0.71	81.48	< 0.05
4	0.47	170.70	< 0.05	0.88	0.64	73.06	< 0.05
5	0.57	202.91	< 0.05	0.91	0.77	88.63	< 0.05
6	0.12	54.00	< 0.05	0.76	0.24	27.69	< 0.05
7	0.45	164.74	< 0.05	0.87	0.67	77.35	< 0.05
8	0.20	82.35	< 0.05	0.72	0.38	43.39	< 0.05
9	0.46	167.13	< 0.05	0.86	0.67	76.68	< 0.05
10	0.58	205.35	< 0.05	0.87	0.79	91.18	< 0.05
11	0.52	188.09	< 0.05	0.84	0.76	87.10	< 0.05
12	0.19	76.78	< 0.05	0.66	0.37	42.55	< 0.05
13	0.47	169.74	< 0.05	0.85	0.66	76.12	< 0.05
14	0.35	131.26	< 0.05	0.76	0.56	64.06	< 0.05
15	0.44	159.30	< 0.05	0.87	0.62	71.71	< 0.05
16	0.24	92.70	< 0.05	0.77	0.41	47.13	< 0.05
17	0.29	109.43	< 0.05	0.77	0.49	56.65	< 0.05
18	0.39	145.30	< 0.05	0.81	0.58	67.02	< 0.05
19	0.36	133.57	< 0.05	0.86	0.53	60.91	< 0.05
20	0.35	131.04	< 0.05	0.83	0.52	60.07	< 0.05
21	0.41	150.96	< 0.05	0.86	0.59	68.07	< 0.05
22	0.37	135.57	< 0.05	0.83	0.57	65.12	< 0.05
23	0.48	173.52	< 0.05	0.86	0.71	81.18	< 0.05
24	0.50	181.09	< 0.05	0.85	0.72	83.32	< 0.05
25	0.37	135.87	< 0.05	0.78	0.58	66.47	< 0.05
26	0.26	101.57	< 0.05	0.74	0.42	48.76	< 0.05
27	0.37	135.52	< 0.05	0.75	0.56	64.65	< 0.05
28	0.39	145.17	< 0.05	0.78	0.64	73.67	< 0.05
29	0.40	147.91	< 0.05	0.81	0.60	68.73	< 0.05
30	0.31	117.83	< 0.05	0.80	0.50	57.90	< 0.05

Table 3.10: Consistency coefficients for 23 observers

Observer	$\bar{\zeta}_{\text{image}}$	Observer	$\bar{\zeta}_{\text{image}}$	Observer	$\bar{\zeta}_{\text{image}}$
1	0.7583	9	0.9208	17	0.8406
2	0.7990	10	0.8354	18	0.8271
3	0.9042	11	0.7896	19	0.8094
4	0.6885	12	0.7708	20	0.8688
5	0.8688	13	0.8906	21	0.7917
6	0.7323	14	0.8354	22	0.7167
7	0.9208	15	0.8604	23	0.8344
8	0.8187	16	0.7406	−	−

the consistency coefficients are significant. However, we removed some images in the comparisons where these coefficients values are low.

For our preference based pairwise rank data collected from the subjective experiment, the values of these coefficients are presented in Table 3.9. From the significant tests, we have noticed that inter-observers' and intra-observers' consistency coefficients for the images in our new database are high except for the images numbered 6, 8, 12, 16, and 26. We then discarded these images and their related data and did not use these in further experiments.

3.4.5 Correlation Analysis

From the subjective experiments, we have derived the preference scores, i.e., the number of times an image enhanced by a particular CE method is preferred over other enhanced images. The subjective preference ranking of the six CE methods is shown in the first row in Table 3.11. We observe that CLAHE is highly preferred whereas GHE and TOPHAT are not preferred most of the time. In relation to CLAHE, we were expecting the MRETINEX to give better results, but surprisingly it was not the case for the database in our subjective experiments. The CEE measures are also calculated for the enhanced images created from the six CE techniques. For NR-CEE measures, we have computed the change rate

(a) EME (b) AME (c) EMEE

(d) AMEE (e) SDME (f) RME

(g) RMSC (h) EC (i) RSE

(j) AMBE (k) IEM (l) VIF

(m) RIQMC (n) DE (o) MICM

(p) LOE
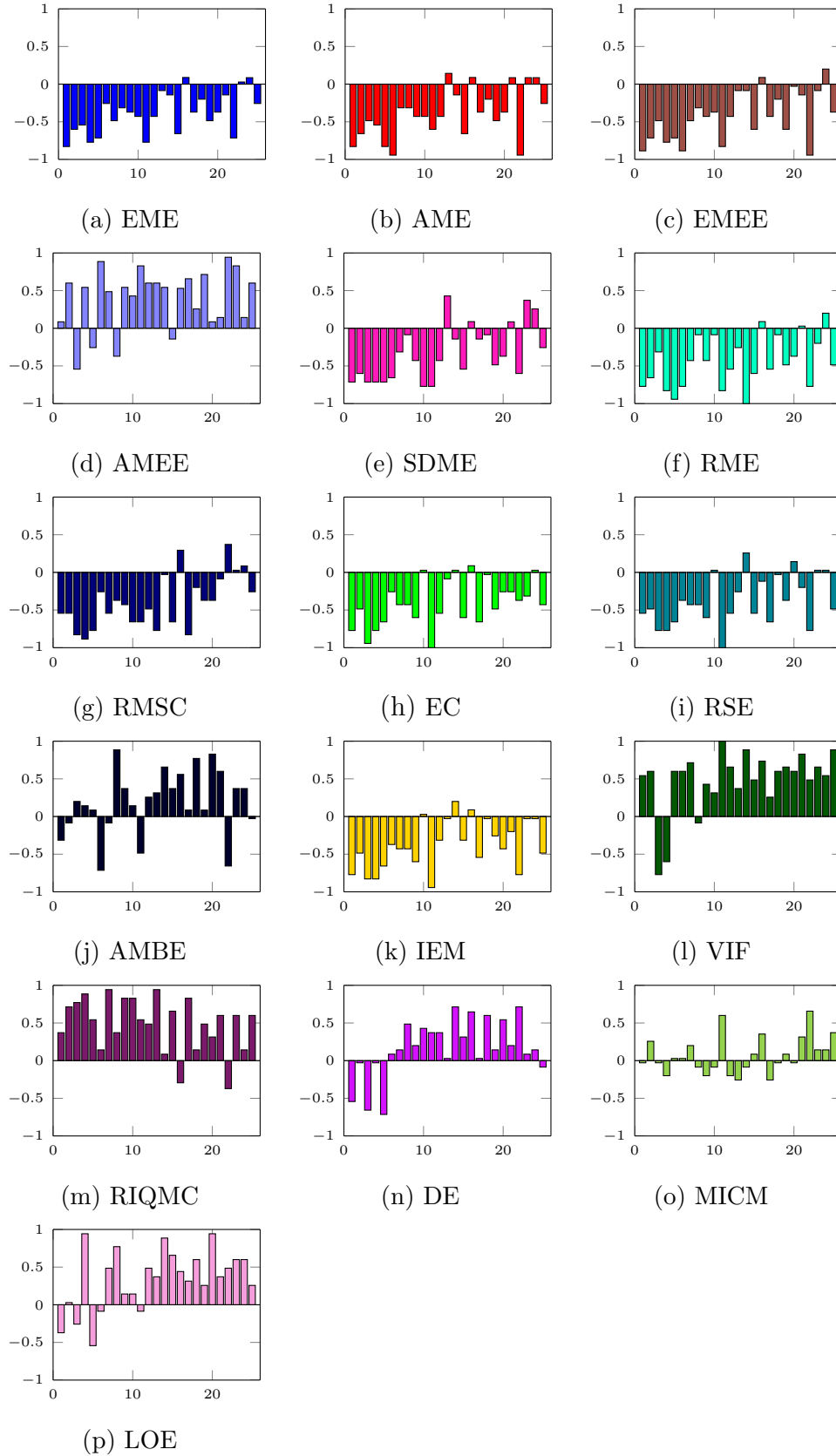
Figure 3.13: SROCC plots for 25 images in the database. The x- and y-axes represent image index and correlation values respectively

106

and then averaged over all the images in the database. The average scores for the CEE measures under study are also shown in Table 3.11. It is worth noting that most of the measures resulted in high/low values for GHE and TOPHAT in contradiction with the subjective preference ranking, and consequently resulted in the reduction of correlation results.

Based on the subjective experimental data, we have used the SROCC and the KROCC to observe the consistency of the CEE measures with the human visual perception [29]. The SROCC and KROCC are widely used non-parametric measures to determine the monotonicity between the ranks of two variables and their values ranges from -1 to $+1$. The values are close to $+1$ in case of strong correlation between the ranks of two variables and -1 in the case of strong disagreement between the two variables. The SROCC and KROCC give zero values when there is no correlation between the ranks. In our study, the aim is to observe how well a CEE measure is consistent in capturing the ranking for the six enhanced versions of each original image in the database. Therefore, before performing the correlations, we must consider, how the change in magnitude of metric values affects the image quality. For some metrics, high values correspond to good quality, whereas for other metrics, the opposite is true (see Section 3.2.2). For preference ranking, the highest score is highly ranked. Whereas, the metrics with high/low values corresponding to good quality are also highly ranked.

We calculate the median and mean correlations for each of CEE measures under study. For each image in the database, we have the ranking scores for its six enhanced versions as well as the quantitative scores. If I_i represents an original image and $I_{i,j}$, its enhanced version processed by method M_j, for $i = 1, 2, \cdots, n_I$ and $j = 1, 2, \cdots, n_J$, for $(n_I = 25, n_J = 6)$. Here n_I and n_J represent the number of original images and the number of CE methods respectively. We compute the

Table 3.11: Results of CEE measures and subjective evaluation for six CE methods

	AEBCE[124]	CLAHE[131]	DCT[132]	GHE[133]	TOPHAT[134]	MRETINEX[135]
Subjective Ranking	2060.5	2115.5	1211.5	686	662	1889.5
EME[97]	0.4910	0.1951	-0.1418	1.1633	1.3178	0.3867
AME[99]	-0.1944	-0.0776	0.0735	-0.2924	-0.3746	-0.0878
EC[114]	0.4301	0.1680	0.2773	0.6705	1.0645	0.5176
AMBE[109]	0.8370	3.8476	47.5236	24.2366	1.2839	22.3352
SDME[100]	-0.1217	-0.0595	0.0150	-0.1424	-0.2434	-0.0918
IEM[113]	1.7198	1.1814	1.3217	1.7572	2.2788	1.5467
VIF[102]	0.7914	1.0468	0.7272	0.8806	0.7856	0.9888
AMEE[99]	-0.0102	0.0188	-0.0079	-0.0300	-0.1516	0.0023
EMEE[98]	1.8851	0.5447	-0.2311	13.1548	24.5326	3.2767
RME[98]	0.1988	0.0353	0.0461	0.2831	0.3513	0.0826
RSE[65]	0.0610	0.0162	0.0338	0.0587	0.0907	0.0491
RMSC[110]	0.0619	0.0481	-0.0118	0.5888	0.2023	0.1651
MICM[25]	-0.2461	-0.0731	-0.1965	0.0133	-0.3157	-0.1484
LOE[112]	35.0732	120.6825	258.8569	134.5581	116.0922	135.6356
DE[111]	0.0143	0.0143	-0.0290	-0.1030	0.0311	0.0366
RIQMC[103]	2.3875	2.1724	2.0288	10.2342	6.8936	6.6620

SROCC for each image using the following relation:

$$\rho_{i,k} = 1 - \frac{6d_{i,j}^2}{n_J(n_J^2 - 1)}, \text{ for } i = 1, 2, \cdots, n_I \tag{3.43}$$

where $d_{i,j}$ represents the difference in the ranks of subjective preferences and objective scores of kth CEE measure for the ith image. The correlation for each image for all the CEE measures are shown in Fig. 3.13. Finally, the median and mean of the SROCC's for the 25 images are computed as a single performance measure of each CEE measure and reported in Table 3.12. Similarly, two different types of KROCC across the images, i.e., τ_{median} and τ_{mean} are also calculated and shown in Table 3.12. The median correlation provides more information compared to the mean correlation, as the median statistic is not affected by the outliers.

The performance of different CEE metrics is also compared with other existing databases with contrast manipulated images and results are shown in Table 3.13.

3.4.6 Results and Discussions

From the correlation results on different databases, we observed that the performance of various CEE metrics might differ on some databases. The metrics which perform better for some databases do not work well on others. The reasons are that the metrics are not adapted to different CE distortions, and the databases do not contain enhanced images affected by various CE distortions.

The comparison of the median correlations between subjective and objective data for the CEE measures is also shown in Fig. 3.14. From Fig. 3.14, we observe that only VIF, RIQMC, AMEE, LOE, AMBE, DE, and MICM metrics have positive correlations with the subjective ranking. The negative correlation of other CEE measures shows the inconsistencies of these measures with the human perception of quality judgment. This is due to the inconsistencies exist between

Table 3.12: Median and mean correlation analysis of CEE measures on the proposed CEED2016 database

Measure	SROCC (ρ)		KROCC (τ)	
	ρ_{median}	ρ_{mean}	τ_{median}	τ_{mean}
EME[97]	-0.3714	-0.3896	-0.3333	-0.2905
AME[99]	-0.4286	-0.3896	-0.2000	-0.2958
EC[114]	-0.4286	-0.4079	-0.3333	-0.3118
AMBE[109]	0.2000	0.1892	0.2000	0.1747
SDME[100]	-0.4286	-0.3325	-0.3333	-0.2425
IEM[113]	-0.4286	-0.3782	-0.3333	-0.2958
VIF[102]	0.6000	0.4797	0.4667	0.3884
AMEE[99]	0.5429	0.3892	0.3333	0.3241
EMEE[98]	-0.4286	-0.4239	-0.3333	-0.3225
RME[98]	-0.4857	-0.4468	-0.3333	-0.3598
RSE[65]	-0.4286	-0.3819	-0.3333	-0.2802
RMSC[110]	-0.4286	-0.3905	-0.3333	-0.3167
MICM[25]	0.0286	0.0713	0.0667	0.0726
LOE[112]	0.3714	0.3377	0.3333	0.2810
DE[111]	0.1429	0.1676	0.0667	0.1161
RIQMC[103]	0.5429	0.4865	0.4667	0.4021

Table 3.13: Median rank correlation values of different CEE metrics calculated on existing contrast-changed databases.

Metrics	CCID2014[130]		DRIQ[101]		TID2013[38]		CSIQ[33]	
	SROCC	KROCC	SROCC	KROCC	SROCC	KROCC	SROCC	KROCC
EME[97]	0.3543	0.2487	0.3221	0.2234	0.4986	0.3714	0.8360	0.6546
AME[99]	-0.0414	-0.0106	0.5367	0.3660	0.4530	0.3430	0.7376	0.5376
EC[114]	0.8286	0.6328	0.7457	0.5504	0.8060	0.6252	0.9532	0.8150
AMBE[109]	0.5965	0.4359	0.0129	0.0010	0.1715	0.1550	0.4963	0.3559
SDME[100]	0.0096	0.0335	0.5860	0.4059	0.4263	0.3221	0.8460	0.6549
IEM[113]	0.8352	0.6406	0.7427	0.5465	0.8065	0.6218	0.9519	0.8138
VIF[102]	0.8349	0.6419	-0.6679	-0.4619	0.7716	0.5795	0.9345	0.7769
AMEE[99]	0.6285	0.4489	-0.4353	-0.2927	0.0734	0.0395	0.4403	0.3040
EMEE[98]	0.1021	0.0807	0.3802	0.2734	0.4908	0.3581	0.8208	0.6129
RME[98]	-0.1625	0.0879	0.5274	0.3706	0.4660	0.3557	0.3290	0.2459
RSE[65]	0.7391	0.5396	0.7154	0.5098	0.7620	0.5678	0.9507	0.8078
RMSC[110]	0.7348	0.5336	0.3449	0.2268	0.7855	0.5922	0.9516	0.8126
MICM[25]	0.6917	0.5016	-0.6305	-0.4252	0.5733	0.4430	0.8005	0.6222
LOE[112]	0.6770	0.4844	-0.7716	-0.5564	0.1440	0.1157	0.7130	0.5316
DE[111]	0.8592	0.6690	0.2830	0.1975	0.7356	0.5598	0.9501	0.8120
RIQMC[111]	-0.8464	-0.6507	-0.3336	-0.2328	-0.8044	-0.6178	-0.9580	-0.8279

the metric values and subjective preferences for the TOPHAT and GHE methods, where these two approaches were ranked worst based on the observers' preferences. We can also draw a conclusion from the correlation analysis, that most of the CEE measures are not well suited for the CE evaluation for GHE and TOPHAT based CE methods. From these results, it is also clear that using simple local features, such as contrast or gradient, in the design of CE measure is not sufficient. It is important to include the color aspects in the design of CEE measures for color images. Through this study, it is shown that CE evaluation is still a very challenging problem. We believe that the introduction of some learning based approaches would offer better solutions to this very challenging problem.

3.4.7 A Multi-Metric Fusion based CEE Measure

It is also evident that a single metric cannot perform very well. This is due to the reason, that no metric is sensitive to different types of artifacts introduced due to CE process. Therefore, we propose to combine some best metrics to benefit from their strengths in quantifying the image contrast. We use a simple weighting based fusion and tune the weights to avoid the limitations of each metric and increase the correlation performance. We use the top four metrics with positive correlations, (i.e., VIF, RIQMC, LOE, and AMBE), and fuse their possible combinations using different weights. We show in Table 3.14, only the top three combinations with high correlations. Compared with the single metrics, the multi-metric fusion results in a substantial increase in correlation performance at the cost of an increase in complexity. From the fusion weights, we observe that LOE metric which captures the naturalness property of an image is considered more important by giving more weights in the fusion process. These observations provide us hints in designing new metrics to consider different quality parameters (e.g., naturalness, lightness, saturation, color-shift, visibility of edges, etc.) in CE applications.

112

Table 3.14: Median correlation results for combining different CEE metrics

Fused Metrics	SROCC	KROCC	Weights
(Q_1) VIF+RIQMC+AMEE+LOE	+0.8286	+0.6445	[0.1,0.1,0.1,0.7]
(Q_2) VIF+RIQMC+LOE	+0.7945	+0.6445	[0.1,0.1,0.8]
(Q_3) VIF+LOE	+0.7714	+0.6000	[0.1,0.9]

3.5 Summary

In this work, a comprehensive psychophysical-based performance comparison of different state-of-the-art CEE measures is presented. The analysis was carried over a new database, that we introduced, which consists of enhanced images using different CE methods most commonly found in practice. Extensive subjective experiments were performed using a balanced pairwise preference-based ranking protocol to rank the CE methods by perceived quality. The correlation between subjective preferences and objective measures showed that most of the existing CEE measures are not well adapted with human perception of enhancement quality. Our analysis revealed that only seven measures, namely VIF, RIQMC, AMEE, LOE, AMBE, DE, and MICM exhibit positive correlations with perceptual quality of contrast enhancement. This is due to the reason that a single metric may be unable to capture various CE artifacts. We demonstrated that multi-metric fusion resulted in substantial improvement in correlation performance. It provides us an insight to consider various CE distortions in designing a new CEE metric. The new database is expected to provide a platform for developing new CEE measures and benchmarking the results without the need for dedicated subjective experiments. Soon, the developed database along with the subjective experimental data will all be made publicly available to the research community.
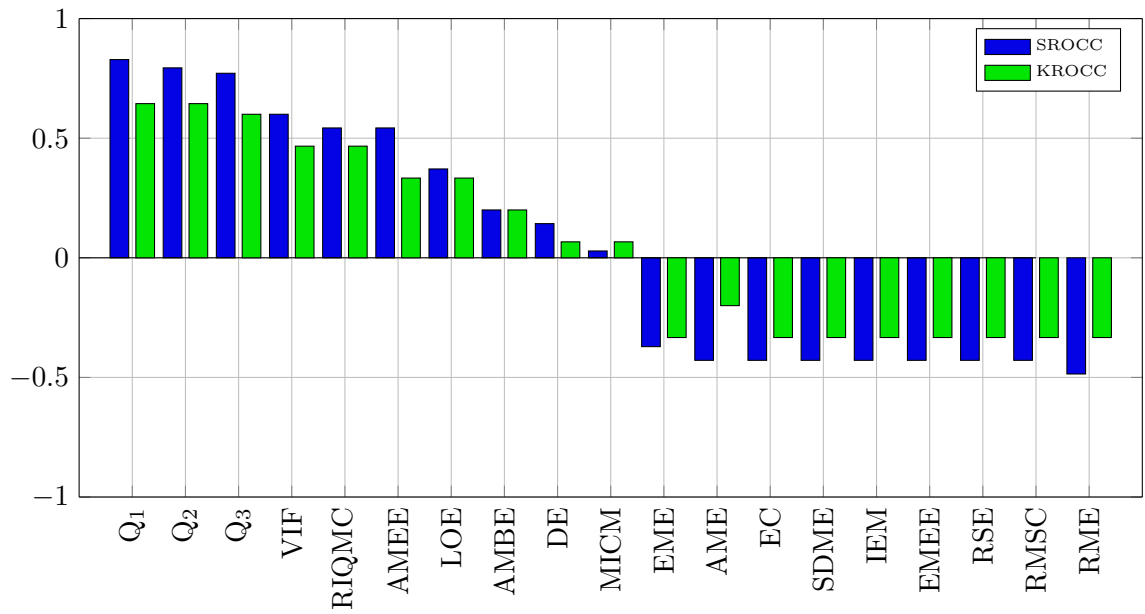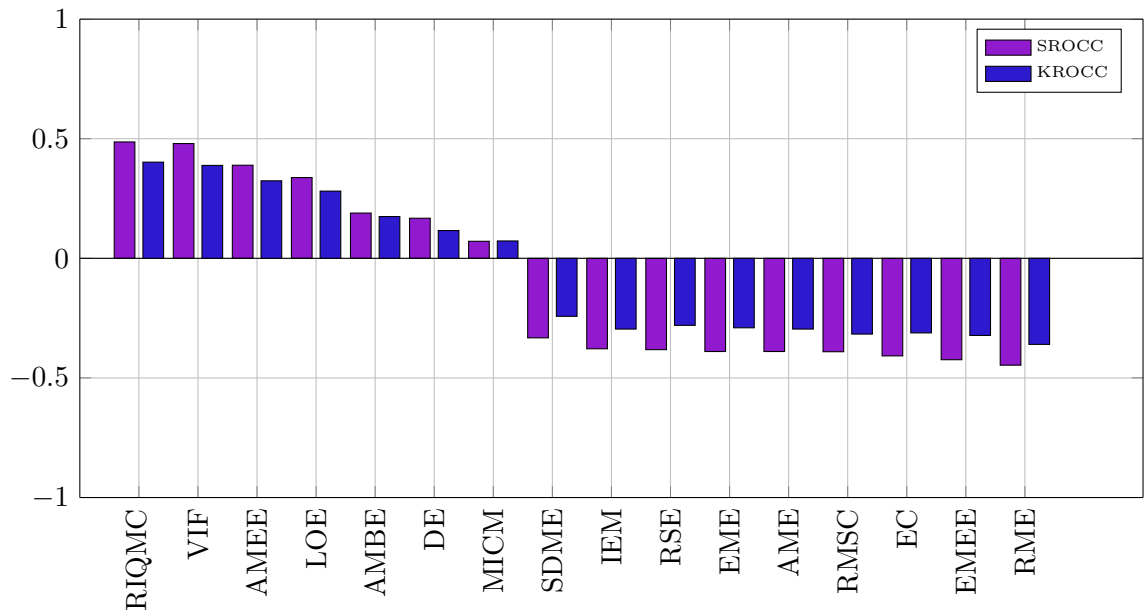
Figure 3.14: Subjective vs objective comparisons of CEE in terms of median correlations, (a) Median correlations, (b) Mean correlations

CHAPTER 4

IQA FOR INPAINTING EVALUATION

4.1 Introduction

Image inpainting is generally defined as the process of restoring missing pixels and damaged regions, or removing unwanted objects in digital images in a plausible way [4]. Considerable research has been carried in developing inpainting algorithms, and a plethora of image inpainting algorithms have been proposed [137, 4, 138, 139].Image inpainting has recently received considerable attention in different areas related to image processing. While the applications of image inpainting are countless, we outline below the most common and practical ones.

- Removing Unwanted Objects: Unwanted objects can be removed from the image using inpainting techniques. The scenario is seen as a special class of image tampering. Fig. 4.1 shows a nice example of image inpainting where the cage in the original image is removed in the inpainted image [3].

- Restoring Photos: The deterioration in photos with the passage of time can be overcome using inpainting. The scratches in the photos resulting from

improper handling can also be removed. This is also the case of restoring images from cultural archives, etc. Fig. 4.2 shows an example in which the scratches in the old photograph have been removed using inpainting [4].

- Photo Retouching: Another widely used application of image inpainting is in the media industry where photos of actors/actresses, models, etc., are manipulated by removing wrinkles, mole marks, or undesirable facial features to make these "more attractive". Fig. 4.3 shows an example of image inpainting where the face is made more attractive by removing some marks using inpainting [5].

- Text Removal: Image inpainting can also be used for removing unwanted text, stamps, copyright logos, etc., in digital images. Fig. 4.4 shows an example of a street image with superimposed text, from which the text is removed in the inpainted image.

(a) (b) (c)
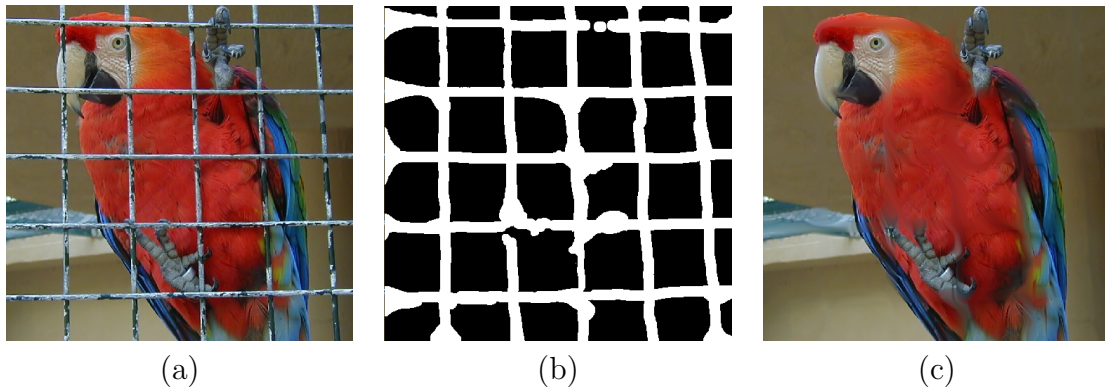
Figure 4.1: An example of image inpainting for object removal (a) original image. (b) binary mask, (c) inpainted image [3].

In a way, image inpainting can be seen as a modified copy-move tampering process which is used to recover or remove some parts of the image without any perceptual loss [11, 140]. It is different from copy-move forgery in a sense that different blocks or regions come from different locations of the image (see Fig. 4.5).

Figure 4.2: An example of image inpainting used in restoration, original image (left), restored image (right) [4].

Figure 4.3: An example of inpainting for photo retouching, original image (left), restouched image (right) [5].

Although a substantial amount of research has been carried out in developing robust inpainting algorithms, very little efforts have been put in developing quality assessment metrics to evaluate the performance of image inpainting (restoration) methods. IIQA is a complex and a challenging problem. The objectives of image inpainting assessment are quite different from those of classical image quality evaluation. Here, for inpainting IQA, the goal is to evaluate the quality of the restored images using either subjective or objective methods. However, traditional IQA fidelity-based metrics have mainly been developed for quantifying distortions in degraded images hence, these are not well suitable for evaluating the quality of restored images and cannot directly be used. This is due to the fact, that the
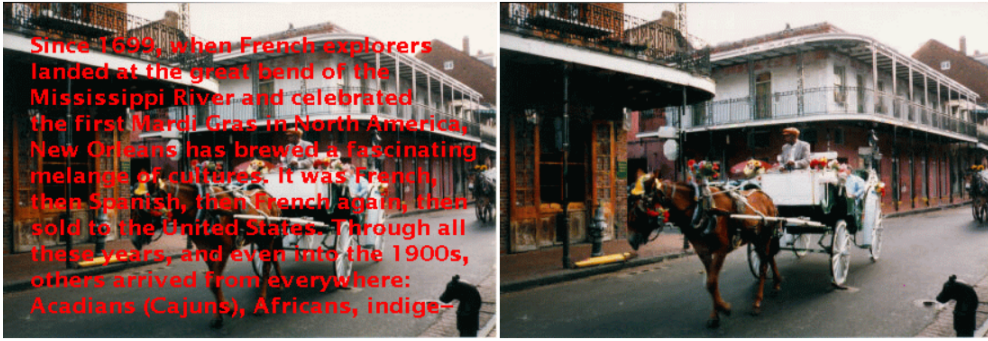
Figure 4.4: An example of image inpainting for text removal, original image (left), restored image (right) [4].
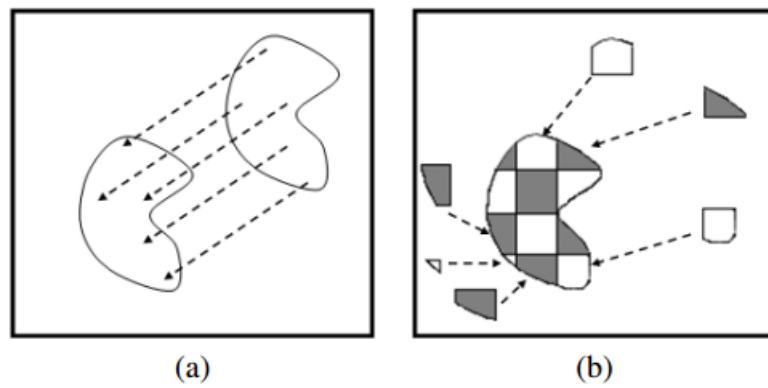
Figure 4.5: Difference between two types of tampering (a) copymove forgery (b) inpainting [6]

restored image in inpainting is different from its original counterpart. In image inpainting, different artifacts are introduced which in turns affect the perceived quality. Among these artifacts, blur is introduced around edges when restoring large inpainted regions. The curved boundaries are not produced correctly as well. This is our focus here, in this chapter, we provide a comprehensive and a critical review of different methods developed for quality assessment of inpainted images. This review will be the first of its kind and is expected to help researchers working in this area to benchmark new inpainting techniques, develop more robust methods for inpainting quality assessment, and benchmark their results.

The rest of the chapter is structured as follows: Section 4.2 provides a discussion of common inpainting algorithms. Section 4.3 briefly discusses the state-of-

the-art IIQA metrics. Different inpainting databases are discussed in Section 4.4. Finally, the chapter is concluded in Section 4.5.

4.2   Image Inpainting Methods

The main objective of inpainting algorithms is to restore the unknown regions to create a more pleasing and realistic feeling about the new image. Among different types of inpainting artifacts, most commonly observed ones are blurring, disconnected edges, inconsistent pieces of texture, etc. Based on our analysis of the state-of-the-art, we propose here to group inpainting algorithms into four broad categories: Exemplar-based, Partial Differential Equation (PDE)-based, Sparsity-based, and Hybrid (combination of Exemplar-, PDE-, or Sparsity-based) approaches. We display in Fig. 4.6 a tree diagram of different classes of most commonly used inpainting algorithms. Since, the chapter aims to provide a critical review of IIQA metrics instead of the inpainting algorithms themselves, we will only provide a brief discussion of each of these categories. Note that in image inpainting, the basic assumption is that pixels in the known and unknown regions should have similar geometrical structures and statistical properties.

In PDE-based or diffusion-based inpainting methods, the local structure information is transferred or diffused from the known region to the unknown (target) region [4]. Several variations of PDE-based methods were introduced based on the flow of texture information in linear, nonlinear, isotropic, or anisotropic directions. The PDE-based methods are well-adopted for restoring long narrow regions (cracks, lines). However, they are not recommended for restoring large unknown texture regions, due to the introduction of blur in the textured regions.

In Exemplar-based inpainting techniques, the structure completion process is carried out using texture synthesis i.e., the target regions are restored by selecting
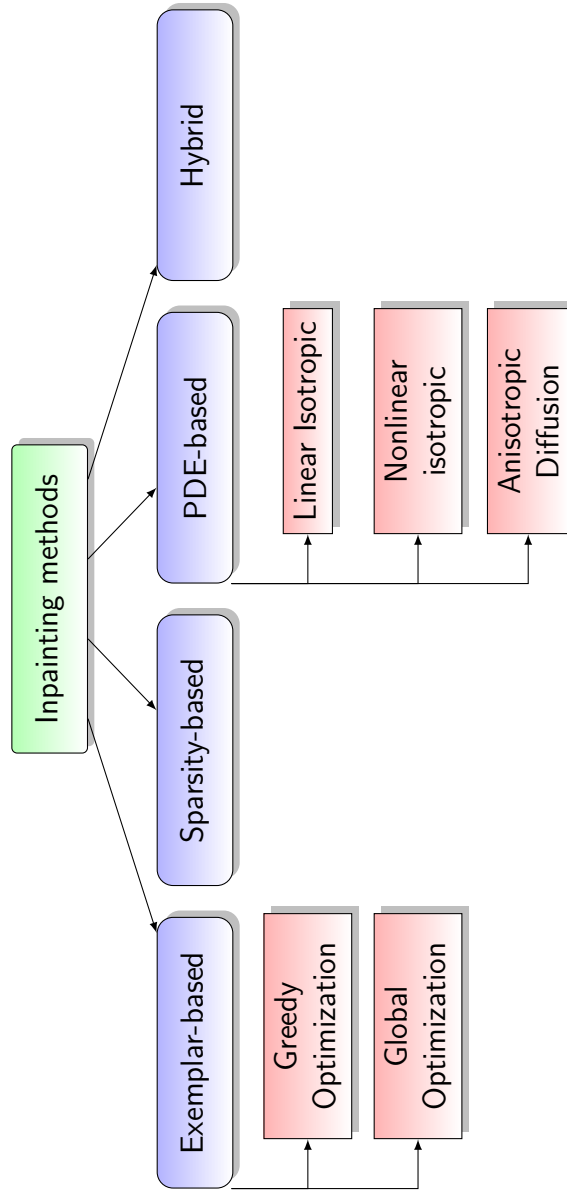
Figure 4.6: Tree diagram of different classes of inpainting techniques

the patches in the known regions similar (in terms of structure) to the partially unknown patches in the target regions [11]. These techniques use greedy and global optimization functions in filling the target regions with the similar known pixels.

Sparse representations of images, over a particular basis (e.g., Discrete Cosine Transform (DCT), Wavelet, etc.) has also been used in image inpainting. In Sparsity-based inpainting methods, the assumption made is that the known and unknown image regions share the same sparse representations.

The Exemplar-based and the Sparsity-based inpainting methods, above, were shown to perform better compared to the PDE-based methods for filling large texture regions. Various Hybrid techniques also exist that combine the strengths and different types of inpainting methods for performance improvement. To demonstrate the effect of different inpainting algorithms, Fig. 4.7 shows a very nice example of inpainting where broken pieces of the kiwi fruit are restored using different inpainting algorithms [7]. The broken area is shown as a green mask. It is clear that Figs. 4.7 (d) and (f) represent more realistic inpainting output compared to the other methods. After this brief survey on commonly used inpainting algorithms, we now move to the focus of the chapter and discuss in more details different types of subjective and objective IIQA metrics, commonly used in the literature.

4.3 Image Inpainting Quality Assessment (IIQA) Measures

Image inpainting methods were initially used for removing missing or damaged areas in an image. The main criterion was that the restored image should be "close" to the original one. The traditional fidelity metrics were used to evaluate
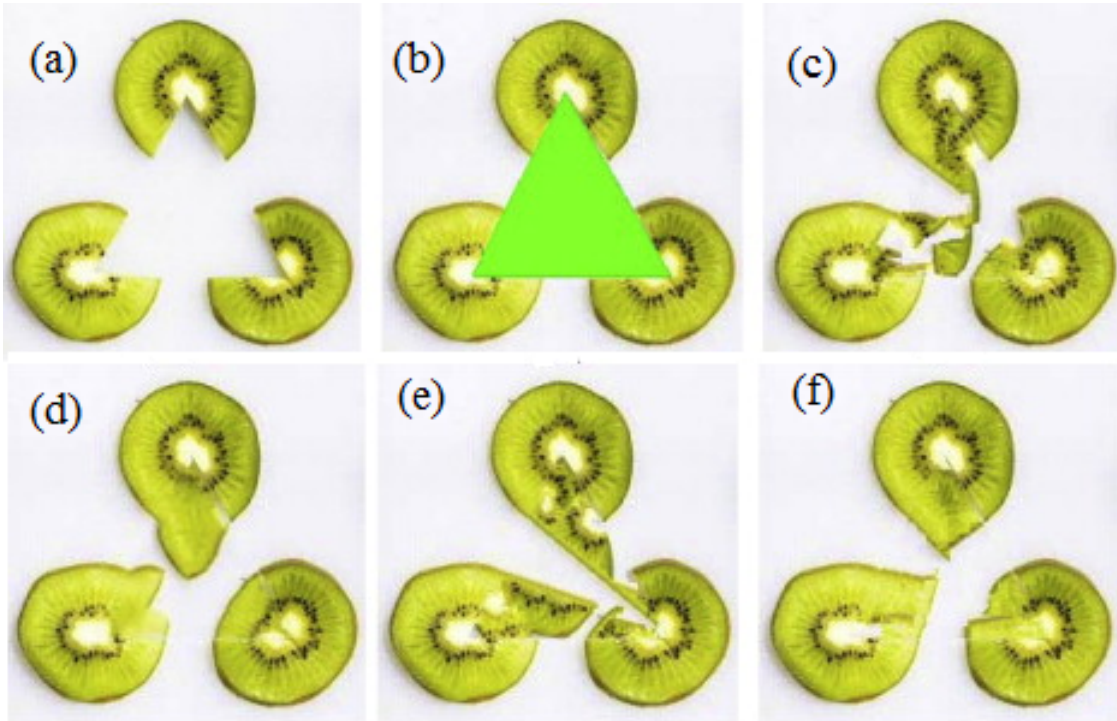
Figure 4.7: An example of broken object restoration using inpainting (a) original image, (b) original image with target region in green color, (c-f) inpainting results from different algorithms [7].

the quality of inpainted images. The MSE and PSNR, which are considered as the most widely used fidelity metrics, were the simplest ones available. Oliveira et al. in [141], for example, used these metrics for quality evaluation of inpainted images. However, both MSE and PSNR are not well correlated with perceptual quality assessment. In inpainting applications, the objective is to restore the original image such that it is more appealing and that the artifacts introduced inside, outside, and around the inpainted regions, are not noticeable/visible.

For performance evaluation of different inpainting algorithms, the metric of choice would be a qualitative judgment averaged over a number of human observers. In this regard, Hays et al. [142] qualitatively evaluated inpainting image quality for the first time using subjective experiments. The purpose of the experiment was just the identification of the original and the tampered (inpainted) images. The proposed method was compared with an exemplar-based approach

[11]. Twenty naive observers participated in the subjective tests and were asked to differentiate between the real image and the inpainted image (fake image). The detection rate for tampered images was achieved as 34%, 64%, and 3% for the images inpainted by their algorithm, [11], and for original images. The purpose of the study was to investigate whether or not the proposed inpainted algorithm produced better perceptual quality image compared to other methods. However, the study did not provide any quantitative ratings of inpainted images.

Subjective assessment methods involve humans and the ratings provided are considered as most reliable and accurate in relation to perceived quality. However, these methods are time-consuming, laborious, and require a significant number of observers to be consistent. They also require a well-controlled environment and lighting conditions. This has motivated researchers in this field to develop alternative objective metrics for inpainting quality assessment without the need of human involvement. Such objective methods use mathematical tools to extract characteristic features from either the reference or the test images or both. These features are then used to get a single quality score for the given image. The aim of objective quality assessment techniques is to predict perceived image quality, the way a human observer perceives it.

Traditional objective quality assessment methods, depending upon the availability of the original image, are grouped into Full Reference (FR), Reduced Reference (RR), and No Reference (NR) methods. In FR methods, the original image is required in addition to the processed image (inpainted image). These are impractical as the original image is usually not available. With NR quality prediction methods, the original image is not available. For RR techniques, partial information about the original images is available in the form of some extracted features. RR techniques are seen as a compromise between FR and NR methods. For the case of inpainting, only NR methods are considered as the original

images are not available. Based on our analysis of the literature, we decided to group IIQA measures broadly into structural-based, saliency-based, and machine learning-based measures. We display in Fig. 4.8 a tree diagram listing the different metrics commonly used for IIQA. The different categories are now discussed in more details.

For completeness, we define the notations we used in this chapter. We represent an image to be inpainted with I_r and the inpainted image with I_{inp}. The original image to be inpainted is decomposed into three distinct regions, (1) The region to be restored or modified by the inpainting algorithm is represented by Ω, (2) The remaining area is denoted by Φ, and (3) The boundary between the two regions is indicated by $\delta\Omega$. Fig. 4.9 shows an image inpainting model where different regions are clearly labeled. After defining the notations, we will now start discussing each of the groups of IIQA mentioned in Fig. 4.8.

4.3.1 Structure-based IIQA measures

In image inpainting, some of the structural details in the original image are either removed or replaced. Inspired by the use of SSIM [48] in traditional IQA, Wang et al. [143] proposed a FR metric using luminance, definition, and gradient similarities to determine a quality index for inpainted images. The metric, defined as Parameter Weight Image Inpainting Quality (PWIIQ), is calculated as follows:

$$\text{PWIIQ} = \big[L(I_r, I_{inp})\big]^{\alpha}\big[D(I_r, I_{inp})\big]^{\beta}\big[G(I_r, I_{inp})\big]^{\gamma} \qquad (4.1)$$

where the terms L, D, and G, represent the variances of image luminance, definition, and gradient similarity between the original and inpainted images. The α, β, and γ are positive parameters used to determine the importance of each term in the final quality score.
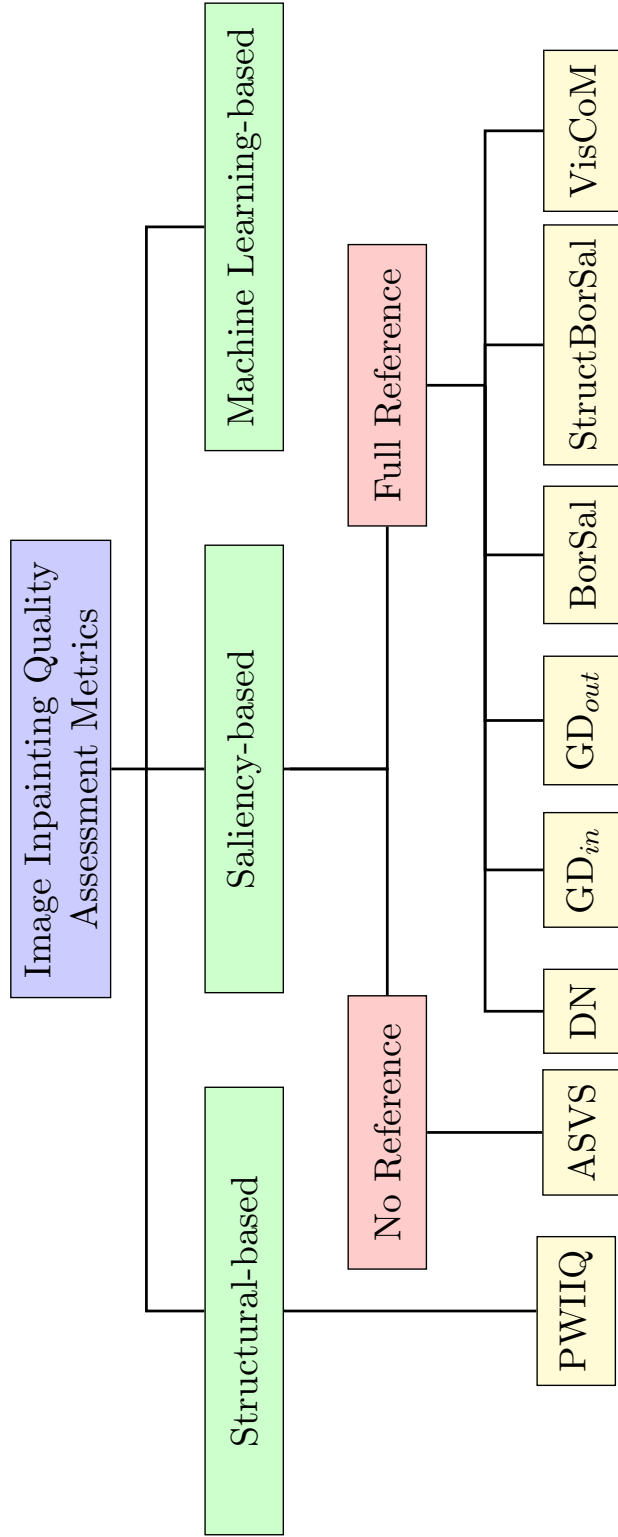
Figure 4.8: Proposed framework for grouping the different IIQA metrics
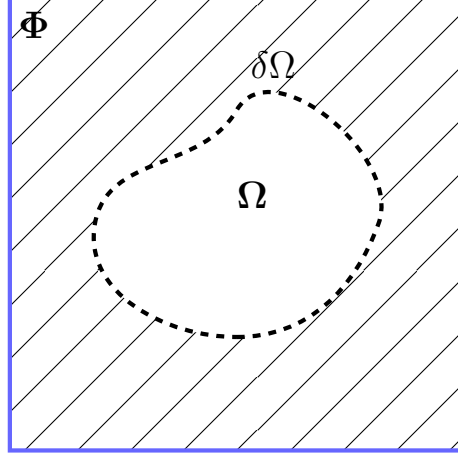
Figure 4.9: A simple model used in image inpainting techniques

For implementation purposes, both the original and the inpainted images are first divided into $b \times b$ fixed-size blocks, and the luminance similarity between the corresponding blocks is computed:

$$l(I_r, I_{inp}) = \frac{2\mu_r \mu_{inp} + K_1}{\mu_r^2 + \mu_{inp}^2 + K_1} \tag{4.2}$$

where μ_r and μ_{inp} represent the mean value of the original and inpainted image blocks respectively, while K_1 is a positive constant with very small value to avoid instability when the denominator is close to zero.

The weighted block means for the original and inpainted images are used. The weights are computed from the symmetrical Gaussian filter window of size 11×11 pixels. The resulting weighted mean is given by:

$$\mu = \sum_{i=1}^{N} \text{w}_i x_i \tag{4.3}$$

where $\text{w} = \left\{ \text{w}_i \text{ such that } \sum_{i=1}^{N} \text{w}_i = 1, \quad i = 1, 2, \cdots, N \right\},$ where N denotes the number of pixels in the window.

The luminance component, L, is computed as the average of the luminance

similarities across all blocks:

$$L(I_r, I_{inp}) = \frac{1}{B_1 \times B_2} \sum_{i=1}^{B_1} \sum_{j=1}^{B_2} l(I_r^{ij}, I_{inp}^{ij}) \tag{4.4}$$

where B_1 and B_2 represent the number of overlapping blocks along the rows and columns of the image.

Secondly, the image definition function, D, is computed as follows:

$$D(I_r, I_{inp}) = \frac{\sum\limits_{i=0}^{W-1} \sum\limits_{j=0}^{H-1} |\mathcal{F}_{inp}^{ij}| - |\mathcal{F}_{inp}^{00}|}{\sum\limits_{i=0}^{W-1} \sum\limits_{j=0}^{H-1} |\mathcal{F}_r^{ij}| - |\mathcal{F}_r^{00}|} \tag{4.5}$$

where $\mathcal{F}(.)$ represents the Fourier transform of an image and $\mathcal{F}_{00}$ is the dc component or overall mean value of an image.

Finally, the gradient component is defined as:

$$G(I_r, I_{inp}) = \frac{2 \sum\limits_{i=0}^{W-1} \sum\limits_{j=0}^{H-1} G_r^{ij} G_{inp}^{ij} + K_2}{\sum\limits_{i=0}^{W-1} \sum\limits_{j=0}^{H-1} \left[G_r^{ij}\right]^2 + \sum\limits_{i=0}^{W-1} \sum\limits_{j=0}^{H-1} \left[G_{inp}^{ij}\right]^2 + K_2} \tag{4.6}$$

where the $G(.)$ represents the gradient magnitude computed from the Sobel filter mask of size 3×3 in the vertical and horizontal directions and K_2 is a small positive constant.

Similarly to its IQA counterpart, the structure-based methods suffer from some serious limitations. Since image inpainting operations do not require the original images, the large inpainted regions may be quite different from the actual ones. Consequently, the structural similarity based methods (e.g., [143]) may fail for images with large inpainted regions. To overcome the drawbacks of structure-based methods, researchers started introducing image saliency to derive new measures

for evaluating quality of inpainted images.

4.3.2 Saliency-based IIQA measures:

Visual saliency plays a significant role in image quality assessment applications. Image saliency is used to highlight the areas towards which the human vision system is more sensitive. Various saliency detection algorithms exist in the literature [144, 145]. Given its importance in IQA, saliency has been used in estimating visibility of different artifacts introduced by the inpainting process. The basic idea is that salient regions change after inpainting. The most prominent IIQA metrics using the concept of saliency are now briefly outlined.

Average Squared Visual Salience (ASVS): In [146], Ardis et al. proposed two objective metrics for quality assessment of inpainted images. The image saliency was used in capturing the distortions introduced during the restoration process. The first metric is the ASVS, which is represented by the normalized sum of squares of the saliency values within the inpainted region. The ASVS metric relates to the noticeability of the inpainted pixels compared to the overall scene. ASVS is a NR metric as it does not require the original image information. It is calculated as follows:

$$\text{ASVS} = \frac{1}{||\Omega||} \sum_{p \in \Omega} |S'(p)|^2 \qquad (4.7)$$

where $S'(p)$ represents the saliency value for the inpainting pixel, p, related within the inpainted region, Ω. High values of the ASVS correspond to more visibility of inpainting related artifacts and reduced perceptual quality [146].

Degree of Noticeability (DN): Ardis et al. in [147], categorized inpainting artifacts into two broad classes, i.e., in-region and out-region artifacts. During the restoration operation in image inpainting, the in-region artifacts occur due

to the introduction of distinct color and structures in the inpainted regions only. These artifacts result in an increased saliency in the inpainted areas and thus disturb attention flow. The ASVS metric relates to the in-region artifacts as it only considers the salient pixels within the inpainted region.

Similarly, the out-region artifacts occur when the local colors or structures are not properly extended to the inpainted region by the inpainted method. These artifacts result in an increase in the saliency of the inpainted region neighborhood and hence decreases attention flow within the inpainted region. The in-region and out-region artifacts are computed as follows:

$$\text{In-region} = \text{ASVS} = \frac{1}{||\Omega||} \sum_{p \in \Omega} |S'(p)|^2 \tag{4.8}$$

$$\text{Out-region} = \frac{1}{||\Phi||} \sum_{p \in \Phi} |S'(p) - S(p)|^2 \tag{4.9}$$

Ardis et al., in [146], took into account both in-region and out-region artifacts and proposed another metric, namely the DN. The DN measure is intended to identify non-noticeable inpainted regions and indicates the change in attention flow in the neighborhood of the inpainted regions. It is calculated as follows:

$$\text{DN} = \frac{||\Omega||}{||\Omega|| + ||\Phi||}\text{in-region} + \frac{||\Phi||}{||\Omega|| + ||\Phi||}\text{out-region} \tag{4.10}$$

Equation (4.10) can further be simplified as follows:

$$\text{DN} = \frac{1}{||\Omega|| + ||\Phi||}\left(\sum_{p \in \Omega} |S'(p)|^2 + \sum_{p \in \Phi} |S'(p) - S(p)|^2\right) \tag{4.11}$$

For both ASVS and DN calculations, the saliency maps are generated using the iLab Neuromorphic Vision Toolkit (iNVT) version 3.1, using scale-4 and discretization of $1 : 16$. The expected visual cortex stimulation was set with 0.1 ms

observation cutoff. Furthermore, four orientation scales, three center scales (2 to 4), and two center-surround channels (3,4) were considered .

Similarly to the ASVS, high values of the DN correspond to more visibility of inpainting related artifacts and reduced perceptual quality. The authors claimed a good correlation for both metrics, with subjective ratings. However, the subjective ratings were not considered reliable as only three observers participated in the psychophysical experiment. Moreover, the overall visual appearance of an image is also ignored while calculating DN and ASVS IIQA metrics.

Gaze Density (GD)-based IIQA measures *:* Following the work in [146], Mahalingam et al. [148] proposed two visual saliency-based metrics for inpainting quality assessment within and outside the inpainted regions. From eye-tracking experimental data, the gaze densities were used to capture the saliency in the original and inpainted images. The motivation was that changes in the saliency map in the inpainted image is related to its perceptual quality.

In their subjective experiments, 45 reference images and 90 modified images were obtained using two different inpainting algorithms. The images were equally distributed into three subsets. Twenty-four naive observers without any prior knowledge of the original and the inpainted images rated the subgroups under ambient lighting conditions and at a distance of 65cm from the display screen. The average gaze distribution was calculated for each image from the eye-tracking experiment. It was observed that the Human Visual System (HVS) is more attracted towards the regions with more noticeable inpainting artifacts. The gaze density was calculated for both inside and outside the inpainted regions using:

$$\text{GD}_{\text{in}} = \frac{1}{||\Omega||} \sum_{p \in \Omega} S'(p) \tag{4.12}$$

$$\text{GD}_{\text{out}} = \frac{1}{||\Phi||} \sum_{p \in \Phi} S'(p) \tag{4.13}$$

The gaze density measures of the inpainted image were normalized by the gaze densities of the original image to account for variations in textures and sizes. The final normalized gaze densities were given by:

$$\overline{\text{GD}}_{\text{in}} = \frac{\sum_{p \in \Omega} S'(p)}{\sum_{p \in \Omega} S(p)}, \quad \text{and} \tag{4.14}$$

$$\overline{\text{GD}}_{\text{out}} = \frac{\sum_{p \in \Phi} S'(p)}{\sum_{p \in \Phi} S(p)} \tag{4.15}$$

The experimental results showed a strong correlation between the rankings from the subjective experiments and the gaze density based measures. However, these methods require the original image and are not suited for practical inpainting applications, where the original image is usually unavailable. Similar to the ASVS and the DN metrics, both GD_{in} and GD_{out} highlight the change in attention flow within and outside the inpainted regions respectively, and do not consider the global visual appearance of the image.

Border Saliency based Measures (BorSal)*:* The measures proposed in [146, 148] considered the in-region and out-region artifacts separately. Oncu et al. [149] showed that saliency map pixels in the neighborhood of the inpainted region are sufficient to capture the changes in saliency due to inpainting. The BorSal metric was proposed to compute the normalized gaze density using the border pixels extended only to three pixels inside and outside the inpainted regions. The six pixels wide border area simultaneously contain information from both in-region and out-region artifacts (see Fig. 4.10). The BorSal metric was computed as follows:

$$\text{BorSal} = \frac{\sum_{p \in Border} S'(p)}{\sum_{p \in Border} S(p)} \tag{4.16}$$

Structural Border Saliency based Measures (StructBorSal)*:* The Bor-Sal IIQA metric accounts for changes in the flow of attention over the inpainted
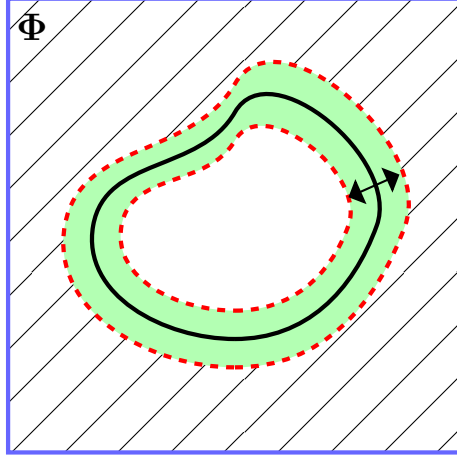
Figure 4.10: A typical model used for BorSal and StructBorSal IIQA metrics computations (shaded region is used for IIQA)

image. Oncu et al. [149] proposed another metric called StructBorSal, to account for the structure information in the whole image and to highlight the artifacts in the restored image. The SSIM_{IPT} based measure [150] was used by taking the geometrical mean of the three SSIM computed for each color channel separately. The StructBorSal metric combines the BorSal metric with the structural measure as follows:

$$\text{StructBorSal} = \text{BorSal} + \text{SSIM}_{IPT} \qquad (4.17)$$

The correlations between the subjective ratings and 14 quality measures (IQA metrics for distortions as well as inpainting) were calculated. The results showed poor performance of existing metrics. The inpainting IQA metrics performed well for images with small and less structured inpainted regions.

The above mentioned saliency-based inpainting IQA metrics, i.e., DN [146], GD_{in} [148] , GD_{out} [148], BorSal [149], and StructBorSal [149] require the original image whereas in the restoration process, usually the original image is not available. The overall visual appearance of the image also plays a significant role in the quality perception. These metrics are also lacking in considering the global visual appearance of an image, limiting their use in practical setups.

Visual Coherence Metric (VisCom): In image inpainting, the reference image is usually not available, therefore the restored pixels rely solely on the surrounding pixels. The restored pixels in the inpainted regions, however, should exhibit consistency with existing pixels. The coherence of the inpainted regions, which is associated with the degree of annoyance of noticeable distortions, is computed by taking the correlation between the inpainted pixels and the original ones. Similarly, the HVS is more sensitive to the edges and contours in an image. The presence of contours and edge details are more attracted by the HVS compared to the remaining regions. The saliency map of a given image relates to the degree of attention in the image and hence can be used to weight the coherence map in evaluating the final quality index.

Trung et al. [151, 152, 153, 10], proposed a FR quality metric using visual coherence and visual saliency of restored regions. The index is computed as follows:

$$\text{VisCoM} = \frac{1}{||\Omega||} \sum_{p \in \Omega} C(p)^{\alpha} S(p)^{\beta} \qquad (4.18)$$

where $C(p)$ and $S(p)$ represent the coherence term and the saliency or structure term respectively. The exponents α and β control the significance of each term in the final quality score.

The coherence term, C, is basically a similarity index between the inpainted regions and the remaining ones in the inpainted image. It is defined as follows:

$$C(p) = \max \left[SIM(\Psi_p, \Psi_q), \qquad \forall \Psi_q \in \Phi, \quad \forall \Psi_p \in \Omega \right] \qquad (4.19)$$

where Ψ_p and Ψ_q represent small patches around pixels, p and q, respectively. The SIM is the similarity function between two patches. Traditional similarity measures, such as the MSE and PSNR, are not well correlated with perceptual quality. The Structural Similarity Metric (SSIM) [48], used in classical IQA, is

133

exploited here to find the coherence between patches of size 7×7. The SIM is defined as:

$$SIM(\Psi_p, \Psi_q) = \frac{(2\mu_p\mu_q + K_1)(2\sigma_{pq} + K_2)}{(\mu_p^2 + \mu_q^2 + K_1)(\sigma_p^2 + \sigma_q^2 + K_2)} \tag{4.20}$$

where μ_p, σ_p and μ_q, σ_q represent the mean and standard deviation of the patches, Ψ_p and Ψ_q, centered at pixels, p and q, respectively, whereas σ_{pq} is the cross correlation between the patches Ψ_p and Ψ_q, and K_1 and K_2 are small positive coefficients to insure stability when denominator is zero.

The local structure term is computed from the saliency values which are further used as weights in the final quality index. Among different saliency detection algorithms, the authors used a simple and computationally efficient method for salient region detection [154]. In [154], color and luminance information were used for saliency detection. For a given image, I, the saliency map was generated using:

$$S' = ||I_\mu - I_G|| \tag{4.21}$$

where I_μ represents the mean value of the original image, and I_G is a Gaussian blurred (5×5 filter mask) version of the original image. The operation is performed in the CIELab color space. The method is simple, computationally efficient, and does not need any downsampling operation during the estimation of the saliency map. Finally, the saliency map, defined in Eq. 4.21, is normalized to the range $[0, 1]$:

$$S(p) = \frac{S'(p)}{\max_I(S')} \qquad \forall p \in \Omega \tag{4.22}$$

The authors in [151, 152, 153, 10] used the visual coherence of recovered regions and visual saliency describing visual importance to develop their index shown in Eq. 4.18. The proposed approach showed promising results but could only handle a limited number of possible inpainting artifacts.

Based on our study of existing approaches, we present in Table 4.1 a summary

of both structure-based and saliency-based IIQA metrics. It is important to note that there exists only one NR-IIQA metric among these metrics. To overcome this limitation, among others, researchers tried to use advanced machine learning approaches in developing robust quality assessment metrics for practical inpainting applications.

4.3.3 Machine Learning based IIQA Measures

Machine learning-based approaches were originally developed for solving classification and regression problems efficiently and provide good approximation of functional relationships between input features and output classes/scores scores in the training session. In testing stage, a set of features is extracted from a given image. The trained model and the extracted features are then used for predicting the quality rating of test image.

Among the first approaches using machine learning for IIQA is the metric proposed by Viacheslav et al. in [155]. The method is an NR approach for IIQA based on natural scene statistics and machine learning. First, the saliency map of the inpainted image is calculated to identify most important perceptual information in the inpainted image. The saliency map is then thresholded using average gaze density computed from the outside inpainted regions using Eq. (4.13) for proto-objects. Then, the DCT is calculated only for the proto-objects and used to train a dictionary of 100 classes, where each word in the dictionary is a DCT coefficient. For each DCT block, the histogram of words is then used as a feature vector. The quality scores collected from the subjective experiment and the extracted features were then used to train a Support Vector Regression (SVR) network and to predict the quality of inpainted images resulting from different algorithms.

The same authors in [156] replaced the DCT based features with the tradi-

Table 4.1: A summary of structure-based and saliency-based image inpainting quality assessment measures

Metric	Year	Type	Regions used	Description	Strengths	Weaknesses
PWIIQ[143]	2008	FR	overall image	statistical features used	simple and fast	requires reference image, fails for images with large holes (inpainted regions).
ASVS[146]	2009	NR	in-region	used when fidelity is not important, high value means more visibility of artifacts, poor quality	reference image is not needed	ignore overall visual appearance of an image
DN [146]	2009	FR	in-region, out-region	used for applications where preservation of original saliency is desired.	highlights change in attention flow beyond the in-painted regions	requires reference image, ignore overall visual appearance of an image
GD_{in} [148]	2010	FR	in-region	values close to 1 means no deviation of attention flow in the inpainted image	highlights change in attention flow within the in-painted regions	requires reference image,
GD_{out} [148]	2010	FR	out-region	same as GD_{in}	highlights change in attention flow beyond the in-painted regions	requires reference image,
BorSal [149]	2012	FR	border region	a single border region around the hole (inpainted region) is used.	fast, a single border region around the hole (inpainted region) is used.	requires reference image,
StructBorSal [149]	2012	FR	border region	in addition to BorSal, also uses structural information.	fast, a single border region around the hole (inpainted region) is used, structural artifacts are also captured	requires reference image, coherence of the inpainted regions with remaining is ignored
VisCoM [151]	2013	FR	overall image	uses visual coherence along with structural information	The overall global visual appearance of image is considered	requires reference image,

tional Local Binary Pattern (LBP) features given their power in describing image structures effectively. The quality scores collected from the subjective experiments and the extracted features are then used to train an SVR for quality prediction. For the subjective experiments, in [156], a database consisting of 300 images with different structures and textures was used. The database also included some real images. The images were restored using a mask and using four different inpainted methods. Ten observers participated in the subjective experiments and rated the quality of the inpainted images on a scale 1-5 (5 for excellent quality, 1 for bad quality). The results showed good correlation with human ratings of quality.

Recently, Markio et al. [157] demonstrated that saliency is not an absolute requirement for assessing inpainting quality. They performed an experiment using a learning-to-rank approach. Instead of determining the absolute scores for inpainted images, the preference order is obtained among inpainted images from different inpainting algorithms. They demonstrated that visual saliency map is useful but not a requirement. Rather, they showed that some features can be used to reflect the changes within and outside the modified areas in an inpainted image. Such features are extracted from gaze measurements using a simple Tobii eye-tracker device. From each original image, twelve inpainted images are generated from two inpainted methods, three patch sizes, and two multiscale parameters. One-hundred eleven original images were used in the experiments. The proposed metric was compared to other existing metrics in terms of prediction accuracy in estimating the preferences order ranking. The authors showed that existing saliency-based IIQA metrics fail in ordering the inpainted images correctly due to the small significant difference in the saliency maps in the inpainted regions. The results showed an improvement of at least 7% over other metrics with 68.65% prediction accuracy. Table 4.2 provides summary of above-discussed machine learning-based IIQA approaches.

Table 4.2: A summary of machine learning based IIQA measures

Method	Year	Type	Feature Description	Regression
Voronin et al. [155, 158]	2014	NR	DCT-based dictionary	SVR, RBF[1] kernel
Voronin et al. [156]	2015	NR	LBP histograms	SVR, EMD[2] kernal
Markio et al. [157]	2016	NR		RankingSVM, RBF kernel

[1] RBF (Radial Basis Function)
[2] EMD (Earth Mover's Distance

4.4 Image Inpainting Quality Assessment Databases

With the tremendous increase of research activities in image inpainting algorithms and applications, it was crucial to develop comprehensive databases for performance evaluation and benchmarking of different inpainting methods. In the literature, usually, the performance of an inpainting algorithm is evaluated on own local images or using standard databases used for standard IQA problems. Given the importance of image inpainting in multimedia applications, publicly-available databases are needed for unbiased performance comparison. In this regards, Tiefenbacher et al. provided, for the first time, a public database namely the Technische Universitt Mnchen Image Inpainting Database (TUM-IID) [8], for objectively estimating quality of inpainted images and performance evaluation of different IIQA metrics. The database contained 17 reference images with diverse texture types and resolution of 640×480 pixels stored in PNG format. Each image in the database is inpainted using four state-of-the-art inpainted methods and for two inpainting regions. Then, each inpainted image in the database was rated by 21 observers using a Single Stimulus (SS) subjective experiment protocol, and the ratings from all observers were averaged to get a single score for each

image. Some sample images, inpainting masks, and inpainted images from public and private databases are shown in Figs. 4.11 and 4.12 respectively. In an effort to summarize existing work in inpainting using different databases (private and public), we present in Table 4.3, the most common experimental setups used in the literature.

4.5 Discussion and Summary

The quality assessment of inpainted images is a complex and challenging problem. It is entirely different from the classical IQA due to different artifacts not commonly observed in other applications. In this work, a critical review of the state-of-the-art IIQA metrics is presented. The study reveals that among the existing IIQA metrics, most of these require original image information. Whereas, image inpainting is usually used in case of unavailability of reference or original image. It is also observed that most of the metrics are designed and validated on private databases consisting of a limited number of images. We have found only one public database [8] with a limited number of images. Seeing the importance of this evolving field, it is extremely desired to develop more public databases consisting of a large number of inpainted images generated from various inpainted methods. This will help in providing fair comparisons among different IIQA metrics, highlighting their shortcomings, and in introducing new efficient quality measures well-correlated with the human perception of quality.
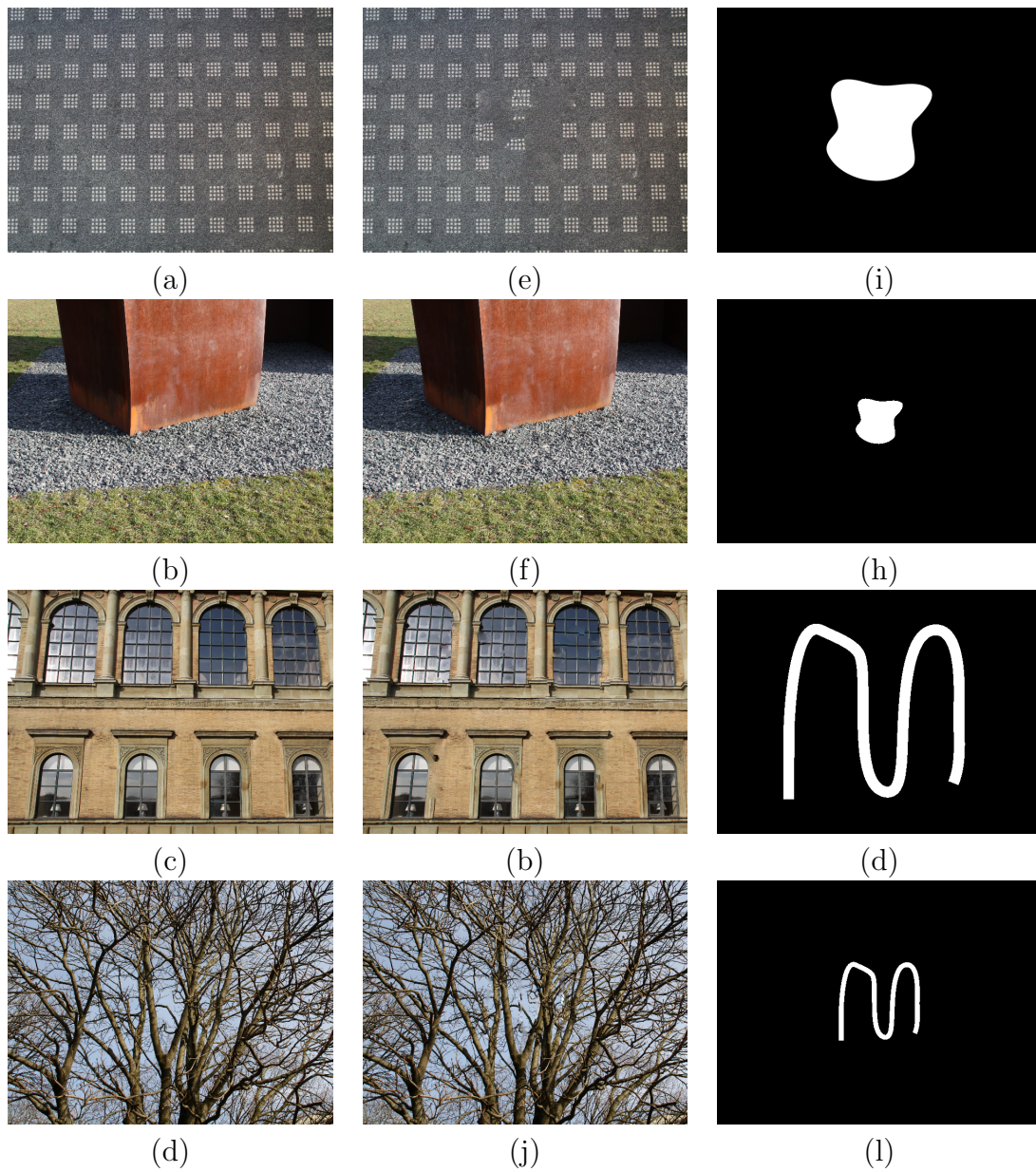
Figure 4.11: Sample images from publicly available TUM-IID [8] database: (a-d) Reference images, (e-h) inpainted images using [9], (i-l) masks used for inpainting.
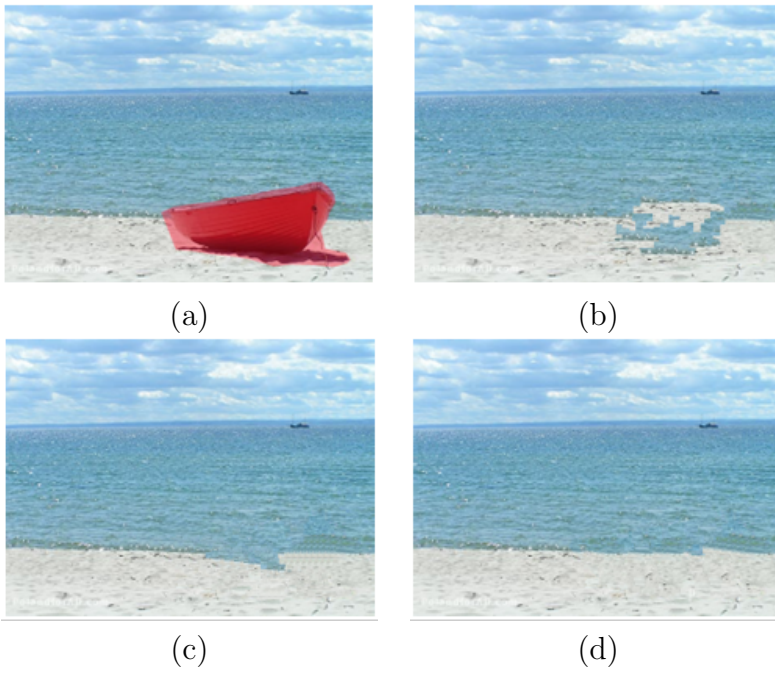
Figure 4.12: Sample local images used in [10] for inpainting quality assessment (a) original image with mask, inpainted image using (b) [11], (c) [12], (d) [13]

Table 4.3: Experimental setups used for private and public inpainting database

Method	[148]	[149]	[146]	[159]	[143]	[8]	[10]	[157]
Year	2010	2012	2009	2015	2008	2015	2013	2016
Original images	45	6	-	300	3	7	8	111
Inpainted images	90	48	-	1200	9	56	-	2466
Image resolution	-	-	-	-	-	640×480	-	-
Subjective method[1]	SS	DS	-	SS	rank-order	SS	web	rank
Score type	gaze density	5-scale	4-scale	5-scale	ranking	5-scale rating	5-scale	5-scale
Screen Resolution[1]	1280×960	-	20cm ht.	-	-	-	-	1280
Observers	24	69	5	10	-	21	$15 - 20$	24
Viewing distance[3]	65cm	web	0.7m	-	-	-	-	60cm
Inpainting methods	2	8	-	4	3	4	-	2
Evaluation measures	2	14	2	4	2	3	-	6
Access	private	private	private	private	private	public	private	private

[1] SS (Single Stimulus), DS (Double Stimulus)
[2] ht. (height)
[3] Web (through web based interface)

CHAPTER 5

CONCLUSION AND FUTURE RECOMMENDATIONS

5.1   Conclusion

In this thesis, a number of contributions were made towards objective quality assessment of image degradations, enhancement, and inpainting. Starting with the quality assessment of distortion, two main contributions were made. A fast NR-IQA metric using fast convolution operations was proposed. The advantages of this technique are simplicity and easy hardware implementations for real-time applications [23]. Another NR-IQA metric was proposed using higher order singular values to quantify blur in color images, which is an important component in the spectrum of distortions. The spatial and inter-channel correlations, in color images, were exploited using tensors to quantify the amount of blur more efficiently and consistently rather than using the traditional luminance component only or the individual color channels as in existing techniques [24].

For quality assessment of enhancement, an extensive literature review of existing state-of-the-art CEE measures was carried. In earlier studies [26, 25], different CEE metrics were analyzed to show how these metrics can help in evaluating

143

artifacts due to CE processes. An NR metric was proposed based on mutual information computed from the gray level co-occurrence matrix and tested with two classical neighborhood-based CE methods. This study was extended to a new dedicated CE database, to see how different CE evaluation measures are consistent with human judgement of image quality enhancement [27, 160].

For quality assessment of inpainted images, a comprehensive and critical review of the state-of-the-art IIQA metrics was presented. The review is the first of its kind and is expected to help researchers working in this area to benchmark new inpainting techniques, develop more robust methods for inpainting quality assessment, and benchmark their results [161].

In addition to the above-mentioned major thesis contributions, additional research work was carried out and some important contributions to the area were made. These additional achievements are the use of Compressed Sensing (CS) for image compression and extensive review of blind image forgery detection techniques.

CS is a recently introduced approach for signal sampling, which allows recovery of sparse signals using fewer measurements, sampled at less than the Nyquist rate. In this regards, two novel image compression algorithms using CS were proposed. The main contribution was the introduction of a sampling scheme based on the optimal representations of the wavelet coefficients to form sparse vectors for CS [162, 163].

Moreover, with the advent of advanced editing tools, we are witnessing a major threat to the multimedia industry, it is becoming easier to alter images/videos in a realistic than it was before. For this reason, a comprehensive literature review of blind image tampering detection techniques was carried [3, 164].

5.2 Future Recommendations

The field of IQA is very important and has diverse applications. Based on the research work discussed in this thesis, the following future recommendations are proposed:

- **IQA for Medical Imaging:** During the past few years, an increasing trend of using e-Health services for remote patient diagnosis was observed. The applications of medical images are Resonance Imaging (MRI), Computed Tomography (CT), and ultrasonic imaging. Medical images may suffer from different types of degradations due to acquisition, transmission or, post-processing operations. It would be desirable to evaluate how image distortions affect the analysis of medical images and the resulting diagnosis rather than the traditional perceptual appeal of natural images. Moreover, the NR-IQA metrics for medical images are important since the reference image is usually not available.

- **Extension of CEE Database:** The proposed CEE database contains only images enhanced from different CE methods. The database is built to evaluate the performance of CEE measures rather than the CE methods. It can be further extended to support different applications such as:

  - Adding more images related to color management. i.e., images enhanced from the methods to improve color appearance, illumination, restoration or other methods used to address problems related to image acquisition.

  - Adding more images created from methods developed to reduce coding artifacts (i.e., ringing, blocking, blurring, quantization, etc.) for performance evaluation of different compression algorithms.

- Including more images created from different denoising methods to evaluate the performance of different denoising methods. The noise can be additive Gaussian white noise, multiplicative, impulse noise, and specular noise.

- Adding more images with different enhancement artifacts (i.e., halo effect, saturation loss, color shift, excessive brightness change, etc.). This can be used to investigate the capabilities of different CEE measures in terms of sensitivity towards these CE artifacts. Currently, there is no database of these kinds exists in the literature.

- **Reduced Time Complexity:** Among different challenges in the field of IQA is the time complexity issue, which is very important for real-time applications. Most of the existing metrics perform well in terms of consistency of their results with the human subjective scores, but they are very complex and computationally inefficient. There is a need to develop objective IQA metrics which are easy to implement and computationally efficient. One of the possible solutions is to use reduced samples from compressed sensing in the design of objective IQA metrics instead of using the whole image information.

- **Novel Applications:** Although substantial research efforts have been put in developing FR, RR, and NR IQA metrics for quantifying distortions and/or enhancement, a very nice application would be to provide information to users playing YouTube videos regarding the quality rating and the types of distortions the video has been subjected to.

- **IQA Software:** It is extremely needed to develop a self-contained software package based on a MATLAB graphical user interface (or other software) which can be used for quality assessment of digital images using state-of-the-

art IQA methods. The software should be fully configurable and it should have the capability to easily add new metrics and to configure each metric with the chosen parameters. It should rate the quality of a given test image in terms of Excellent, Very Good, Good, Fair, and Poor quality image based on the correlation calculated between the subjective scores and the selected metric. The package should support researchers working in IQA filed in terms of comparing the new metrics with other metrics integrated into the software package.

- **Human Face Beauty Assessment:** Digital image retouching is widely used in fashion and multimedia industry, and a growing interest is witnessed in developing tools like Adobe Photoshop, to make photos look more natural and attractive. Although, there exists a lot of image retouching detection methods, limited work has been carried in developing objective metrics to quantify beautification or enhancement of human faces automatically. This is particularly true for females in the entertainment industry. The quality assessment of retouched photos without any information about the original image is considered a challenging task, will be regarded as a significant contribution, and can open new research directions.

REFERENCES

[1] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, sep 2009.

[2] I. N. M. Systems, "QUALINET Multimedia Databases v5 . 0," pp. 1–128, 2014.

[3] M. A. Qureshi and M. Deriche, "A bibliography of pixel-based blind image forgery detection techniques," *Signal Processing: Image Communication*, vol. 39, pp. 46–74, nov 2015.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH '00*. New York, USA: ACM Press, 2000, pp. 417–424.

[5] "Inpaint photo restoration software." [Online]. Available: http://www.theinpaint.com/

[6] D. T. Trung and A. Beghdadi, "Blind Inpainting Forgery Detection," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Atlanta, GA, USA: IEEE, dec 2014, pp. 1019–1023.

[7] R. Rodriguez-Sánchez, J. A. García, and J. Fdez-Valdivia, "Image inpainting with nonsubsampled contourlet transform," *Pattern Recognition Letters*, vol. 34, no. 13, pp. 1508–1518, 2013.

[8] P. Tiefenbacher, V. Bogischef, D. Merget, and G. Rigoll, "Subjective and objective evaluation of image inpainting quality," in *2015 IEEE International Conference on Image Processing (ICIP)*. Quebec City, Canada: IEEE, sep 2015, pp. 447–451. [Online]. Available: http://www.mmk.ei.tum.de/tumiid

[9] A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro, "A Comprehensive Framework for Image Inpainting," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2634–2645, oct 2010.

[10] T. T. Dang, A. Beghdadi, M. C. Larabi, U. Paris, and U. D. Poitiers, "Inpainted image quality assessment," in *Visual Information Processing (EUVIP), 2013 4th European Workshop on*, Paris, France, jun 2013, pp. 76–81.

[11] A. Criminisi, P. Perez, and K. Toyama, "Region Filling and Object Removal by Exemplar-Based Image Inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, sep 2004.

[12] Q. Zhang and J. Lin, "Exemplar-based image inpainting using color distribution analysis," *Journal of information science and engineering*, vol. 28, no. 4, pp. 641—-654, 2012.

[13] T.-T. Dang, C. Larabi, and A. Beghdadi, "Multi-resolution patch and window-based priority for digital image inpainting problem," in *3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2012, pp. 280—-284.

[14] ITU-T Recommendations G.1011, "Reference guide to Quality of Experience assessment methodologies," International Telecommunications Union, Tech. Rep., 2008.

[15] ITU-R Recommendations BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Geneva, Switzerland, Tech. Rep., 2012. [Online]. Available: https://www.itu.int/en/ITU-R/

[16] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of Four Subjective Methods for Image Quality Assessment," *Computer Graphics Forum*, vol. 31, no. 8, pp. 2478–2491, dec 2012.

[17] S. Winkler, F. Dufaux, D. Barba, and V. Baroncini, "Special issue on Image and Video Quality Assessment," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 467–468, aug 2010.

[18] A. Beghdadi, M. C. Larabi, A. Bouzerdoum, and K. M. Iftekharuddin, "A survey of perceptual image processing methods," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 811–831, 2013.

[19] J. Lu, "Contrast enhancement of medical images using multiscale edge representation," *Optical Engineering*, vol. 33, no. 7, p. 2151, jul 1994.

[20] E. Lee, S. Kim, W. Kang, D. Seo, and J. Paik, "Contrast Enhancement Using Dominant Brightness Level Analysis and Adaptive Intensity Transformation for Remote Sensing Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 1, pp. 62–66, jan 2013.

[21] R. Schettini and S. Corchs, "Underwater Image Processing: State of the Art of Restoration and Image Enhancement Methods," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–15, 2010.

[22] L. K. Choi, J. You, and A. C. Bovik, "Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, nov 2015.

[23] M. A. Qureshi and M. Deriche, "A fast no reference image quality assessment using laws texture moments," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*.   Atlanta, GA, USA: IEEE, dec 2014, pp. 979–983.

[24] M. Qureshi, M. Deriche, and A. Beghdadi, "Quantifying blur in colour images using higher order singular values," *Electronics Letters*, vol. 52, no. 21, pp. 1755–1757, oct 2016.

[25] A. Beghdadi, M. A. Qureshi, and M. Deriche, "A critical look to some contrast enhancement evaluation measures," in *2015 Colour and Visual Computing Symposium (CVCS)*.   Gjovik, Norway: IEEE, aug 2015, pp. 1–6.

[26] M. A. Qureshi, M. Deriche, A. Beghdadi, and M. Mohandes, "An information based framework for performance evaluation of image enhancement methods," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*.   Orleans, France: IEEE, nov 2015, pp. 519–523.

[27] M. Qureshi, A. Beghdadi, B. Sdiri, M. Deriche, and F. A. Cheikh, "A Comprehensive Performance Evaluation of Objective Quality Metrics For Contrast Enhancement Techniques," in *European Workshop on Visual Information Processing (EUVIP)*, 2016, pp. 1–5.

[28] ITU-R Rec. BT.500-13, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," International Telecommunications Union, Geneva, Switzerland, Tech. Rep., 2012.

[29] S. Siegel, *Nonparametric Statistics for The Behavioral Sciences*, 2nd ed., CLIFFQRD and T. MoROA, Eds. McGraw-Hill, 1956.

[30] Z. Wang and A. Bovik, *Modern image quality assessment (synthesis lectures on image, video, and multimedia processing)*, S. Rafael, Ed. CA: Morgan Claypool, 2006.

[31] H. R. Sheikh, Z. Wang, L. Cormack, and A. Bovik, "LIVE Image Quality Assessment Database Release 2," 2004. [Online]. Available: http://live.ece.utexas.edu/research/quality

[32] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. November 2015, pp. 30–45, 2009. [Online]. Available: http://www.computervisiononline.com/dataset/tid2008-tampere-image-database-2008

[33] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, jan 2010. [Online]. Available: http://vision.okstate.edu/csiq/

[34] P. Le Callet and F. Autrusseau, "Subjective quality assessment IRCCyN/IVC database," 2005. [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb/

[35] U. Engelke, M. Kusuma, H.-J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 525–547, aug 2009.

[36] U. Engelke, H.-J. Zepernick, and M. Kusuma, "Wireless Imaging Quality Database," 2010. [Online]. Available: http://www.bth.se/tek/rcg.nsf/pages/wiq-db

[37] Y. Horita, K. Shibata, and K. Yoshikazu, "MICT Image Quality Evaluation Database," 2008. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mictdb.html

[38] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, jan 2015.

[39] X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID:IQ - A New Image Quality Database," in *International Conference on Image and Signal Processing (ICISP)*, ser. Lecture Notes in Computer Science, vol. 8509. Cherbourg, Normandy, France: Springer International Publishing, jul 2014, pp. 193–202. [Online]. Available: http://www.colourlab.no/cid

[40] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Hakkinen, "CID2013: A Database for Evaluating No-Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 390–402, jan 2015.

[41] S. Winkler, "Analysis of Public Image and Video Databases for Quality Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, oct 2012.

[42] H. Sheikh, M. Sabir, and A. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, nov 2006.

[43] "VQEG- Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II," Tech. Rep., 2003. [Online]. Available: http://www.vqeg.org

[44] Z. Wang, H. R. Sheikh, and A. C. Bovik, *Objective video quality assessment.* CRC Press, 2003, vol. 45, no. September.

[45] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, may 2011.

[46] D. M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," *ISRN Signal Processing*, vol. 2013, pp. 1–53, 2013.

[47] A. C. B. Zhou Wang, "Mean Squared Error: Love it or leave it?" *IEEE SIGNAL PROCESSING MAGAZINE*, no. January, pp. 98–117, 2009.

[48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[49] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 9–13, 2003.

[50] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. II, no. March, pp. 573–576, 2005.

[51] Y. Chun-Ling, G. Wen-Rui, and P. Lai-Man, "Discrete wavelet transform-based structural similarity for image quality assessment," in *2008 15th IEEE International Conference on Image Processing*, San Antonio, CA, oct 2008, pp. 377–380.

[52] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[53] L. Zhang, Y. Shen, and H. Li, "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, oct 2014.

[54] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 668–695, 2014.

[55] Z. Wang and E. P. Simoncelli, "Reduced-Reference Image Quality Assessment Using A Wavelet-Domain Natural Image Statistic Model," *PIE 5666, Human Vision and Electronic Imaging X*, pp. 149–159, 2005.

[56] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 18, no. 7, pp. 1409–1423, 2009.

[57] W. Xue and X. Mou, "Reduced reference image quality assessment based on Weibull statistics," *2010 2nd International Workshop on Quality of Multimedia Experience, QoMEX 2010 - Proceedings*, pp. 1–6, 2010.

[58] A. Chetouani, A. Beghdadi, M. Deriche, and A. Bouzerdoum, "A reduced reference image quality metric based on feature fusion and neural networks," in *EUSIPCO*, 2011, pp. 589–593.

[59] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," *Proceedings. International Conference on Image Processing*, vol. 3, pp. 8–11, 2002.

[60] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.* IEEE, 2003, pp. 469–472 vol.1.

[61] C.-Y. Wee and R. Paramesran, "Measure of image sharpness using eigenvalues," *Information Sciences*, vol. 177, no. 12, pp. 2533–2552, 2007.

[62] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Quality of Multimedia Experience*, 2009.

[63] X. Marichal, W.-Y. Ma, and H. Zhang, "Blur determination in the compressed domain using DCT information," in *International Conference on Image Processing (ICIP'99)*, 1999, pp. 386–390.

[64] J. Caviedes and S. Gurbuz, "No-reference sharpness metric based on local edge kurtosis," *Image Processing. 2002. Proceedings. . . .*, 2002.

[65] A. Chetouani, A. Beghdadi, and M. Deriche, "A new reference-free image quality index for blur estimation in the frequency domain," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, no. January 2016. Ajman: IEEE, dec 2009, pp. 155–159.

[66] C. T. Vu, T. D. Phan, and D. M. Chandler, "S3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 934–945, 2012.

[67] Zhou Wang, H. Sheikh, and A. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proceedings. International Conference on Image Processing*, vol. 1.   IEEE, 2002, pp. I–477–I–480.

[68] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 14, no. 11, pp. 1918–1927, 2005.

[69] J. Zhang, S. H. Ong, and T. M. Le, "Kurtosis-based no-reference quality assessment of JPEG2000 images," *Signal Processing: Image Communication*, vol. 26, no. 1, pp. 13–23, 2011.

[70] Hanghang Tong, Mingjing Li, Hong-Jiang Zhang, Changshui Zhang, Jingrui He, and Wei-Ying Ma, "Learning No-Reference Quality Metric by Examples," in *11th International Multimedia Modelling Conference*.   IEEE, 2005, pp. 247–254.

[71] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 305–312, 2011.

[72] Chaofeng Li, A. C. Bovik, and Xiaojun Wu, "Blind Image Quality Assessment Using a General Regression Neural Network," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, may 2011.

[73] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 21, no. 7, pp. 3129–38, 2012.

[74] J. Kumar, D. Doermann, P. Ye, and L. Kang, "Unsupervised feature learning framework for no-reference image quality assessment," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1098–1105, 2012.

[75] a. K. Moorthy and a. C. Bovik, "A Two-stage Framework for Blind Image Quality Assessment," *IEEE International Conference on Image Processing*, vol. 17, no. 5, pp. 2481–2484, 2010.

[76] ——, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[77] M. Saad, A. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, 2010.

[78] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, aug 2012.

[79] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 21, no. 12, pp. 4695–708, dec 2012.

[80] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 1146–1153.

[81] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, nov 2014.

[82] A. Chetouani, A. Beghdadi, A. Bouzerdoum, M. Deriche, and P. Orléans, "A New Scheme for No Reference Image Quality Assessment," *3rd International Conference on Image Processing Theory, Tools and Applications*, no. 2012, 2012.

[83] K. I. Laws, "Textured image segmentation," 1980.

[84] D. Bailey and R. Hodgson, "Range filters: Localintensity subrange filters and their properties," *Image and Vision Computing*, vol. 3, no. 3, pp. 99–110, 1985.

[85] D. F. Specht, "A general regression neural network," *Neural Networks, IEEE Transactions on*, vol. 2, no. 6, pp. 568–576, 1991.

[86] R. Ferzli and L. Karam, "A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717–728, apr 2009.

[87] N. D. Narvekar and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, sep 2011.

[88] A. Shnayderman, A. Gusev, and A. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 422–429, feb 2006.

[89] Q. Sang, H. Qi, X. Wu, C. Li, and A. C. Bovik, "No-reference image blur index based on singular value curve," *Journal of Visual Communication and Image Representation*, vol. 25, no. 7, pp. 1625–1630, 2014.

[90] Q.-B. Sang, X.-J. Wu, C.-F. Li, and Y. Lu, "Blind Image Blur Assessment Using Singular Value Similarity and Blur Comparisons," *PLoS ONE*, vol. 9, no. 9, p. e108073, sep 2014.

[91] Y. Wang, W. Liu, and Y. Wang, "Color Image Quality Assessment Based on Quaternion Singular Value Decomposition," in *2008 Congress on Image and Signal Processing*.   Sanya, China: IEEE, may 2008, pp. 433–439.

[92] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, jan 2000.

[93] D. Letexier and S. Bourennane, "Adaptive Flattening for Multidimensional Image Restoration," *IEEE Signal Processing Letters*, vol. 15, pp. 229–232, 2008.

[94] A. Rajwade, A. Rangarajan, and A. Banerjee, "Image Denoising Using the Higher Order Singular Value Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 849–862, apr 2013.

[95] C. Cheng and H. Wang, "Quality assessment for color images with tucker decomposition," in *19th IEEE International Conference on Image Processing*. Orlando, FL: IEEE, sep 2012, pp. 1489–1492.

[96] A. Le Négrate, A. Beghdadi, and H. Dupoisot, "An image enhancement technique and its evaluation through bimodality analysis," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 1, pp. 13–22, jan 1992.

[97] S. Agaian, K. Panetta, and A. Grigoryan, "Transform-based image enhancement algorithms with performance measure," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 367–382, mar 2001.

[98] K. Panetta, C. Gao, S. Member, S. Agaian, and S. Member, "No reference color image contrast and quality measures," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 3, pp. 643–651, 2013.

[99] S. S. Agaian, B. Silver, and K. a. Panetta, "Transform Coefficient Histogram-Based Image Enhancement Algorithms Using Contrast Entropy," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 741–758, mar 2007.

[100] K. Panetta, Yicong Zhou, S. Agaian, and Hongwei Jia, "Nonlinear Unsharp Masking for Mammogram Enhancement," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 6, pp. 918–928, nov 2011.

[101] C. T. Vu, T. D. Phan, P. S. Banga, and D. M. Chandler, "On the quality assessment of enhanced images: A database, analysis, and strategies for augmenting existing methods," in *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*. IEEE, apr 2012, pp. 181–184.

[102] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, feb 2006.

[103] K. Gu, G. Zhai, W. Lin, S. Member, and M. Liu, "The Analysis of Image Contrast : From Quality Assessment to Automatic Enhancement," *IEEE Transactions on Cybernetics*, pp. 1–14, mar 2015.

[104] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-Reference Quality Assessment of Contrast-Distorted Images Based on Natural Scene Statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 1–1, 2014.

[105] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a High Dynamic Range display," *ACM Transactions on Graphics*, vol. 24, no. 3, p. 640, jul 2005.

[106] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Transactions on Graphics*, vol. 29, no. 6, p. 1, dec 2010.

[107] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 626–639, 2012.

[108] Z. Chen, T. Jiang, and Y. Tian, "Quality Assessment for Comparing Image Enhancement Algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH: IEEE, jun 2014, pp. 3003–3010.

[109] S.-D. Chen and A. R. Ramli, "Minimum mean brightness error Bi-histogram equalization in contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 1310–1319, nov 2003.

[110] E. Peli, "Contrast in complex images," *Journal of the Optical Society of America A*, vol. 7, no. 10, pp. 2032–2040, oct 1990.

[111] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, p. 3, jan 2001.

[112] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, sep 2013.

[113] V. L. Jaya and R. Gopikakumari, "IEM : A New Image Enhancement Metric for Contrast and Sharpness Measurements," *International Journal of Computer Applications*, vol. 79, no. 9, pp. 1–9, oct 2013.

[114] A. Saleem, A. Beghdadi, and B. Boashash, "Image fusion-based contrast enhancement," *EURASIP Journal on Image and Video Processing*, no. 1, p. 10, 2012.

[115] K. Panetta, C. Gao, and S. Agaian, "No reference color image contrast and quality measures," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 3, pp. 643–651, aug 2013.

[116] H. Ibrahim and N. Pik Kong, "Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, nov 2007.

[117] T. Arici, S. Dikbas, and Y. Altunbasak, "A Histogram Modification Framework and Its Application for Image Contrast Enhancement," *Image Processing, IEEE Transactions on*, vol. 18, no. 9, pp. 1921–1935, 2009.

[118] K. Subr, A. Majumder, and S. Irani, "Greedy Algorithm for Local Contrast Enhancement of Images," in *IInternational Conference on mage Analysis and Processing (ICIAP)*. Cagliari, Italy: Springer, 2005, pp. 171–179.

[119] G. Boccignone and a. Picariello, "Multiscale contrast enhancement of medical images," *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 10–13, 1997.

[120] L. Tao, "Adaptive and integrated neighborhood-dependent approach for nonlinear enhancement of color images," *Journal of Electronic Imaging*, vol. 14, no. 4, p. 043006, oct 2005.

[121] A. Choudhury and G. Medioni, "Perceptually motivated automatic color contrast enhancement," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops.* IEEE, sep 2009, pp. 1893–1900.

[122] K. Panetta, E. Wharton, and S. Agaian, "Parameterization of logarithmic image processing models," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 2007.

[123] K. Panetta, A. Samani, and S. Agaian, "Choosing the Optimal Spatial Domain Measure of Enhancement for Mammogram Images," *International Journal of Biomedical Imaging*, vol. 2014, pp. 1–8, aug 2014.

[124] A. Beghdadi and A. Le Negrate, "Contrast enhancement technique based on local detection of edges," *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 2, pp. 162–174, may 1989.

[125] M. Deriche, "An Image Quality Index Based on Mutual Information and Neural Networks," *Arabian Journal for Science and Engineering*, vol. 39, no. 3, pp. 1983–1993, mar 2014.

[126] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973.

[127] S. W. Zucker and D. Terzopoulos, "Finding structure in Co-occurrence matrices for texture analysis," *Computer Graphics and Image Processing*, vol. 12, no. 3, pp. 286–308, 1980.

[128] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in *Proceedings of SPIE - The International Society for Optical Engineering,*

B. E. Rogowitz and T. N. Pappas, Eds., vol. 5007, Santa Clara, CA, jun 2003, pp. 87–95.

[129] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgatholer, "Global Contrast Factor - a New Approach to Image Contrast," *Computational Aesthetics in Graphics, Visualization and Imaging*, pp. 159–168, 2005.

[130] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Subjective and objective quality assessment for images with contrast change," in *2013 IEEE International Conference on Image Processing*. Melbourne, VIC: IEEE, sep 2013, pp. 383–387.

[131] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*. San Diego, CA, USA: Academic Press Professional, Inc., 1994, pp. 474–485.

[132] J. Mukherjee and S. Mitra, "Enhancement of Color Images by Scaling the DCT Coefficients," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1783–1794, oct 2008.

[133] R. Hummel, "Image enhancement by histogram transformation," *Computer Graphics and Image Processing*, vol. 6, no. 2, pp. 184–195, apr 1977.

[134] S. Mukhopadhyay and B. Chanda, "A Multiscale Morphological Approach to Local Contrast Enhancement," *Signal Processing*, vol. 80, pp. 685–696, dec 2010.

[135] S. Chen and A. Beghdadi, "Natural Enhancement of Color Image," *EURASIP Journal on Image and Video Processing*, vol. 2010, pp. 1–19, aug 2010.

[136] M. G. Kendall and B. B. Smith, "On the Method of Paired Comparisons," *Biometrika*, vol. 31, no. 3/4, pp. 324–345, mar 1940.

[137] C. Guillemot and O. Le Meur, "Image Inpainting: Overview and Recent Advances," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 127–144, 2014.

[138] Z. Chen, C. Dai, L. Jiang, B. Sheng, J. Zhang, W. Lin, and Y. Yuan, "Structure-aware image inpainting using patch scale optimization," *Journal of Visual Communication and Image Representation*, vol. 40, pp. 312–323, 2016.

[139] S. Basterrech and V. Snášel, *Proceedings of the First International Scientific Conference Intelligent Information Technologies for Industry (IITI'16)*, ser. Advances in Intelligent Systems and Computing, A. Abraham, S. Kovalev, V. Tarassov, and V. Snášel, Eds. Cham: Springer International Publishing, 2016, vol. 450.

[140] Z. Liang, G. Yang, X. Ding, and L. Li, "An efficient forgery detection algorithm for object removal by exemplar-based image inpainting," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 75–85, jul 2015.

[141] M. M. Oliveira, B. Bowen, R. McKenna, and Y.-S. Chang, "Fast Digital Image Inpainting," *International Conference on Visualization, Imaging and Image Processing*, no. Viip, pp. 261–266, 2001.

[142] J. Hays and A. A. Efros, "Scene completion using millions of photographs," *Communications of the ACM*, vol. 51, no. 10, p. 87, oct 2008.

[143] S. Wang, H. Li, X. Zhu, and P. Li, "An Evaluation Index Based on Parameter Weight for Image Inpainting Quality," in *2008 The 9th International*

Conference for Young Computer Scientists. Hunan: IEEE, nov 2008, pp. 786–790.

[144] X. Ma, X. Xie, K.-m. Lam, J. Hu, and Y. Zhong, "Saliency detection based on singular value decomposition," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 95–106, oct 2015.

[145] X. Ma, X. Xie, K.-M. Lam, and Y. Zhong, "Efficient saliency analysis based on wavelet transform and entropy theory," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 201–207, jul 2015.

[146] P. A. Ardis and A. Singhal, "Visual salience metrics for image inpainting," in *Proc. SPIE 7257, Visual Communications and Image Processing*, M. Rabbani and R. L. Stevenson, Eds., San Jose, CA, jan 2009, p. 72571W.

[147] P. A. Ardis, C. M. Brown, and A. Singhal, "Inpainting quality assessment," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011002, jan 2010.

[148] V. V. Mahalingam and S.-c. S. Cheung, "Eye tracking based perceptual image inpainting quality analysis," in *2010 IEEE International Conference on Image Processing.* Hong Kong: IEEE, sep 2010, pp. 1109–1112.

[149] A. I. Oncu, F. Deger, and J. Y. Hardeberg, "Evaluation of Digital Inpainting Quality in the Context of Artwork Restoration," in *Proceeding of the 12th international conference on computer vision*, 2012, pp. 561–570.

[150] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of Spatial Gamut Mapping Algorithms," in *14th Color and Imaging Conferenc (CIC)*, nov 2006, pp. 56–61.

[151] T. T. Dang, A. Beghdadi, and M.-C. Larabi, "Visual coherence metric for evaluation of color image restoration," in *2013 Colour and Visual Computing Symposium (CVCS)*. Gjovik, Norway: IEEE, sep 2013, pp. 1–6.

[152] D. T. Trung, A. Beghdadi, and C. Larabi, "Perceptual quality assessment for color image inpainting," in *2013 IEEE International Conference on Image Processing (ICIP)*. Melbourne, VIC: IEEE, sep 2013, pp. 398–402.

[153] D. T. Trung, A. Beghdadi, and M.-c. Larabi, "Perceptual Evaluation of Digital Image Completion Quality," in *21st European Signal Processing Conference (EUSIPCO 2013)*, Marrakech, sep 2013, pp. 1–5.

[154] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, no. Ic. Miami, USA: IEEE, jun 2009, pp. 1597–1604.

[155] V. Viacheslav, F. Vladimir, M. Vladimir, G. Nikolay, S. Roman, and F. Valentin, "Low-level features for inpainting quality assessment," in *2014 12th International Conference on Signal Processing (ICSP)*, vol. 53, no. 9. Hangzhou: IEEE, oct 2014, pp. 643–647.

[156] V. Voronin, V. Marchuk, E. Semenishchev, S. Maslennikov, and I. Svirin, "Inpainted image quality assessment based on machine learning," in *WSCG 2015 Conference on Computer Graphics, Visualization and Computer Vision*, 2015, pp. 167–172.

[157] M. Isogawa, D. Mikami, K. Takahashi, and A. Kojima, "Eye Gaze Analysis and Learning-to-Rank to Obtain the Most Preferred Result in Image Inpainting," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3538—-3542.

[158] V. A. Frantc, V. V. Voronin, V. I. Marchuk, A. I. Sherstobitov, S. Agaian, and K. Egiazarian, "Machine learning approach for objective inpainting quality assessment," in *Proc. SPIE 9120, Mobile Multimedia/Image Processing, Security, and Applications*, S. S. Agaian, S. A. Jassim, and E. Y. Du, Eds., may 2014, p. 91200S.

[159] V. V. Voronin, V. A. Frantc, V. I. Marchuk, A. I. Sherstobitov, and K. Egiazarian, "Inpainted image quality assessment based on machine learning (report)," in *Proceedings of the SPIE*, K. O. Egiazarian, S. S. Agaian, and A. P. Gotchev, Eds., mar 2015, p. 93990U.

[160] M. A. Qureshi, A. Beghdadi, and M. Deriche, "Towards the design of a consistent image contrast enhancement evaluation measure (submitted)," *Signal Processing: Image Communication*, 2016.

[161] M. A. Qureshi, "A critical Review of Image Inpainting Quality Assessment Techniques (submitted)," *Journal of Visual Communication and Image Representation*, 2016.

[162] M. A. Qureshi and M. Deriche, "A new wavelet based efficient image compression algorithm using compressive sensing," *Multimedia Tools and Applications*, vol. 75, no. 12, pp. 6737–6754, jun 2016.

[163] M. Deriche, M. A. Qureshi, and A. Beghdadi, "An image compression algorithm using reordered wavelet coefficients with compressive sensing," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*. Orleans, France: IEEE, nov 2015, pp. 498–503.

[164] M. Ali Qureshi and M. Deriche, "A review on copy move image forgery detection techniques," in *2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*. IEEE, feb 2014, pp. 1–5.

169

Vitae

Name	:	Muhammad Ali Qureshi
Nationality	:	Pakistani
Date of Birth	:	March 15, 1976
Present Address	:	King Fahd Univeristy of Petroleum and Minerals, P.O. Box 6611, Dhahran 31261, Saudi Arabia.
Email	:	*ali.qureshi@iub.edu.pk*
Permanent Address	:	House No.208, Street No.8, Mazhar Farid Colony, Sadikabad, District Rahim Yar Khan, Punjab, Pakistan.
Contact Numbers	:	+966582022525, +923018675677
Academic Background	:	M.Sc. Telecommunication Engineering, NWFP U.E.T, Peshawar, Pakistan, June 2008. B.Sc. Electrical Engineering, U.E.T Lahore, Pakistan, April 2000.