ROMA TRE

UNIVERSITÀ DEGLI STUDI

DOCTORATE IN APPLIED ELECTRONICS

XXIX OF DOCTORATE CYCLE

QUALITY OF EXPERIENCE OF MULTIMEDIA SERVICES

PhD student (Pradip Paudyal)

_____
signature

Tutor (Prof. Marco Carli)

_____
signature

Coordinator of the PhD program
(Prof. Enrico Silva)

_____
signature

QUALITY OF EXPERIENCE OF MULTIMEDIA SERVICES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF UNIVERSITÀ DEGLI STUDI ROMA TRE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Pradip Paudyal
May 2017

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Professor Marco Carli)   Principal Adviser

Approved for the ....

_____

Dedication

In the loving memory of my late grandfather Devidatta Paudyal, I dedicate my research work to him. May he rest in peace!

Acknowledgments

I would like to thank all those people and institutions who made this dissertation possible. I am indebted to them for the cooperation and help; however, in the following, only few of them are mentioned.

First and foremost, I woe my deepest sense of gratitude to my supervisor Professor Marco Carli, who provided all the support, encouragement, motivation, patience, guidance, opportunities, and freedom one could wish for. I would also thank Professor Federica Battisti for her continued support, guidance, and unflinching encouragement. Both of the professors kindly supervised me in every step of my PhD with utmost devotion and dedication to this point.

Moreover, I would like to thank Professor Alessandro Neri for his cooperation and novel ideas. I would like to express the special thank Professor Federica Pascucci for the help and support. I would also like to thank Professor Patrizio Campisi for encouragement and motivation.

During my PhD, I got lots of opportunities to work with researcher in different countries; in particular, I would thank to COST Action IC1105 3D-ConTourNet for the Short Term Scientific Mission (STSM) funding. I would like to thank Professor Paulo Nunes, ISCTE, Instituto Universitario Lisboa, Portugal, for the cooperation, support, and guidance provided during my research period in his lab. I would also like to thank Professor Lus Ducla Soares in the same institution for the ideas and cooperation. Furthermore, I would like to thank Caroline Conti, ISCTE, Instituto Universitario Lisboa, Portugal, for cooperation and support.

Similarly, I would like to express my gratitude to Professor Mårten Sjöström, Department of Information and Communication Systems (IKS), Mid Sweden University, Sundsvall, Sweden, for accepting me as a visiting researcher in his lab, and the support and guidance provided during my research stays in Sweden. I would also like to thank Dr. Roger Olsson in the same research lab for cooperation and help.

Furthermore, I would like to extend my sincerest thank to professor Patrick Le Callet, IRC-CYN/IVC Lab, Polytech Nantes/Universit de Nantes, France, for opportunity, cooperation, supervision during my research stay at IRCCYN/IVC Lab, and continuing encouragement and support from the beginning of my PhD. I would also thank to all the researchers at the lab, particularly Dr. Jesús Gutiérrez and Lukas Krasula, for the help and cooperation.

Abstract

Nowadays, multimedia services are becoming a part of the individual's life; and thus, many ongoing efforts have been given to develop and optimize the technology for multimedia content creation, distribution, and consumption. As a result of advancement in the technology, there is a rivalry among the standards, services, and thus, the service providers. In this context, assurance of the consumer satisfaction by providing a required level of quality for the provided service is of crucial importance for the service providers. From another viewpoint, providing a better Quality of Experience (QoE) to the provided service could be a differentiation strategy to achieve the competitive advantage. To provide the best QoE, all the involving entities in the multimedia communication chain should pose a defined quality level. Moreover, the QoE centric operations and optimization techniques should be incorporated in the multimedia communication system. For QoE centric operation and optimization, the assessment of the QoE of the service is an important step, and thus, the QoE assessment is a scope of this dissertation.

Among multimedia services, the video communication is becoming a dominant service. Moreover, the latest advancement in imaging technology–light field imaging, is expected to be a next generation imaging technology. Due to the possibilities of a wide range of applications provided by the light field imaging, lots of attention is pulled from industry and academia by this technology. Therefore, this dissertation is mainly focused on two types of multimedia content: video and light field image.

Every step involved in multimedia communication such as acquisition, representation, encoding, network, and rendering/presentation, produces the artifacts, and ultimately degrades the quality. The main aim of this dissertation is to propose the novel theories and frameworks for assessing the QoE of video and light field image.

First part of the dissertation is focused on video QoE assessment. Before designing a quality assessment metric, and for optimizing the multimedia communication networks, the knowledge of the impact of transmission impairments on video QoE is crucial. In communication environment, the spatial-temporal sensitive video content need to be transmitted over a noisy and bandwidth limited wired or wireless channel. The network produced impairments introduce the artifacts, and the artifacts ultimately degrade the video QoE. The achieved results show that there is a significant impact of the impairments: packet loss, jitter, and bandwidth on video QoE. However, the initial

delay introduced by the network does not adversely affect the QoE. From the analysis results it is noticed that, the video QoE and the level of quality distortion introduced by the network impairments is also depending on the video scene i.e. video content itself. Therefore, a study is performed to understand the impact of visual content on the QoE. The dataset designed to study the impact of transmission impairments in video QoE, is also used to benchmark the state-of-the-art quality measures. The achieved results strengthen the need of a No-Reference metric for video communication services. Therefore, a blind QoE assessment metric for video communication services is proposed. The performance evaluation of the proposed metric evidences its superiority over existing metrics.

Next, a big revolution in image acquisition systems is given by the introduction of the light field imaging. The basic idea behind the light field imaging systems is the use of a micro-lenses array positioned at the focal point of the camera lens, in front of the image sensor. In this way, it is possible to record multiple views of a scene by using a single camera in a single shot, thus avoiding problems related to calibration and camera synchronization. The micro lens array records information on the incident light direction at different positions, i.e. it records the light field. The availability of low cost acquisition devices turned the light field theory into practice, thus allowing novel applications of the imaging systems. Currently, many ongoing efforts have been given towards the optimization and standardization of the system. In this context, the knowledge of the quality of the processed images is important.

The work has been started by defining the quality related issues, and quality distortion model of the light field technology; which is significantly different compared to the traditional 2D/3D visual technology. Next, the increasing interest towards this media calls for methods to protect these data from manipulations and unauthorized reproduction or diffusion. For this purpose, a watermarking scheme is designed by exploiting a tradeoff between quality, robustness, and capacity for copyright protection. Evaluating the quality issue, that is the imperceptibility of the watermark or its impact on the cover data, is a challenging task. This becomes more challenging for light field data, because there is no validated or standard quality assessment protocol available for light field data. Therefore, our next aim is to devise a quality assessment framework for light field image. Towards, this aim a Reduced Reference light field image quality assessment framework is proposed. The achieved analysis results show that the predicted quality scores are very close to the corresponding subjective opinion scores.

The lack of a validated subjected quality assessment framework for light field image pushes us to work in this topic–to propose a quality assessment framework for light field images, particularly for the images captured by using the commercially available light field cameras such as Lytro Illum and Raytrix. Therefore, an extensive study was performed by considering the Lytro Illum images. During the study, four encoding methods, including standard image compression methods and recently the proposed light field image compression methods are considered. As a result of the study, a new light field image quality dataset is available for the research community. Moreover, the dataset is

used to study the impact of compression artifacts on light field image QoE and to benchmark the existing quality metric, when applied for light field image. Following, this study, a subjective quality assessment framework is proposed for light field field imaging.

Publications and deliverables

Some of the contributions of this dissertation are also published in the following:

International Journals:

- Pradip Paudyal, Federica Battisti, Mårten Sjöström, Roger Olsson, and Marco Carli. "Towards the Perceptual Quality Evaluation of Compressed Light Field Images." IEEE Transactions on Broadcasting (2017).

- Pradip Paudyal, Federica Battisti, and Marco Carli. "Evaluation of the Effects of Transmission Impairments on Perceived Video Quality by exploiting ReTRiEVED Dataset." Journal of Electronic Imaging, SPIE (2017).

- Pradip Paudyal, Federica Battisti, and Marco Carli. "Impact of video content and transmission impairments on quality of experience." Multimedia Tools and Applications, Springer (2016).

- Pradip Paudyal, Marco Carli, and Federica Battisti. "Reduced Reference Quality Assessment of Light Field Image". IEEE Transactions on Image Processing (Submitted).

- Federica Battisti, Marco Carli, Patrick Le Callet, and Pradip Paudyal. "On the impact of asymmetric HEVC encoding of stereo video sequences on the perceived quality". IEEE Transactions on Broadcasting (Submitted).

International Conference Proceeding:

- Pradip Paudyal, Federica Battisti, and Marco Carli. "Effect of visualization techniques on subjective quality of light field images" IEEE International Conference on Image Processing (ICIP), 2017.

- Pradip Paudyal, Jesús Gutiérrez, Patrick Le Callet, Marco Carli, and Federica Battisti. "Characterization and selection of light field content for perceptual assessment." 9th International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2017.

- Pradip Paudyal, Yiwei Liu, Federica Battisti, and Marco Carli. "Video Quality of Experience metric for streaming services." Electronic Imaging, 2016.

- Pradip Paudyal, Roger Olsson, Mårten Sjöström, Federica Battisti, and Marco Carli. "SMART: a light field image quality dataset." In 7th International Conference on Multimedia Systems (MMSys), ACM, 2016.

- Pradip Paudyal, Federica Battisti, Alessandro Neri, and Marco Carli. "A study of the impact of light fields watermarking on the perceived quality of the refocused data." In 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), IEEE, 2015.

- Federica Battisti, Marco Carli, Yiwei Liu, Alessandro Neri, and Pradip Paudyal. "Distortion-based no-reference quality metric for video transmission over IP." In International Symposium on Signals, Circuits and Systems (ISSCS), IEEE, 2015.

- Pradip Paudyal, Federica Battisti, and Marco Carli. "Study of the effects of video content on quality of experience." In SPIE Electronic Imaging. International Society for Optics and Photonics, 2015.

- Pradip Paudyal, Federica Battisti, and Marco Carli. "A study on the effects of quality of service parameters on perceived video quality." In 5th European Workshop on Visual Information Processing (EUVIP), IEEE, 2014.

- Federica Battisti, Marco Carli, and Pradip Paudyal. "QoS to QoE mapping model for wired/wireless video communication." In Euro Med Telco Conference (EMTC), IEEE, 2014.

Deliverables:

- Pradip Paudyal. "Watermarking in Light Field Images." ECOST-STSM-IC1105-130715-062788 Scientific Report, 2015 (with Prof. Paulo Nunes, ISCTE, Instituto Universitario Lisboa, Portugal, Period: 13-07-2015 to 24-07-2015).

- Pradip Paudyal,"Creation of a Light Field database for LF image quality assessment." ECOST-STSM-IC1105-090315-057813 Scientific Report, 2015, (with Prof. Mrten Sjstrm, Mid Sweden University, Sundsvall, Sweden, Period: 13-07-2015 to 24-07-2015).

Contents

List of Tables

List of Figures

Acronyms

ACR Absolute Category Rating

ACR-HR Absolute Category Rating with Hidden Reference

ANOVA Analysis of Variance

AVC Advanced Video Coding

AWGN Additive White Gaussian Noise

BER Bit Error Rate

BIQI Blind Image Quality Index

BLIINDS-II BLind Image Integrity Notator using DCT Statistics-II

BRISQUE Blind/Referenceless Image Spatial QUality Evaluator

CDF Cumulative Distribution Function

CDN Content Delivery Network

CF Colorfulness

CI Confidence Interval

CIF Common Intermediate Format

DCR Degradation Category Rating

DMOS Difference Mean Opinion Score

DSCQS Double Stimulus Continuous Quality Scale

DSIS Double Stimulus Impairment Scale

DWT Discrete Wavelet Transform

ETSI European Telecommunications Standards Institute

FR Full Reference

FVV Free Viewpoint Video

GLCM Gray Level Co-occurrence Matrix

GUI Graphical User Interface

HDTV High Definition Digital Television

HEVC High Efficiency Video Coding

HRC Hypothetical Reference Circuit

HVS Human Visual System

IP Internet Protocol

IQA Image Quality Assessment

IQM Image Quality Metric

IRU-R BT ITU-R Broadcasting service

ISO International Organization for Standardization

ITU International Telecommunication Union

ITU-R ITU Radiocommunication Sector

ITU-T ITU Telecommunication Standardization Sector

JND Just Noticeable Difference

JPEG Joint Photographic Experts Group

LF Light Field

MJPEG Motion JPEG compression

MOS Mean Opinion Score

MPEG Moving Picture Experts Group

MSE Mean Squared Error

MSE Mean Squared Error

NETEM Network Emulator

NIQE Naturalness Image Quality Evaluator

NR No Reference

PC Pair Comparison

PCA Principle Component Analysis

PCM Pair Comparison Matrix

PIS Processed Image Sequences

PLCC Pearson's Linear Correlation Coefficient

PLR Packet Loss Rate

PSNR Peak Signal to Noise Ratio

PVS Processed Video Sequence

QCIF Quarter Common Intermediate Format

QoE Quality of Experience

QoL Quality of Life

QoS Quality of Service

QoSE Quality of Service Experienced

RR Reduced Reference

SAMVIQ Subjective Assessment of Multimedia Video Quality

SDTV Standard Definition Digital Television

SI Spatial perceptual Information

SIF Source Input Format

SNR Signal-to-Noise Ratio

SNOW Wavelet-based compression using the Snow codec

SRC Source Sequence

SRCC Spearman Rank Correlation Coefficient)

SS Single Stimulus

SSCQE Single Stimulus Continuous Quality Evaluation

SSIM Structural Similarity Index

TI Temporal perceptual Information

UHD Ultra High Definition

VoD Video on Demand

VQA Video Quality Assessment

VQEG Video Quality Experts Group

VQM Video Quality Metric

VR Virtual Reality

Chapter 1

Introduction

1.1 Multimedia Communication: quality assessment

The advancement of multimedia ecosystem, from content production, content delivery, to content rendering, allows users to consume multimedia services on any device, in any place, and at any time. In the Internet era, the multimedia services are becoming a part of our daily life. The quality of provided multimedia service is making the difference in service providers. It is useful to notice that, every involving steps, capturing, processing, communication, presentation and/or storage, introduce artifacts and ultimately degrade the quality of content.

The term "quality" is easy to understand, but difficult to define. Quality is influenced by many factors such as technology and human. Quality is determined by the context: system, human, and business. It is more about individual perception, thus the term perceptual quality, later on Quality of Experience (QoE), is used. To guarantee the defined level of QoE, a quality centric management approach is needed for service providers. In particular, continuous QoE assessment of the provided service is crucial to optimize the multimedia communication system. The scope of this dissertation is to develop QoE assessment theories and methods for multimedia services.

1.2 Motivation

1.2.1 Multimedia content

The use of broadband networks allows multimedia content to be ubiquitously created, stored, transmitted and shared among users with a multitude of devices. Among multimedia services, demand of video services is very high [44]. The size of capturing video content is increasing: to watch the amount of video that will cross global IP networks, each month in 2020, it would take for an individual more than 5 million years, and in every second, nearly a million minutes of video content will

cross the network. IP video traffic will be 82% of all consumer internet traffic by 2020 and it was more than 70% in 2015. Internet video surveillance traffic nearly doubled in 2015 compared to 2014 and it is expected that the traffic will further increase tenfold between 2015 and 2020. Internet video to TV grew 50% in 2015, and will continue to grow at a rapid pace and increasing 3.6 fold by 2020. Consumer Video on Demand (VoD) traffic will nearly double by 2020; and Ultra High Definition (UHD) video will be 20.7% of IP VoD traffic in 2020. Virtual Reality (VR) traffic quadrupled in 2015 compared to 2014, and VR traffic will increase 61 fold by 2020. Among multimedia contents, increasing demand of video traffic is expected to cover 75% of all the contents in Content Delivery Network (CDN) by 2020. Meanwhile, by 2020 mobile traffic is expected to increase by more than 80% with respect to 2010 and by 175% by 2025 with respect to 2020 [190]. These results show the increasing demand of video communication, and thus video has been selected as a multimedia service in this study.

The visual scene is inherently three dimensional (3D) including color, texture, and depth; depth has a significant impact on human visual perception. For the last decades, 3D video technology has developed rapidly for a wide range of applications such as 3D video communication (3DTV) and Free Viewpoint Video (FVV). The 3DTV offers a 3D depth impression of the observed scenery and FVV allows for interactive selection of viewpoint and direction within a certain operating range. The use of multi-camera, and the needs of camera calibration information for synchronization added complexity in these techniques. Meanwhile, the 3D TV technology and FVV technology are continually maturing to provide immerse experience to the end user. However, the consumer acceptance of such a technology is very far from what researchers are expecting. Despite of recent advances in 3D digital technology, including auto-stereoscopic display, for solving the some of the critical human discomfort associated with 3D technology, some intrinsic eye fatigue will always exist with stereoscopic 3D technology. This could be due to the fact that, in general, these technologies rely on the brain processing: to fuse the two disparate images for creating the 3D effect. As a result of such a processing after prolonged viewing hazards such as eye strain, fatigue, and headache are experienced by the general users. Due to the fact that, users are required to focus on the screen plane but to converge their eyes to a point in space in a different plane, and producing an unnatural viewing [12].

To solve these issues, technology is further advanced, and Light Field (LF) camera came into the market: the LF imaging technology is considered a next generation imaging technology. Many applications can benefit from this new technology such as photography, astronomy, robotics, medical imaging, and microscopy. The most appealing applications of light imaging are interactive rendering, where focus, exposure, and depth of field can be adjusted after the picture is acquired. Moreover, the LF camera recorded (angular and spatial information of a scene) can be exploited for a wide range of applications such as 3D rendering, parallax, and VR.

Now, industry and academic researchers believe that, this LF imaging is becoming a driver

technology in the imaging field. This claim is justified by the statement: "*Light-field technology is an industry game changer, enabling new, creative possibilities to meet consumers' expectations for immersive digital cinema and media*", given by Dr. Siegfried Foessel, Fraunhofer Digital Media Alliance.

Moreover, the interest in this technology is witnessed by the efforts of Joint Photographic Experts Group (JPEG) committee that launched a new activity, JPEG PLENO [56]. The aim is to develop a standard framework for the representation and exchange of new imaging modalities such as LF imaging. The MPEG [222] also started the third phase of free viewpoint television, in August 2013, aiming to devise super multi-view (360 degree display), free navigation, and integral photography (LF imaging) for full parallax applications. And thus, in this dissertation, the LF image is also considered as a multimedia content for the study.

1.2.2   Multimedia quality assessment

Every step involving in multimedia ecosystem: capturing, encoding, distribution, decoding, to presentation, produces artifacts and ultimately degrades the quality of the content. Since the following issues repeatedly appear, no matter with the multimedia content type, the knowledge of degraded quality or quality level is crucial.

- How to evaluate the videos generated from our algorithms/systems?

- How do we know our algorithm/system is creating an improvement between the input and output videos, and by how much?

- How can we know one algorithm/system performs better than another, and by how much?

- What should be the quality criterion for which the design of our algorithms/systems should be optimized?

The knowledge of quality level is not only limited to train, test, and benchmark the processing algorithms, also highly important for the adaptation of a new technology from both users and industry point of view. The new trend is shifting the focus of quality assessment of compliance with system design goals to fulfillment of user needs or expectations [225]. Thus, service providers need to devise a strategy for continuous assessment of the perceived quality of the service and for performing system optimization object to automatically provide the required QoE level.

Since, human is the ultimate receivers in most image or video processing systems, the goal of the quality assessment is to evaluate the quality of image or video as perceived by an average human observer. Most reliable Image Quality Assessment (IQA) or Video Quality Assessment (VQA) methods are based on collecting human judgements, measured through Mean Opinion Score (MOS). However, with the exponential increase of the volume of image or video data being generated daily, it becomes impossible to address these quality issues in a timely manner by subjective visual

testing, which is slow, cumbersome and expensive [247]. Moreover, it is non applicable in practical applications, such as dynamically monitoring and adjusting the quality, optimizing algorithms and parameter settings of video communication systems. Therefore, the availability of an objective Image Quality Metric (IQM) or video Quality Metric (VQM) able to give a close approximation of the subjective scores is important.

1.3 Statement of Problem

As mentioned before, this dissertation is dealing with two types of multimedia contents: 2D video and LF image.

1.3.1 Video

The VQA topic is well investigated in the literature. Meanwhile, my ongoing efforts have been given to devise the objective VQM, however, there are many challenges and issues regarding the applicability of the metrics. The state-of-the-art works have the following shortcomings:

- perceptual impact of key transmission impairments on the video communication is not studied;

- impact of video content on the QoE is not studied;

- there are no video quality datasets: test videos and annotated subjective scores, designed in a real video communication environment. The dataset has a particular value to study the perceptual quality of transmission impairments and video content, and to test the performance of the VQM; and

- performance of the state-of-the-art VQM is not good enough, and thus there is a need of a VQM.

1.3.2 Light field image

On the other hand, the rapidly developing LF technology and consumer interest towards this technology is pushing the need for quality evaluation of such contents. To the best of our knowledge, very few works [238] are performed in this direction. In particular,

- LF camera provides a grid of elemental images, in which the content of each elementary image is similar to its neighbors. Therefore, compression of LF content is needed for storage and transmission. The lossy compression may introduce visible artifacts, however, the analysis of the overall perceived quality is not studied;

- perceptual quality of LF images based on their content is not studied;

- performance of existing 2D IQMs, when applied to LF image is not evaluated, and no metric is designed for LF images;

- LF image quality dataset is not available with test images and corresponding subjective quality ratings;

- given the novelty of the imaging system, there is no validated subjective quality assessment procedure for LF image, and thus, there is a need of an IQA framework for LF image;

- securing multimedia data from undesired manipulation is a widely investigated topic in the state-of-the-art. A large number of techniques have been developed for protecting images, videos, and also audio from malicious attacks. The increasing interest towards LF media calls for methods for protecting these data from manipulations and unauthorized reproduction or diffusion.

1.4 Objectives

The research work presented in this dissertation aims to improve the QoE when dealing multimedia content such as video and LF image. The approach is to understand the impact of transmission impairments on QoE and to develop the video and LF IQM. The main objectives are:

concerning video,

1. to investigate the impact of key transmission impairments and video content on QoE;

2. to study the correlation between QoS parameters to QoE by exploiting the state-of-the-art mapping models;

3. to investigate the performance of the state-of-the-art objective VQMs;

4. to develop a No-Reference VQMs to the video communication environment; and

5. to create video quality datasets and make available for research community, it is also needed for the above mentioned objectives.

concerning LF image,

1. to study the theoretical background and LF image communication chain–leading towards the definition of the quality issues in LF imaging;

2. to study the impact of compression artifacts on LF image QoE;

3. to develop a LF IQA framework, since there is no quality assessment framework available for LF image;

4. to create LF image quality datasets and make it available for research community, it is also needed for the above mentioned objectives; and

5. to devise an objective quality assessment metric for LF image;

6. to designed a novel embedding scheme tuned to LF image for copyright protection.

1.5   Scope

The work in this dissertation falls within telecommunication filed, particularly in signal and image processing; and address one of the key problem in multimedia communication: quality assessment. The broad scope of the dissertation is to devise the theories and tools needed for end-to-end quality assessment.

The proposed quality metrics could be used to benchmark the processing algorithms and communication system. The designed video and LF image quality datasets could be used to train, test, and benchmark the new algorithms. Furthermore, adopted LF IQA framework will be a valuable breakthrough for the LF image processing community.

1.6   Contributions

This dissertation is focused to deal with the quality issues of video and LF image. Therefore, the contributions are grouped in two parts: video QoE and LF image QoE.

1.6.1   Video quality of experience

In the first part of this dissertation, video QoE and related issues are presented. In brief,

- a video quality dataset is created by considering key transmission impairments: delay, jitter, packet loss, and bandwidth limitation, and making it freely available for research community;

- the impact of transmission impairments on video QoE is presented;

- effect of video content in the presence of transmission impairments is presented, and the video content and its influence on QoE is studied by using low level video content attributes, such as spatial-temporal perceptual information, video motion, colorfulness information, frame resolution, and HVS characteristics;

- the performance of the state-of-the-art VQMs is evaluated on the proposed video quality dataset, the achieved results evidence the need of a new VQM;

- by exploiting the proposed dataset, QoE estimation models (by using QoS information) is evaluated;

- finally, a No-Reference VQM is proposed to the video communication services.

1.6.2 LF image quality of experience

The second part of this dissertation is focused to deal the issues and challenges arisen with the development of new imaging technology: LF imaging. Briefly, the major contributions are:

- an introduction of LF imaging, together with possible image quality distortion model, and quality related issues and their implications on the QoE evaluation are presented;

- a novel watermark embedding scheme tuned to LF image is presented;

- to study the performance of state-of-the art IQMs, when applied to LF image, and to benchmark the proposed LF IQM, a LF image quality dataset is designed and making it freely available for the research community;

- a Reduced-Reference IQA framework is proposed for LF image;

- a LF image quality dataset is designed by considering the hand held LF camera (Lytro Illum) recorded images;

- finally, a subjective quality assessment framework for LF image is presented.

1.7 Thesis Outline

The organization of this dissertation is shown in Figure 1.1. The thesis is organized in three parts. First part includes the background information to the dissertation. The second part of the dissertation is focused for video QoE, and the quality assessment of LF imaging is included in part three. In brief;

Chapter 2 reviews the theoretical background, including multimedia communication and factors affecting the multimedia quality, definition of quality towards QoE, and QoE management framework for multimedia communication. In particular, this chapter highlight the QoE influencing factors and the need of quality assessment for QoE centric multimedia operation.

Chapter 3: This chapter is focused to develop the theoretical background for assessing the multimedia QoE. In particular, the adopted subjective assessment framework is presented by considering validated subjective and objective quality assessment methods.

Chapter 4: In this chapter, a video quality dataset–with a brief description of the adopted procedure to create the dataset is reported. Moreover, the preliminary results exploited from the dataset are also included.

Chapter 5: This chapter presents the impact of transmission impairments on video QoE. Moreover, the effect of video content on video QoE in the presence of the transmission impairments is

Part-I: Background Study

Chapter 2: Multimedia quality of experience

Chapter 3: Quality of experience assessment

Part –II: Video quality of experience

Chapter 4: A dataset: to study the video quality of experience

Chapter 5: Impact of impairments on quality of experience

Chapter 6: Video quality of experience metric

Part –III: Light field image quality of experience

Chapter 7: Light Field imaging

Chapter 8: Light Field image watermarking

Chapter 9: Quality evaluation of light field images

Chapter 10: Subjective quality assessment of light field images

Chapter 11: Perceptual quality of compressed light field images

Chapter 12: Conclusion

Figure 1.1: An overview of the thesis outline.

presented. Finally, the analysis of the effects of content related low level attributes on video QoE is presented.

Chapter 6: This chapter proposes a No-Reference VQM to video communication services. Analysis of the achieved performance gain over the state-of-the-art metrics is also reported.

Chapter 7: This chapter presents the theoretical background of LF imaging–highlighting the distortions introduced by the steps involving in the LF image communication chain. In particular, this chapter highlight the quality of experience issues arisen with the development of this imaging technology.

Chapter 8: This chapter proposes a watermarking scheme by exploiting the features of LF image. The robustness of the proposed method are evaluated by using subjective and objective measure–the achieved results are also reported.

Chapter 9: A Reduced Reference IQA framework for LF image is presented in this chapter. Moreover, the performance of the proposed metric is evaluated on a newly designed LF image quality dataset. The brief description of the dataset and performance analysis of the proposed method and the state-of-the-art metrics, when applied for LF image, are included.

Chapter 10: In this chapter, a generic subjective quality assessment framework comprises of source sequence selection, distortion model, and selection of the subjective quality assessment protocol, is presented. Following the framework, a LF image quality dataset (SMART) is created. Next,

a LF image processing experimental setup is proposed for the subjective experiment of LF image.

Chapter 11: This chapter includes the preliminary results achieved from the SMART dataset, in particular, the impact of compression artifact in LF image QoE, impact of image scene on LF QoE, and the performance of 2D IQMs, when applied to LF image.

Chapter 12: This chapter presents the concluding remarks of this dissertation.

Chapter 2

Multimedia Quality of Experience

In this chapter, a brief overview of multimedia communication and the factors affecting multimedia quality are given. Then, a review of the terminologies related to multimedia QoE is presented.

As a contribution to the dissertation, this chapter provides a conceptual framework for assessing the QoE of multimedia services. In particular, the needs and uses of multimedia QoE assessment are highlighted. QoE influencing factors followed by a QoE model are presented. The QoE model provides the conceptual framework to devise efficient and effective metrics.

2.1   Key definitions

In this dissertation, some terms related to multimedia quality are detailed:

- *application:* set of activities performed to respond to the needs of the users in a given situation for purposes such as business, education, personal communication or entertainment (ITU-T Rec. F.700 [113]);

- *multimedia:* an adjective and must be attached to a noun which provides the context. For example, multimedia service or application, multimedia terminal, multimedia network and multimedia presentation (ITU-T Recommendation F.700 [113]);

- *multimedia application:* an application that requests the handling of two or more representation media (information types) simultaneously which constitute a common information space. Examples are cooperative document editing, long distance meetings, remote surveillance, medical document remote analysis and teletraining (ITU-T Rec. F.700 [113]);

- *multimedia service:* telecommunication services that handle two or more types of media in a synchronized way from the user's point of view. A multimedia service may involve multiple

parties, multiple connections, and the addition or removal of resources and users within a single communication session (ITU-T Rec. F.700 [113]);

- *effectiveness:* accuracy and completeness with which users achieve specified goals (ISO 9241-11 [27]);

- *efficiency:* resources expended in relation to the accuracy and completeness (ISO 9241-11 [27]);

- *experience:* encounter of a human being with a system, having a defined beginning and end (ETSI TS 103 294 [60]);

- *satisfaction:* freedom from discomfort, and positive attitudes towards the use of the product (ISO 9241-11 [27]);

- *service:* group of functions provided by an organization or by an application to a user through an interface (ITU-T Rec. E.860 [110]);

- *context of use:* users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used (ISO 9241-11 [27]);

- *quality:* outcome of a subjective evaluation process (ETSI TS 103 294 [60]);

- *Quality of Experience (QoE):* degree of delight or annoyance of the user of an application or service (ETSI TS 103 294 [60]);

- *QoE feature:* perceivable, recognized and nameable characteristic of the individual's experience of a service which contributes to its quality (ETSI TS 103 294 [60]).

2.2   Multimedia Communication

In general, multimedia has been used over the decades to indicate any kind of "new media" being manipulated or displayed. Multimedia refers to data or information being transferred over the network(s); may be composed of one or more of the text, audio, image, animation, video, and interactive content.

Specifically, multimedia is any combination of text, graphic art, sound, image, animation, and video delivered to the users by computer or other electronic means, and it is the richly presented sensation [235]. It is the integrated manipulation of continuous or discrete media such as text and graphics; the manipulation refers to the act of capturing, processing, communication, presentation or storage [257].

The multimedia communication indicates all the steps involving from capturing to presentation, as shown in Figure 2.1. In networked multimedia applications, various entities need to cooperate to allow multimedia content to be presented at the sink or user interface with required guarantees
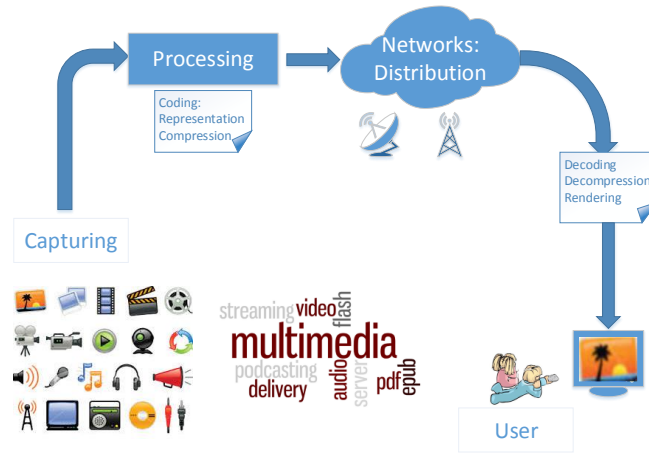
Figure 2.1: Block diagram of multimedia communication–it includes all the steps: capturing, representation, encoding, networking, decoding, and storage/presentation.

or QoE. In the context of communication services, QoE is influenced by content, network, device, application, user expectations and goals, and context of use [131].

2.3   Factors affecting multimedia quality
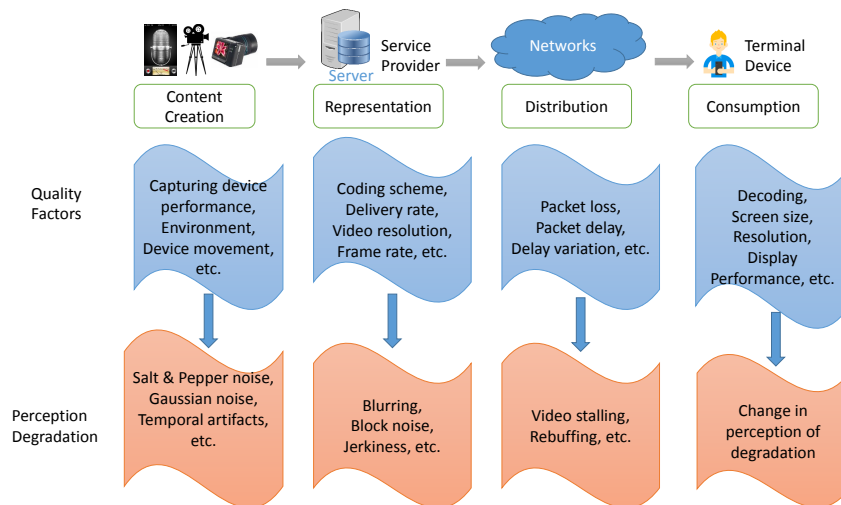
Figure 2.2: Quality degradation factors in video communication.

In multimedia communication many factors may influence the quality: from the quality of multimedia content to the quality of user terminal [88]. As an example, in the context of video communication, major causes of the quality deterioration are presented in Figure 2.2. The quality of multimedia content is depending on acquisition device, process, and environment. Quality of captured video is depending on shooting conditions such as focus, brightness, contrast, etc. and camera performance. Next, captured content undergoes processing steps before transmission. Raw video is encoded; the encoding methods (MPEG2, MPEG4, H.264/AVC, HEVC, etc.) and selected coding parameters (encoding bit rate, spatial resolution, frame rate, etc.) affect the quality. During the transmission of encoded video over a wired or wireless media, network produced impairments such as jitter and packet loss will cause lowering the visual quality. Finally, at the user terminal, decoding, image quality realized by the terminal or display capabilities affect the quality and its perception.

2.4 Perceived Quality: from quality to quality of experience

This section elaborates the concept of QoE for establishing a conceptual basis to this dissertation. As shown in Figure 2.3, QoE mostly deals with the human and contextual factors, whereas the Quality of Service (QoS) is referring more for technical and network aspects. Brief description of the terminologies is presented in following subsections.
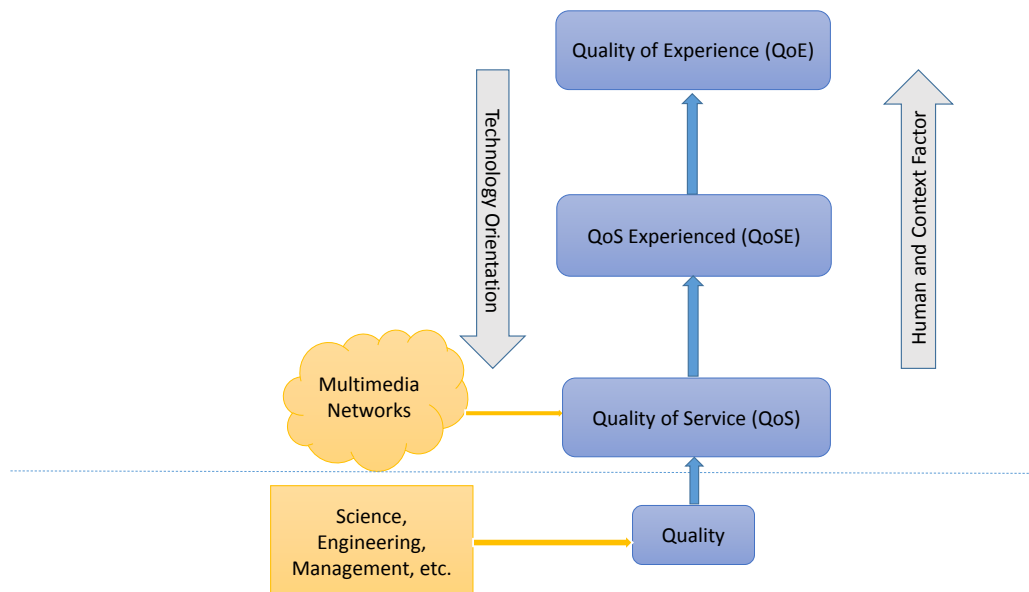
Figure 2.3: Quality to multimedia QoE: conceptual framework for defining the quality towards QoE.

2.4.1 Quality

The Quality is one of *Aristotle's Categories*–places every object of human apprehension under one of ten categories: substance, quantity, **qualification or quality**, relative or relation, where or place, when or time , being-in-a-position, having or state, doing or action, and being affected or affection [10]. Quality is a buzzword used in different sectors, such as business and industry, to suit their specific perceived needs. Some of the definitions for quality are:

- to the field of quality management, ISO 9000 defines quality as the "degree to which a set of *inherent characteristics* fulfills *requirements*" [93];

- ISO 8402 defines quality as "the totality of *characteristics* of an entity that bear on its ability to satisfy *stated and implied needs*" [67];

- in Qualinet white paper [131] quality is expressed as the outcome of an individual's *comparison* and judgement process, and it includes perception, reflection about the perception, and description of the outcome.

These definitions indicate that the quality of something can be assessed by comparing a set of *inherent characteristics* with the set of *requirements*. For example, if characteristics meet the requirements can be considered as high quality otherwise low quality, and thus, the quality can be considered as a relative term.

2.4.2 Quality of service

In telecommunication, term quality is expressed in terms of QoS. Similar to ISO 8402 definition of quality, in ITU-T Rec. E.800 [191] QoS is expressed as "the totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service." This definition of QoS is focused for telecommunication services, and does not cover the user characteristics and context of use. Mostly, it deals with physical and measurable performance parameters of the network or delivery system, and thus QoS is considered as a service providers approach.

Another definition of QoS mentioned in ITU-T Rec. E.800– "the collective effect of service performances, which determines the degree of satisfaction of a user of the service", provides a wide range of applicability of the term QoS. The use of QoS is not limited to, telecommunications also in broadband, wireless, and multimedia system; particularly in IP related systems.

ITU-T Rec. G.1000 [194] emphasizes the consumer perception of QoS: QoS is viewed from customers and service provider's viewpoint, as shown in Figure 2.4. QoS is defined from four different prospectives: customers QoS requirements, service provider offerings of QoS (or planned/targeted QoS), QoS achieved or delivered, and customer survey ratings of QoS. The level of quality required
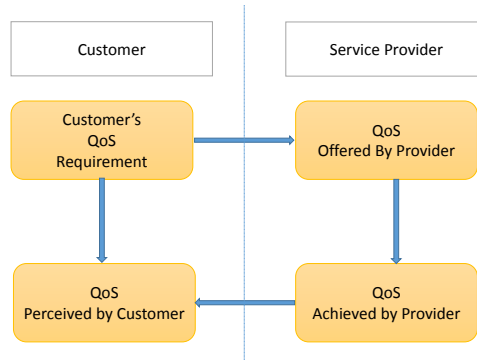
Figure 2.4: ITU-T Rec. G.1000: The four viewpoints of QoS.

for a particular service (which may be expressed in non-technical language) is customers QoS requirements, and it has a particular value to the service provider for planning the network service level. The level of quality expected to be offered to customers by the service provider is QoS offered by the service provider, and level of quality is expressed in terms of values assigned to QoS parameters.

QoS can be assessed in two ways: 1) objective measurements of physical attributes of a circuit, system, network, etc. and 2) subjective measurement by customer surveys. From the system design point of view, objective measure of QoS is crucial, thus many QoS parameters have been exploited. The QoS parameters are different to the services: fixed telephony service (service delivery time, call success rate, billing accuracy, etc.) mobile telephony service (call success rate, blocked calls, dropped calls, etc.), and data network and internet service (bandwidth, latency, BER, jitter and jitter variations, throughput, etc.) In today's converged (all-IP) network, the internet service related QoS parameters are more relevant, thus the parameters are considered as key QoS parameters.

2.4.3 Quality of service experienced

The concept of QoS experienced (QoSE) is introduced in ITU-T Rec. E.800; defined as the level of quality that customers/users believe they have experienced. The QoSE comprises two components: quantitative and qualitative. Quantitative component is influenced by the complete end-to-end system effects (network infrastructure) and qualitative component can be influenced by user expectations, ambient conditions, psychological factors, application context, etc. The level of QoSE can be expressed as opinion ratings. The definition of QoSE is leading to the concept of QoE.

2.4.4   Quality of experience

QoE is the overall performance of an end-to-end networked system from the user's perspective. In particular, QoE extends the concept of QoS by encompassing additional factors: source quality, encoding quality, network performance (QoS), decoding and error correction, and presentation quality, affect the users' perception of multimedia presentation quality [179].

ITU-T Rec. P.10/G.100 [195] defines QoE as "the overall acceptability of an application or service, as perceived subjectively by the end-user". The acceptability is the outcome of a decision which is partially based on the QoE [64]. QoE includes complete end-to-end system effects (client, terminal, network, services infrastructure, etc.) and overall acceptability is influenced by user expectations and context. The QoE implicitly promises the individual engagement. The experience is an individual stream of perception and interpretation of one or multiple events. The event is an observable occurrence and it is determined in space (i.e. where it occurs), time (i.e. when it occurs), and character (i.e. what can be observed) [131].

Qualinet [131] defines QoE as the "degree of delight or annoyance to the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state". The application is a software and/or hardware that enables usage and interaction by a user for a given purpose. Such purpose may include entertainment or information retrieval, or other, and service is an episode in which an entity takes the responsibility that something desirable happens on the behalf of another entity.

The concept of QoE, introduce the user centric quality management approach. The term experience promises individual agreement, and term QoE involves aspects related to not only subjective perception, but also user behavior and needs, appropriateness, context, usability and human factors of the delivered content [55]. QoE Influence Factor (IF) is any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the QoE for the user [60]:

- human IFs are any variant or invariant property or characteristic of a human user. The characteristic can describe the demographic and socioeconomic background, the physical and mental constitution, or the users emotional state;

- system IFs refer to properties and characteristics that determine the technically produced quality of an application or service. They are related to system performance, and in the context of multimedia services, media capture, coding, transmission, storage, rendering, and reproduction/display, as well as to the communication of information between the service and the user;

- context IFs are the factors that embrace any situational property to describe the user's environment in terms of physical, temporal, social, economic, task, and technical characteristics of

devices.

Most common QoE IFs are presented in Figure 2.5 and briefly reported in the following:
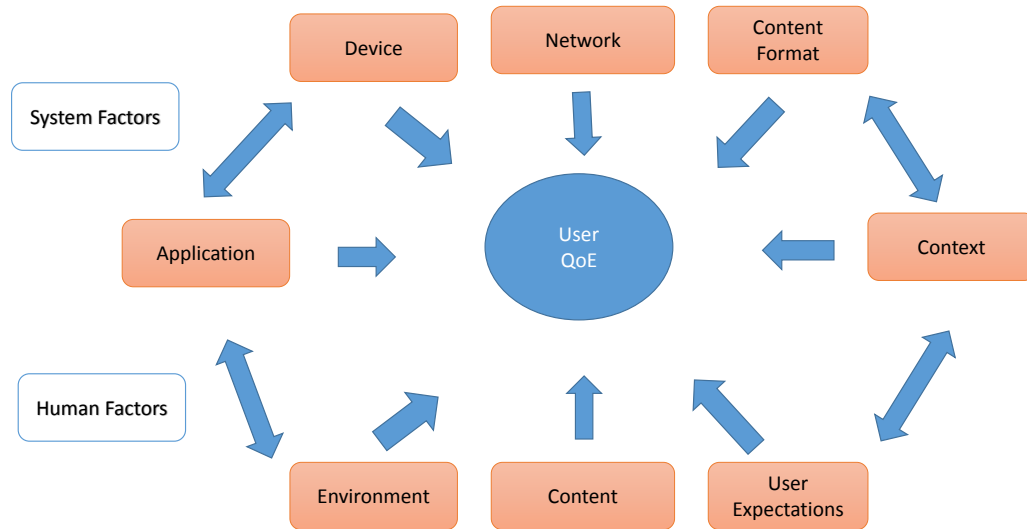
Figure 2.5: Factors influencing QoE [43].

Content

As mentioned before, multimedia communication covers a wide range of content types, and in this contribution mainly images and videos are considered. First, source content quality is an important factor for better QoE. Next, visual content characteristics also influence the QoE, since, it is a major source of information [52]. At the semantic level, the content may be represented by high level features, such as indoor, outdoor, sports, movie, and even different types of movies such as action or documentaries etc. In the video, the content features such as bit rate, frame rate, resolution, 2D/3D, spatial information, and temporal information i.e. motion have a direct influence on QoE. Meanwhile, different multimedia devices and applications should support different content characteristics for a smooth playback.

Device

Quality of captured visual content is crucial for better QoE, and it can be achieved with the proper selection of acquisition device and by following defined acquisition procedure. On the other hand, visual content is consumed through different devices such as televisions, personal computers, laptops,

tablets, and smart phones; the consumption devices comprise different features (screen resolution, viewing angle and distance, and sound quality), and thus the consuming device also influence the QoE.

Application

QoE could be differ based on usability and interactivity of an application. The usability rating is higher when the service is easier to use [26]. In particular, multimedia applications should be self-intuitive and easy to use. In the context of video communication, most applications have different buffering scheme, encoding, and decoding, which affect playback of video and the overall video QoE. Moreover, the applications are varied based on the used device and environment.

Environment

Environment properties such physical (where), temporal (when), character (what), social (with whom), technical (iterativeness), and economic (cost), also influence the overall user perception towards the content. For an example, watching a football match at home, or in a bar, and at morning, afternoon, and evening, results different perception about the content. As presented in [208], results obtained from controlled experiments (in lab) and crowdsourcing experiment are different. To produce the reliable and reproducible results (QoE) controlled environment is needed.

Network

Network mostly reflect to a distribution channel of multimedia communication system. The network can be a wired (copper cable and fiber) or wireless. Usually, network condition is expressed in terms of QoS parameters, such as Packet Loss Rate (PLR), delay, jitter, and bandwidth. The results presented in [175] show that the video QoE is influenced by QoS parameters.

Content Format

Usually, we encode the visual content for making it compatible to the delivering over the networks. There are different encoding formats and each of them have different features and capability. For an example, the video signal can be encoded by Moving Picture Experts Group ((MPEG), H.264 or MPEG-4 Advanced Video Coding (MPEG-4 AVC), and H.265/High Efficiency Video Coding (HEVC) to distribute over the network. The HEVC encoding provides high QoE compare to other methods for the same level compression ratio [199].

User Expectations

Individual user has his or her own requirement or expectations on the quality. The user profiles, such as age, sex, interest, skills, frame of mind, experience, etc., to determine the requirement of

the quality. User attitude and expectation play a vital role in determining the QoE [131]. For an example, perception towards the horror movies could be different from the persons of different ages: child, young, and old. Thus, different opinion scores are given for the same test image or video during subjective experiment. To express the average behavior of opinion scores and to represent differences in the opinion scores given to the test material, the MOS is plotted with corresponding confidence interval.

Context

Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves [9]. The contextual factors are strongly related to human and system factors, and have a significant influence on QoE, and are very difficult to consider during the quality evaluation process. From the user's point of view, context influence factors could be the environmental properties, service properties, economic properties, and social properties. In particular, location, temperature, heart rate, eye movement, the amount of sweat, social context, people nearby, light, background noise, age, gender, etc. are considered as the user and environment related contextual parameters. On the other hand, screen size, design, layout, resolution, button placement, input/output methods, appeal, usability, etc. are the system related contextual parameters [157].

Mostly, image and/or video (analog and digital) with textual information is expressed as multimedia. For the content, two of our human senses: sight and hearing, are primarily engaged. 60% of human communication is nonverbal and most of us perceive the world through the combination of five senses: sight, hearing, touch, taste, and smell. In this situation, multimedia experiences fail to convey the sensation, and this concept of *Mulsemedia* – multiple sensorial media, engage three (or more) of our senses, is presented in [74]. Typically, the QoE assessment is based on vision and hearing, and thus *Quality of sensory experience* for mulsemedia is defined, and concept is further elaborated as *Quality of Life (QoL)*– as the general well-being of individuals and societies in [225].

2.5   Theoretical quality of experience model

Section 2.4 presents the theoretical background of QoE, which is helpful to understand the factors that influence the QoE. Whereas, this section presents an intuitive and systematic way to identify the factors; the factors are categorized into four multi-dimensional spaces (shown in Figure 2.6):

- Application space (A): Application space represents the application or/and service configuration factors (media encoding, resolution, sample rate, frame rate, buffer sizes, etc.) and content related factors (temporal or spatial information, 2D/3D content, color depth, etc.). Therefore, the space is a multi-dimensional; $A = (a_1, a_2, a_3, ..., a_y)$.
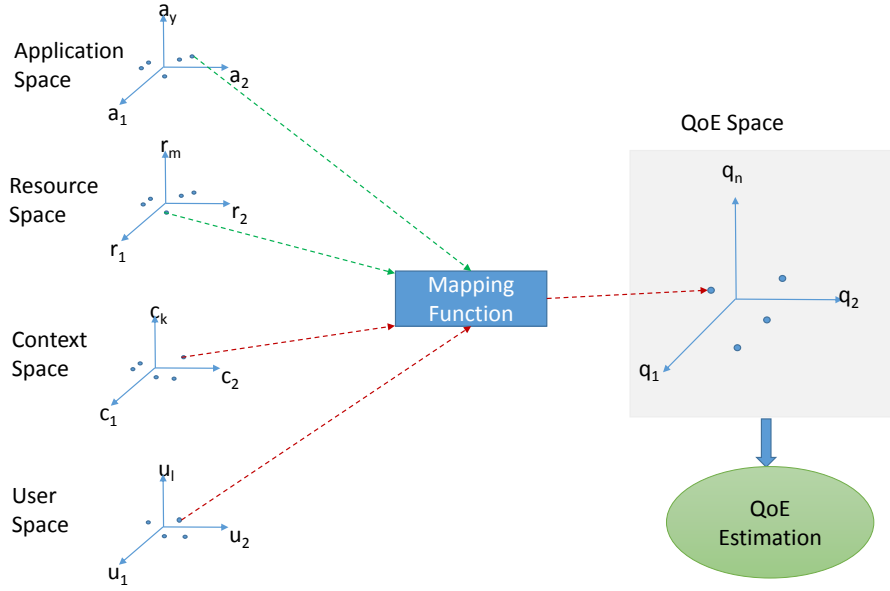
Figure 2.6: QoE modeling: $ARCU$ model.

- Resource space (R): This space composed of dimensions representing the characteristics and performance of the technical system(s) and network resources used to deliver the service such as network QoS parameters (delay, jitter, loss, error rate, and throughput) and system resources (server processing capabilities and end user device capabilities– computational power, memory, screen resolution, user interface, battery lifetime, etc.). $R = (r_1, r_2, r_3, ..., r_m)$.

- Context space (C): The C composed of dimensions indicating the situation in which a service or application is being used such as ambient conditions (lighting conditions, noise, etc.), user location, time, task/purpose, economic context, and service level agreements. $C = (c_1, c_2, c_3, ..., c_k)$.

- User space (U): The U composed of dimensions related to the specific user of a given service or application such as user's demographic data, user preferences, requirements, expectations, prior knowledge, mood, and motivation. $U = (u_1, u_2, u_3, ..., u_l)$.

This model is first presented in [214], and is adopted by ETSI in [60]. The dimensions in each space may have different scales, thus for convenience, the model is expressed as the direct sum of these spaces (Equation 2.1).

$$ARCU = A \oplus R \oplus C \oplus U. \tag{2.1}$$

In an initial point, the mapping function can be envisioned as having the form $Q : ARCU \rightarrow Q$, where $Q = \mathbb{R}^n$ and n is the number of dimensions in QoE space. Mathematically, if the QoE space is composed of n dimensions, $Q = (q_1, q_2, q_3, ..., q_n)$ with corresponding assigned weight factors $W = (w_1, w_2, w_3, ..., w_n)$, then $QoE_{integral} = f(w1q, w2q2, ...wnqn)$. This model provides a theoretical approach for modeling the multimedia QoE.

2.6 Quality of experience management for multimedia services

In the context of emerging multimedia services, a user-centric approach need to be monitored to design the whole process of multimedia system: content production, content delivery, service activation, content consumption, service management, and updating [18]. Indeed, the quality of the user experience, is the determining factor for success or failure of the novel multimedia services.
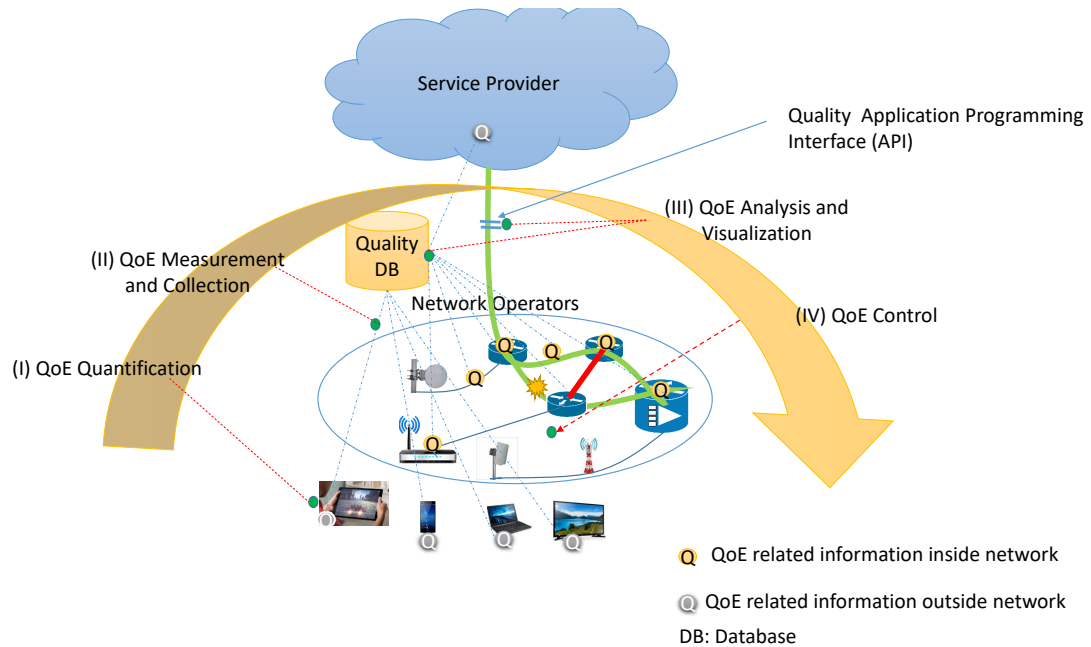
Figure 2.7: Overview of QoE centric operation for multimedia applications.

To guarantee a defined level of QoE for the user, a QoE centric operation need to be devised. The QoE centric operation for multimedia application can be accomplished in four steps: QoE quantification, QoE measurement and collection, QoE analysis and visualization, and QoE control [88], as shown in Figure 2.7.

- QoE quantification: This step clarifies service quality criteria and service management guidelines by elucidating human and cognitive characteristics with respect to the quality of multimedia services. Particularly, it establishes a methodology to estimate QoE by using the quality-related information from the network, servers, and terminals. It clarifies the relationship between QoE and network quality measures (packet loss/delay, delay fluctuation, etc.) and application quality measures (sound quality, picture quality, response time, etc.).

- QoE measurement and collection: Since, QoE is affected by inside and outside network quality factors, information of the service, use environment, and terminal conditions are crucial in order to identify and deal with QoE degradation on a user-by-user basis. This stage efficiently measure and collect the quantified QoE information.

- QoE analysis and visualization: For analysis and visualization of collecting data from inside and outside the network, big data analysis techniques such as machine learning and deep neural network can be exploited. The resulting information includes the degree of network congestion, state of service provision, quality degradation, etc.

- QoE control: Finally, necessary actions are taken when detecting QoE degradation with the aim of improving QoE or avoiding such degradation. It uses the analyzed and visualized QoE information to optimize QoE according to, network conditions by allocating network resources appropriately.

In the presented QoE centric operation, QoE measurement is most important and basic operation.

Chapter 3

Quality of experience assessment

This chapter presents the quality assessment methods: subjective and objective. As the contributions to the dissertation, this chapter provides: i) a generic subjective quality assessment framework constituting the required steps to be followed such as source content selection criteria, content characterization attributes, and subjective quality assessment methods; and ii) the need of an objective quality metric is highlighted, and the basic approaches of designing the objective metrics are reported.

3.1 Introduction

QoE is a multi-modal and multi-dimensional problem; it is difficult to quantify and measure. However, for the design of QoE centric multimedia communication system, and for service providers, it is important to quantify and measure the QoE in the statistical and interpretable values.

As shown in Figure 3.1, there are subjective and objective methods for assessing the QoE. In general, ultimate user of the multimedia content is human, thus the goal of the quality assessment is to evaluate the quality of the visual content as perceived by an average human observer. QoE is significantly influenced by human factors, thus more accurate and reliable methods for evaluating the multimedia QoE are based on collecting human judgements: *subjective method*. The subjective method is based on surveys and interviews. In general, *qualitative* and *quantitative* techniques are used to evaluate the QoE in subjective method. The qualitative technique uses behavior, verbal words, and observations to record the users opinion about the QoE; again the issue is quantification of QoE. Whereas, quantitative technique use surveys and user studies to record the human perceptions, feeling, and cognition in the forms of quantifiable data such as numbers and levels. In this dissertation, the quantitative technique is used for the subjective QoE assessment.

In subjective QoE assessment, visual content to be evaluated is shown to the group of people, i.e. the human observers ask for their opinion to the test material. The opinion scores collected
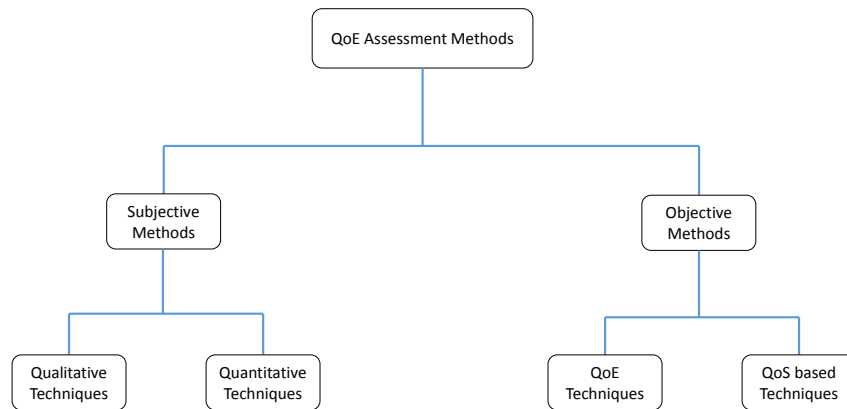
Figure 3.1: Quality of experience assessment methods.

from subjects for the test material are used for a wide range of applications. For example, opinion scores are used to understand the response of HVS to visual signal; the response/features can be used for optimizing the signal processing algorithms. However, the subjective quality assessment technique is time consuming and not suitable in practical applications. Therefore, a common goal is to devise an objective quality metric. From another viewpoint, to design, train, test, and benchmark the objective quality metric, test material and annotated subjective quality scores are important.

An objective method, the QoE is estimated by analyzing features of content to be evaluated. As presented before, in Chapter 2, basic definitions of the quality give the idea for assessing the quality–the quality of something can be assessed by comparing a set of inherent characteristics with the set of requirements. In literature, the objective QoE is devised based on two techniques: *QoE techniques*– by exploiting the system and human factors such as signal features, human visual system, and perceptual models, and *QoS based techniques*–by using mapping models, to estimate QoE from QoS parameters (e.g. by exploiting a logarithmic relationship between the change in perception and change in physical stimulus). Based on the availability of reference signal for QoE estimation, the objective metrics are broadly classified in Full Reference, Reduced Reference, and No Reference; these are briefly described in Section 3.3.

3.2 Subjective quality assessment

In this dissertation, many subjective experiments were scheduled for collecting the user opinion scores to the test materials. In literature, many subjective quality assessment methods are available, among them the mostly used and relevant to this dissertation are reported. The selection of a particular

protocol always depends on the purpose of the study and test materials. Therefore, a generic overview of the adopted subjective quality assessment protocols are briefly reported in the following.

3.2.1   Source scene selection

The proper selection Source Sequences (SRCs) is a key operation for obtaining reliable results; and it should be based on video characteristics and on the purpose of the experiment [180]. In particular, image features such as spatial information, color information, and brightness are important parameters. By using these parameters, it is possible to quantify the distortions suffered by data compression or transmission over a bandwidth limited channel. Therefore, selected SRCs should span a wide range of content features.

In the literature, several efforts [178, 120, 255, 61] have been made for image content classification. In brief, many *low level* image features such as contrast, brightness, and edges can be used for image classification. The use of many low level features is resulting the large number of image content clusters [178]. At the same time, the classification of image content based on *high level* (indoor, outdoor, etc.) features are complex, due to the fact that high level features can be considered as a combination of low level ones, and thus, there is no standard procedure for image content definition [120].

From another viewpoint, it is not easy to map the extractable video features such as color, texture, shape, structure, layout, and motion, into semantic concepts, high level features such as indoor and outdoor, people, documentary, news reader, or car-racing scenes, though the effective way for scene content analysis is to use the attributes extractable from the sources [52].

In this context, in [255] the authors use spatial, temporal, and colorfulness information for image and video content analysis, and a survey of available image quality datasets is presented based on these features. Similarly, image content is explained with the help of color, texture, shape, position, and dominant edges of image objects and regions in [61]. In this dissertation, key quality attributes spatial-temporal perceptual information, colorfulness, contrast, correlation, homogeneity, brightness, hue, and saturation are used to characterize and select the reference images or image scenes. The features were considered based on the image quality attributes and HVS characteristics [61].

Spatial perceptual information

As a perceptual indicator of the spatial information of the scene, Spatial perceptual Information (SI) is used. The SI measures the amount of spatial details of each frame and it is higher for spatially complex scenes. SI measurement is based on the Sobel filter. According to this, each video frame at time $n(Fn)$ is first filtered with the Sobel filter $Sobel(Fn)$ and standard deviation over the pixels std_{space} in each Sobel-filtered frame is then computed and applied for each frame. The maximum value in the time series $\max_{time}$ is chosen to represent the spatial information of the scene. The SI is computed by Equation 3.1,
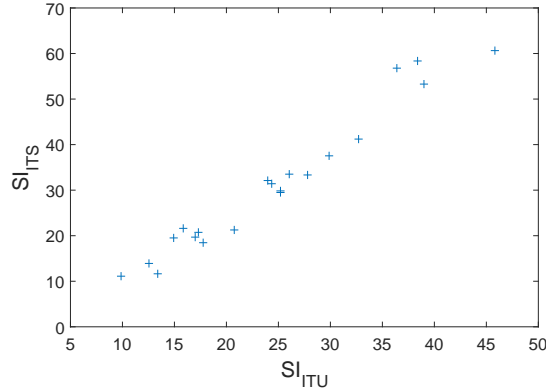
Figure 3.2: Relationship between the ITU and ITS recommended SI.

$$SI = \max_{time}\{std_{space}[Sobel(F_n)]\}. \tag{3.1}$$

The SI metric, Equation 3.1, (proposed for video) can be adopted for image: considering image as a frame of the video. The luminance, Y, of the image is first filtered by using a *Sobel* filter. The standard deviation over the pixels in each filtered image is then computed as an SI.

$$SI_{ITU} = \sigma_{space}[Y_{Sobel}], \tag{3.2}$$

where, SI_{ITU} is spatial information metric, σ_{space} is the standard deviation over the pixels, and Y_{Sobel} is the *Sobel* filtered luminance plane of the image.

Moreover, a spatial information filter (SI_{ITS}) is also proposed by the Institute for Telecommunication Science (ITS) to estimate the image spatial information. The filter is similar to the classical Sobel filter, where separate horizontal and vertical filters are applied, then the total edge energy is computed as the Euclidean distance [97] .

Both SI_{ITU} and SI_{ITS} were tested for the SMART LF dataset [176]. The achieved results show that, for all-focused 2D images the correlation between two metrics is around 98% (shown in Figure 3.2). Therefore, in this dissertation, to estimate the spatial perceptual information of image the SI_{ITU} is used.

Temporal perceptual information

Temporal perceptual Information (TI) indicates the amount of temporal changes of a video sequence and it is higher for high motion sequences. Its measurement is based on motion difference feature, $M_n(i,j)$, which is the difference between the pixel values (of the luminance plane) at the same location in space but at successive times or frames. $M_n(i,j)$ as a function of time is defined as

(Equation 3.3):

$$M_n(i,j) = F_n(i,j)F_n 1(i,j), \tag{3.3}$$

where, $F_n(i,j)$ is the pixel at the ith row and j^{th} column of nth frame in time. The measure of TI is computed as the maximum over time (max_{time}) of the standard deviation over space (std_{space}) of $M_n(i,j)$ over all i and j. Accordingly TI is computed by Equation 3.4:

$$TI = max_{time}std_{space}[M_n(i,j)]. \tag{3.4}$$

Colorfulness

Colorfulness (CF) is the main perceptual attribute underlying the video perceptual quality and the naturalness of the signal. As a perceptual indicator of the variety and intensity of colors in the image the colorfulness metric is used [87]. The CF is computed by Equation (3.5),

$$M_{CF} = \sigma_{rgyb} + 0.3\mu_{rgyb}, \tag{3.5}$$

where, M_{CF} is colorful metric, $\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}$, $\mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2}$, $rg = R - G$, and $yb = 0.5(R+G) - B$, σ is the standard deviation and μ is the mean value. The R, G, and B are the red, green and blue colour channels of the image pixels. For video, an average value of the frame colorfulness is used.

Contrast

Contrast is one of the most important parameters that has been used to evaluate the perceptual quality [250]. This is because the meaningful visual information is conveyed by the contrast. For example, a largely uniform picture carries little or no information [142]. The contrast for each frame is computed by using Gray Level Co-occurrence Matrix (GLCM). The average value of frame contrast is used as an indicator for video.

In brief, the GLCM is used to express how often the different combinations of pixel brightness values occur in an image. It is created from a grayscale image, and calculates how often a pixel with gray level (tone) value i occurs either horizontally (0^0), vertically (90^0), or diagonally (bottom left to top right, -45^0; top left to bottom right, -135^0) adjacent pixels with the value j. GLCM is a matrix where the number of rows and columns is equal to the number of gray levels (G) in the image. The matrix element $P(i,j|\Delta x, \Delta y)$, occurs within a given neighborhood, one with intensity i and other with intensity j. In another viewpoint, the matrix element $P(i,j|d,\theta)$ contains second order statistical probability values for the change between gray levels i and j at a particular displacement distance d and at a particular angle θ [13]. In [86] set of 28 textural features are extracted from GLCM. Some of them are:

$$Energy = \sum_i \sum_j g_{ij}^2, \tag{3.6}$$

$$Contrast = \sum_i \sum_j (i - j)^2 g_{ij}, \tag{3.7}$$

$$Homogenity = \sum_i \sum_j \frac{1}{1 + (i - j)^2} g_{ij}, \tag{3.8}$$

$$Entropy = -\sum_i \sum_j g_{ij} \log_2 g_{ij}, \tag{3.9}$$

where, $g_{ij} = (i, j)^{th}$ element of $GLCM$.

Hue, saturation, and color value

Hue (H), saturation (S), and brightness are major attributes that have been used to express the colorfulness of video [213]. Hue is the most obvious characteristic of a color and it indicates a specific wavelength. The saturation can be expressed as the purity of a color, i.e. high saturation colors look rich and full, and low saturation colors look dull and grayish. The color value (V) is the lightness or darkness of a color. The average scores of each component are used for frames, and finally mean value of the frames is used to express the hue, saturation, and color value for video. Given R, G, and B (each on domain [0 1]), the equivalent H, S, and V (each on domain [0 1]) is computed as [215, 68]:

$$V = max(R, G, B); \tag{3.10}$$

$$S = \frac{V - X}{V}, \tag{3.11}$$

where, $X = \min(R, G, B)$. Let, $r = \frac{V-R}{V-X}$; $g = \frac{V-G}{V-X}$; $b = \frac{V-B}{V-X}$;

- If $R = V$ then $H =$(if $G = X$ then $5 + b$ else $1 - g$);

- If $G = V$ then $H =$ (if $B = X$ then $1 + r$ else $3 - b$);

- else $H =$ (if $R = X$ then $3 + g$ else $5 - r$);

$$H = \frac{H}{6}. \tag{3.12}$$

Brightness

The HVS is more sensitive to luminance (brightness) than color. Brightness is used to describe the intensity of the color, and thus indicating the amount of light. The average score of the luminance

component (Y) of the frame is computed, and mean value of the frames is used to express the brightness of the video.

Generally, in RGB color space all the three color channels are equally important. However, it is also possible to represent a color image efficiently by separating the luminance from the color information. Therefore, YCbCr color space is popular to represent the color image efficiently. The luminance component (Y) cab be calculated as a weighted average of R, G, and B.

$$Y = k_r R + k_g G + k_b B, \tag{3.13}$$

where $k = \{k_i | i = r, g, b\}$ are the weight factors. As mentioned in ITU-R BT.709 [196], $k_r = 0.2126$, $k_g = 0.7152$ and $k_b = 0.0722$. Moreover, the color information is representation as:

$$Cb = B - Y; Cr = R - Y; Cg = G - Y. \tag{3.14}$$

Content selection procedure

The applied scene selection procedure is as follows:

- the low level image quality attributes are computed;

- scene with high value of the attributes are selected–as a result the primary list of possible SRCs is produced; and

- from the primary list, the SRCs are selected keeping in the mind that the final list of SRCs must cover the variety of the content category (high level features).

3.2.2 Number of source sequences

Having a large number of SRCs is preferable in order to obtain a complete understanding of the phenomena under investigation. However, in some applications, in particular subjective quality assessment and pilot test of processing algorithms, the number of SRCs is limited by processing time, duration of the experiment, and available number of the subjects for the experiment.

As specified in ISO 20462 standard [121], to get relative quality values in Just Noticeable Differences (JNDs), the selected attributes should appear in at least three images. As consequence, the number of SRCs considered in this work is always grater than three.

3.2.3 Subjective quality assessment method

In literature, many ongoing efforts have been given to develop the subjective quality assessment protocols for different media: voice, image, and video. Similarly, ITU produced many recommendations depending on services and system under study (some of them are reported in Table 3.1). In particular, the available assessment methods have different strengths and weakness, and the performance

Recommendation	Title	Scope
ITU-R BT.500 [109]	Methodology for the subjective assessment of the quality of television pictures	Broadcast TV signals image and video
ITU-T P.800 [192]	Methods for subjective determination of transmission quality	transmission quality
ITU-T P.910 [112]	Subjective VQA methods for multimedia applications	video-telephony, video-conferencing, and video-on-demand
ITU-T P.911 [203]	Subjective audiovisual quality assessment methods for multimedia applications	audio and video
ITU-R BT.1438 [33]	Subjective assessment of stereoscopic television pictures	3D image and video
ITU-R BT.2021 [231]	Subjective methods for the assessment of stereoscopic 3DTV systems	3DTV and QoE factors: depth perception, visual comfort, etc.
ITU-R BT.1129/ BT.710 [230]	Subjective assessment of standard definition digital television (SDTV/HDTV)systems	SDTV and HDTV Video
ITU-R BT.1210 [229]/ BT.802-1 [101]	Test materials to be used in subjective assessment	Test material

Table 3.1: ITU Recommendations used in this dissertation for image and VQA.

of a particular method is depending on the content type and the artifacts under investigation. Subjective quality assessment techniques are broadly categorized in single stimulus and double stimulus. In single stimulus technique, test signals are shown to the subjects one at a time for evaluating the quality. Whereas, pairs of test sequences are shown to the subjects at a time in double stimulus technique. To reduce the impact of the pre-introduced artifacts, the double stimulus paradigm is an appropriate choice. Based on the purpose of study, image resolution, number of test sequence, and available number of subjects for the experiment, the particular subjective quality assessment technique is selected.

Absolute category rating

Absolute Category Rating (ACR) method is defined in ITU-T Rec. P.800 and ITU-T Rec. P.910, it is a single stimulus rating method, where the subject is presented once with the stimuli, then asked to rate the stimuli on a discrete five point scale, as shown in Table 3.2. If needed, more discriminative power, the ITU-T P.910 also allows the possibility to use a nine-level scale, eleven-level scale, and continuous scale. In this method, the test sequences are presented one at a time and are rated independently on a categorical scale. For quality rating, after each presentation the subjects are asked to evaluate the quality of the sequence shown. The time pattern for stimulus presentation and voting is shown in Figure 3.3.

Another most important feature provided by ITU-T Rec. P.910 is ACR with hidden reference (ACR-HR). The ACR-HR is developed to eliminate the effect of differences in the content of an

5 Level	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 3.2: ACR quality rating scale.
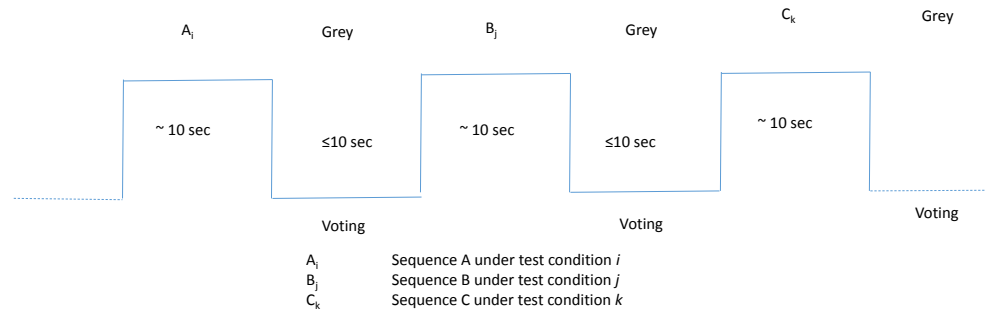
Figure 3.3: Stimuli presentation in ACR method.

image/video used for the evaluation. In this technique, the reference is also presented one at a time and are rated independently on a categorical scale: it is a hidden reference condition. During the data analysis, a differential quality score (DMOS) is computed between each test sequence and its corresponding (hidden) reference. This technique has a particular value, to remove the impact of pre-introduced artifacts and content in the perceived quality.

ACR is mostly used subjective quality assessment method for image and videos. VQEG is using the ACR method to validate the objective metrics [78, 79]. Results presented in [119] show that among the subjective quality assessment methods, DSCQS, ACR, DCR, and ACR-HR, in terms of stability and assessment time the ACR method is most effective for video. The ARC method is an appropriate choice, if test image/video posses a wide range of quality, low to high. In this dissertation, ACR method is adopted for evaluating the QoE of video.

Degradation category rating

Degradation Category Rating (DCR) method is defined in ITU-T Rec. P.910 and ITU-T Rec. P.800, also appears in ITU-R Rec. BT.500-3 under the name Double Stimulus Impairment Scale (DSIS). DCR presents stimuli to subjects in pairs, the first stimulus presented in each pair is always the source reference, without any impairments and second one is the same source but impaired by the

5 Level	Quality
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table 3.3: DCR quality scale.

test conditions. The subjects are asked to rate the impairment of the second stimulus with respect to the reference. The time pattern for stimulus presentation and voting is shown in Figure 3.4. The ITU-T Rec. P.910 and ITU-R Rec. BT.500 use the scales as presented in Table 3.3. Moreover, it is also possible to use any of the scales provided by ACR such as five scales, nine scales, and/or continuous scale.
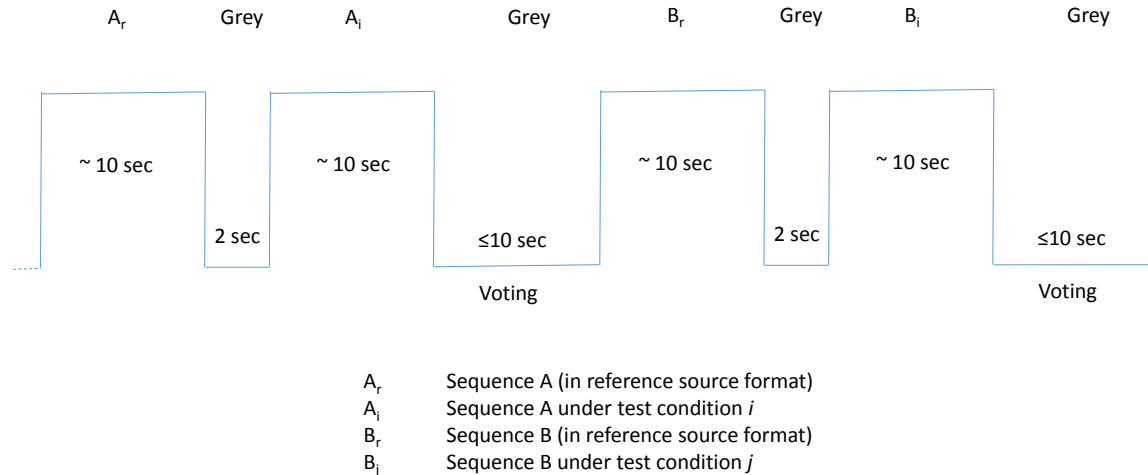
Figure 3.4: Stimuli presentation in DCR method.

For reduced picture formats such as CIF, QCIF, and SIF, the reference and the test sequence can be simultaneously displayed on the same monitor. In this arrangement, the reference should be placed always on the same side (e.g., left), and the subjects must be aware of the relative positions of the reference and test conditions. The advantages of using this technique are: i) it reduces the duration of the test, ii) it is easier for the subjects to evaluate the differences between the stimuli, and iii) under the same test conditions the number of presentations is halved; the attention of the subjects is higher. However, it requires particular precautions, e.g. two sequences must be perfectly synchronized (both must start and stop at the same frame and that the displaying must

be synchronized), in order to allow the subjects to avoid bias due to the way of presentation.

DCR method with simultaneous presentation has a particular value to assess quality of LF images because the size of the LF image (aperture image/view) is considerably small (434×625 for Lytro Illum image).

Pair comparison

The Pair Comparison (PC) method is presented in ITU-T Rec. P.910. In this method the test sequences are presented in pairs (as shown in Figure 3.5), consisting of the same sequence being presented first through one system under test and then through another system. The systems under tests (A, B, C, etc.) are generally combined in all the possible $n(n-1)$ combinations(AB, BA, CA, etc.) for a n number of sequences. Moreover, all the pairs of sequences should be displayed in both the possible orders (e.g., AB, BA). After each pair, a judgment is made on which element in a pair is preferred in the context of the test scenario, i.e. observer selects the one that has better image quality. The result of the PC experiment is a PC Matrix (PCM), that contains the number of times that each option was preferred over the other option. This method has a particular advantage for reduced resolution image; the pair of the test sequences can be displayed simultaneously on the same monitor.
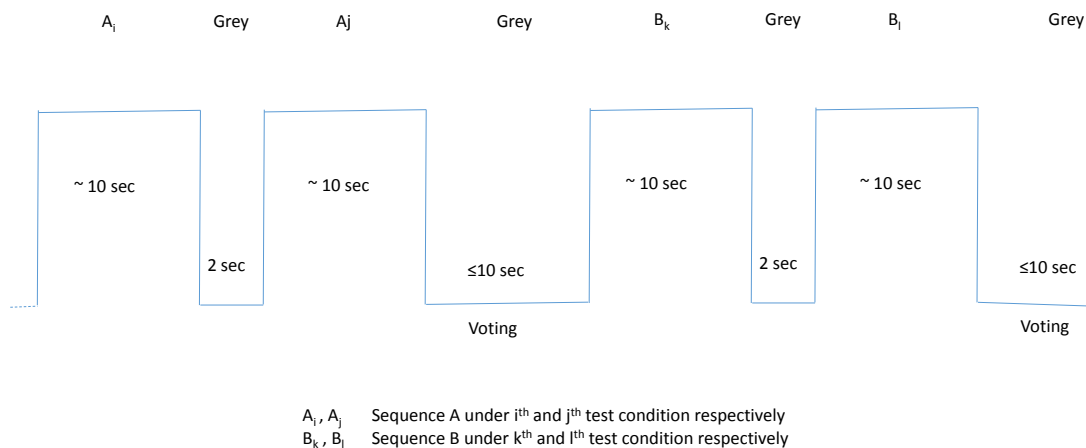
Figure 3.5: Stimuli presentation in PC method.

The main advantage of the PC is its high discriminatory power, which is of particular value when several test items are nearly equal in quality. Moreover, the forced-choice PC method was found to be the most accurate among other widely used methods: single stimulus, double stimuli, and pairwise similarity judgment method [152]. In this dissertation, to evaluate the quality of compressed LF images the PC method is used. However, repeated presentation of the same test image limits the use of this technique, if number of test conditions are large.

| A_r | Grey | A_i | Grey | B_r | Grey | B_i | Grey |

~ 10 sec   ~ 10 sec   ~ 10 sec   ~ 10 sec

2 sec   ≤10 sec   2 sec   ≤10 sec

Voting   Voting

A_r   Sequence A (in reference source format)
A_i   Sequence A under test condition *i*
B_r   Sequence B (in reference source format)
B_j   Sequence B under test condition *j*

(a) Stimuli presentation

12   13   14   15
A   B   A   B   A   B   A   B

Excellent

Good

Fair

Poor

Bad
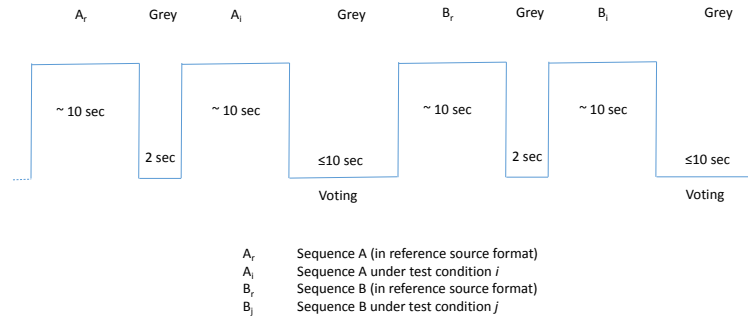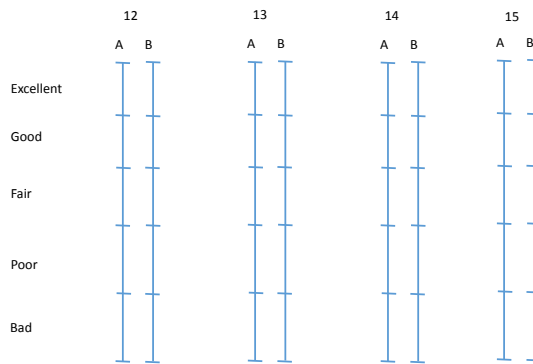
(b) Quality rating scale

Figure 3.6: Stimuli presentation and quality scale in DSCQS.

Double stimulus continuous quality scale

Double Stimulus Continuous Quality Scale (DSCQS) is defined in ITU-R Rec. BT.500. In this method test sequences are presented in pairs: the reference and the impaired, as shown in Figure 3.6. Subjects are asked to assess the overall picture quality of each presentation by inserting a mark on the vertical quality scale. The unimpaired one is included to serve as a reference, however the observers are not told which is the reference sequence. In the series of tests, position of reference is changed randomly. This method has a particular value, if it is not possible to provide test conditions that exhibit the full range of quality. It can be used for the evaluation of a new system or of the effects of network impairments on quality.

Subjective assessment of multimedia video quality

Subjective Assessment of Multimedia VIdeo Quality (SAMVIQ) method is a modified DSCQS [108]. Briefly, in this method, the subject is given several versions of a sequence resulted by combining different image processing features such as codec type, image format, bit-rate, temporal updating, zooming, etc. When all versions of the sequence are rated by the subject, the following sequence content can be then accessed. Different versions are selectable randomly by the subject and provides the possibility to stop, review and modify the score. Including explicit reference (i.e. unprocessed) several versions of the same sequence that include both processed and unprocessed (i.e. a hidden reference) sequences are included, and an observer can view the explicit reference whenever he/she wants, as shown in Figure 3.7. The continuous quality scale (0 to 100 annotated by 5 quality items linearly arranged: excellent, good, fair, poor, and bad) is used to provide a measurement of the intrinsic quality of test video sequences.
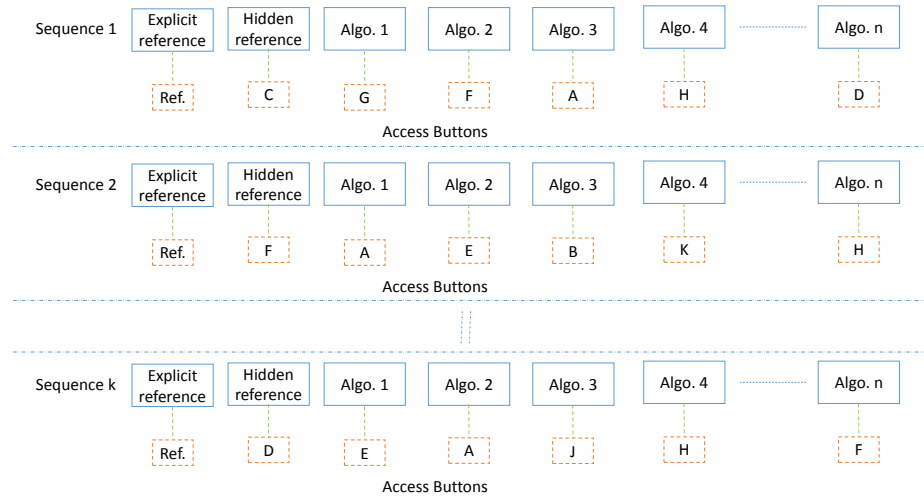
Figure 3.7: SAMVIQ: Test organization example.

3.2.4 Opinion scores processing

Outliers Detection

To detect and remove scores given by the subjects whose score is very far from the mean behavior, an outlier detection technique [112] is adopted. In brief, for each test sequence (k), mean $(\bar{x}_k)$, standard deviation (s_k), and *Kurtosis* coefficient $(\beta 2_k)$ are computed. $\beta 2_k$ is given by:

$$\beta 2_k = \frac{m_4}{m_2{}^2}, \tag{3.15}$$

and

$$m_x = \frac{\sum_{i=1}^{N}(x_{ik} - \bar{x}_k)^x}{N},$$ (3.16)

where, N is the number of subjects, x_{ik} is the judgment given by the ith user for k^{th} test visual signal and $\bar{x}_k$ is the average score given by all the subject to k^{th} visual signal. For each observer i, find p_i and q_i.

That is:

if $2 \leq \beta2_k \leq 4$, and

    if $(x_{ik} \geq \bar{x}_k + 2s_k)$, then: $p_i = p_i + 1$,

    if $(x_{ik} \leq \bar{x}_k - 2s_k)$, then: $q_i = q_i + 1$,

else

    if $(x_{ik} \geq \bar{x}_k + \sqrt{20}s_k)$, then: $p_i = p_i + 1$,

    if $(x_{ik} \leq \bar{x}_k - \sqrt{20}s_k)$, then: $q_i = q_i + 1$,

    where, s_k is the standard deviation on the subjective scores given for k^{th} visual signal. Finally for each subject, if $\frac{p_i + q_i}{N} > 0.05$ and $\frac{p_i - q_i}{p_i + q_i} < 0.3$ then the observer i is rejected.

Mean opinion score estimation

After the outlier detection, the MOS score for k^{th} video is calculated by Equation 3.17:

$$MOS_k = \frac{1}{N} \sum_{i=1}^{N} x_{ik},$$ (3.17)

where, N is the number of subjects and x_{ik} is the judgment given by i^{th} user for k^{th} video. Moreover, in order to remove the hidden reference, Difference MOS (DMOS) [109] for k^{th} test video sequence has been computed by Equation 3.18.

$$DMOS_k = \frac{1}{N} \sum_{i=1}^{N} r_{ik} - x_{ik},$$ (3.18)

where, r_{ik} is the judgment given by i^{th} user for reference of kth video.

The MOS values are in the range 1 to 5 while normalized DMOS scores are in the range 0 to 5. In these scales, higher MOS value and lower DMOS scores represent better perceived quality and vice versa.

Confidence Interval (CI)

When we present the results of a test, all mean scores should have an associated confidence interval which is derived from the standard deviation and size of each sample [109]. Therefore, the 95%

Confidence Interval (CI) is defined as:

$$[MOS_K - \delta_k, MOS_K + \delta_k], \tag{3.19}$$

where, $\delta_k = 1.96 \frac{S_k}{\sqrt{N}}$. Moreover, the standard deviation, S_k is computed as

$$S_k = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_{ik} - MOS_k)^2}. \tag{3.20}$$

3.2.5 Pair comparison data analysis

Pair comparison data

The result of the pair comparison experiment is a count matrix, C (also referred as a PC Matrix (PCM)), that contains the number of times that each option was preferred over the other option [227].

$$C_{i,j} = \begin{cases} \# \quad \text{times} \quad \text{option} \quad i \quad \text{preferred} \quad \text{over} \quad \text{option} \quad j, & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j. \end{cases} \tag{3.21}$$

Moreover, it is assumed that each paired comparison is independent, and thus, we don't need to know the order of the comparisons occurred. In other words, different pairs may have different number of comparisons.

Models for comparative judgment

As mentioned before, the result of the subjective experiment is the PCM matrix. In the field of perceptual quality evaluation, it is important to put the results in a continuous rating scale. For this purpose, a model based approach needs to be used. The model should be able to map the stimuli to metrics, which represent the magnitude of sensation [129]. In literature, there are two most common models (Thurstones model [223, 148] and Bradley-Terry model [31]) for analyzing paired comparison data (Equation 3.21). In the context of image processing, the Bradley-Terry model is widely used, and thus, the Bradley-Terry (BT) model is the scope of this dissertation. The BT model is also known as Bradley-Terry-Luce model (BTL) for Ducan Luce's extension to multiple variables [147]. A detailed description of the BT score and Confidence Interval (CI) estimation procedure is available in [253]. The original BT developed a model as giving each option a rating, π_i which satisfies

$$P(\text{choice} \quad A \quad \text{over} \quad B) = \frac{\pi_A}{\pi_A + \pi_B}. \tag{3.22}$$

In [147], choice axiom is formulated, to make a comparison of more than 2 objects; again each
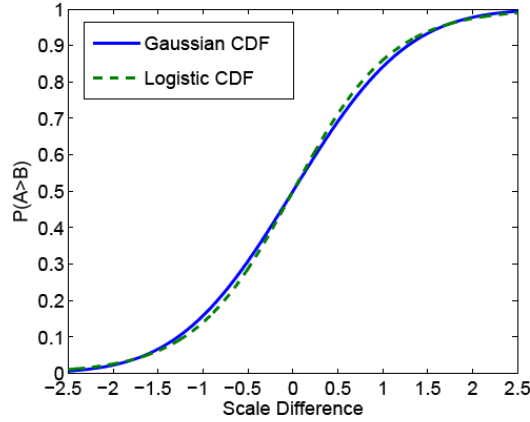
Figure 3.8: Gaussian vs Logistic CDF [227].

rating π_i must satisfy,

$$P(\text{choice} \quad A \quad \text{out} \quad \text{of} \quad A, \quad B, \quad \text{and} \quad C) = \frac{\pi_A}{\pi_A + \pi_B + \pi_C}. \tag{3.23}$$

Now, the variable $\pi_i = exp(\mu_i/s)$ and s is a scale parameter. Then the Equation 3.24 becomes:

$$P(\text{choice} \quad A \quad \text{over} \quad B) = P(A > B) = P(A - B > 0) = \frac{\exp(\mu_A/s)}{\exp(\mu_A/s) + \exp(\mu_B/s)} \tag{3.24}$$

$$= \frac{1}{2} + \frac{1}{2}\frac{\mu_A - \mu_B}{2s}. \tag{3.25}$$

Equation 3.25 can be expressed as: $1 - F_{A-B}(0)$, where F is the Cumulative Density Function (CDF) of the random variable $A - B$. It is consistent with Equation 3.25 to assume that $A - B$ is a logistic random variable with mean $\mu_A - \mu_B$ and scale factor s. In comparison to Thurstones model, the BT model assumes the random quality difference $A - B$ has a logistic distribution where the Thurston assumes that the random quality difference is Gaussian distribution. The logistic CDF is very similar to the Gaussian CDF (as shown in Figure 3.8), and thus the result produced by both models is very similar. Therefore, in this dissertation the BTL model is selected for computational simplicity. In other words, we don't have to compute the *erf* function for the inverse Gaussian CDF. Moreover, the BT has more stronger analytical capabilities and provides more statistical procedures than the TM (Case V) [83].

Then, the results presented in [29, 149, 227] show that if A and B have Gumbel distributions of qualities then $A - B$ is logistic function. Next, by inverting the Equation 3.25 the quality difference is estimated as: $\mu_{AB} = \mu_A - \mu_B$ and $P(A > B)$ is estimated as the empirical count proportion

$C_{A,B}/(C_{A,B} + C_{B,A})$. The *logit* (inverse logistic CDF) follows a closed-form expression (since, $tanh^{-1}(x) = 1/2[ln(1 + x) - ln(1 - x)])$, so BTL quality difference estimate is

$$\widehat{\mu_{AB}} = s(ln(C_{A,B}/(C_{A,B} + C_{B,A})) - ln(1 - C_{A,B}/(C_{A,B} + C_{B,A}))), \tag{3.26}$$

where, $s = \sqrt{3}/\pi$, for comparative results with Thurstone model.

Exact Test

An exact test is a statistical significance test. One of the most popular is a Fishers exact test, and it is used in the analysis of contingency table. In particular, the test is useful to examine the significance of the association (contingency) between the two kinds of classification. Another test that is used in the analysis of contingency tables is Barnards test. The Barnards test is a statistical significance test of the null hypothesis of independence of rows and columns in a contingency table, and it is more powerful than Fishers exact test for 2×2 contingency tables [22]. In our work the Barnards test has been used to check whether the probability, P_{ij}, (scores given for a test sequence) is significantly different from a probability of 0.5 (i.e., whether the observers are undecided) or not [137].

In detail, for a pair $\{I_i, I_j\}$, let us assume a_{ij} out of n_{ij} observers choose I_i and a'_{ij} observers choose I_j out of n'_{ij} observers. To test the significant difference between the probabilities that observers chose I_i over I_j, the Barnards test has been used. In this scenario, input matrix for the test could be $\left\{ \begin{array}{cc} a_{ij} & a'_{ij} \\ n_{ij} - a_{ij} & n'_{ij} - a'_{ij} \end{array} \right\}$, and the output of the test is a p-value. For example, at a 95% confidence, p-value < 0.05 means there is significant difference between the probabilities that observers chose I_i over I_j.

3.2.6 Data analysis tools

Box plot

The box plot (box and whisker diagram) is used to display the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum. On each box, the central mark is the median score, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points the algorithm considers to be not outliers, and the outliers are plotted individually [153].

Correlation coefficients

To analyze the correlation between the variables such as estimated subjective score and ground truth subjective scores (MOS), the correlation coefficients: Pearson's Linear Correlation Coefficient (PLCC), Spearman Rank Correlation Coefficient (SRCC), and Kendalls Tau Correlation Coefficient

(KTCC) are used. PLCC evaluates the linear relationship between two continuous variables. SRCC evaluates the monotonic relationship between two continuous or ordinal variables. Finally, KTCC rank correlation is a non-parametric test that measures the strength of dependence between two variables [40]. We included KTCC, since it is less sensitive to error and discrepancies in the data compared to the SRCC. High values of coefficients indicate best correlation between the variables. Moreover, the P_{value} is result of the hypothesis test of no correlation against the alternative that there is a nonzero correlation. The small P_{value}, say less than 0.05, indicates that the correlation is significantly different from zero.

Principal component analysis

In Principal Component Analysis (PCA) [117], positive or negative relationship between the variables and the principal component is expressed as a vector direction. The correlation between variables is measured in terms of the angle between them, and a small angle corresponds to a higher correlation [85]. In this work, the PCA is performed by using the MATLAB function *princomp* and FactoMineR Software [130].

Analysis of variance

Analysis of Variance (ANOVA) [217] test is used for comparing the means of two or more groups of data and determines whether any of those means are significantly different from each other or not. Particularly, it tests if the null hypothesis is accepted, that is the group means are equal. During the testing, test statistic is measured with the help of F-distribution (Fisher-Snedecor distribution), indicated as a F_{value}, and if the probability (p_{value}) for the F-statistic is smaller than the significance level, then the test rejects the null hypothesis i.e. accept alternative hypothesis (at least one of the group means is significantly different from the others). In this article, significance level, α of 0.05 has been considered.

3.3 Objective quality assessment

As mentioned before, the subjective quality evaluation demands significant time and effort for its implementation, and the dedicated evaluation equipment is required, and it gives the best measure of QoE. However, it is difficult to apply for quality monitoring and management, because it is time consuming, and not useful for many practical and real-time applications. Therefore, there is a strong demand of an objective quality assessment technique for estimating subjective quality from physical feature related to signal, service, and network (e.g. encoding rate, IP packet transfer time/fluctuations and IP packet loss ratio, etc.). The necessity for obtaining the automatic estimates of perceived quality of multimedia services is leading to the development of objective QoE metric.

The signal fidelity measure has been used for more than 50 years as MSE. It compares two signals and provides a quantitative score that describes the degree of similarity or level of error/distortion between them. Generally, one signal is distorted or contaminated signal, for which we want to measure the level of distortion, and another signal is a reference or pristine original. Mathematically, MSE between signals x and y is

$$MSE(x,y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 \,, \tag{3.27}$$

where, $x = \{x_i | i = 1, 2, 3, ..., N\}$ and $y = \{y_i | i = 1, 2, 3, ..., N\}$ are two finite length discrete signals.

MSE has been used in almost every field of signal processing from image processing, filter design, signal compression, restoration, reconstruction, classification, etc. to solve the optimization problems. Because of its simplicity and clear physical meaning, it is a convention. However, in many applications, the MSE exhibit weak performance and has been widely criticized for serious shortcomings, especially when dealing with perceptually important signals such as speech and images [249]. For better performance, in image processing, the MSE is converted into PSNR as:

$$PSNR = 10\log_{10} \frac{L^2}{MSE}, \tag{3.28}$$

where L is the dynamic range of image pixel intensities such as for images that have allocations of 8 bits/pixel of grayscale, $L = 2^8 - 1 = 255$. PSNR metric has been devised and used over decades to measure the quality of the visual signal. Equation 3.28 indicates the PSNR is a simple modification in MSE by taking into account the logarithmic relationship between physical stimulus and human perception. Furthermore, the performance of visual quality metrics has been improved by considering HVS characteristics [57].

In the above mentioned quality metrics, MSE and PSNR, to estimate the quality of distorted signal, complete knowledge of the reference signal is needed. In some applications, for an example, to estimate the quality of receiving image at user side the availability of complete knowledge of reference image is not possible, and thus, due to the requirement of reference signal limit the applications of such a metric.

Based on the availability of the reference signal, the level of processing complexity, and the accuracy they provide, the objective metrics are categorized into three groups: Full Reference (FR), Reduced Reference (RR), and No Reference (NR) [193].

Full Reference objective quality evaluation

In this technique, quality metric requires the access of both test visual signal and corresponding original signal in the same physical location. Furthermore, this technique assumes that the original sequence has a maximum quality. In particular, to estimate the objective quality scores a comparison
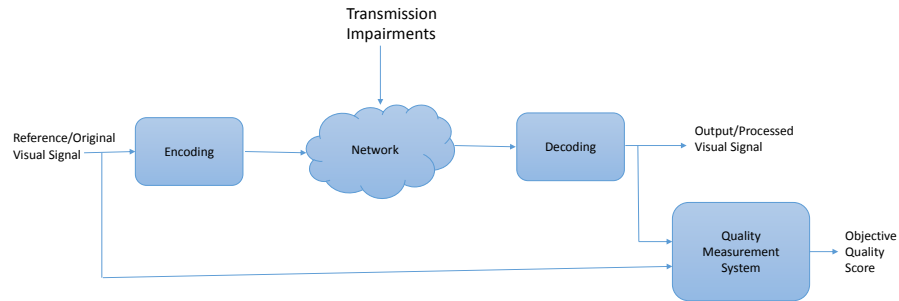
Figure 3.9: Full Reference objective quality assessment framework.

between input or reference video at the input of the system and the processed signal at the output of the system is performed (as shown in Figure 3.9).

By comparing the test visual signal to the original signal, the FR metrics are able to compute the distortion in the test sequence accurately. However, the challenge is, to find the particular type of the distortion that reduce the visual quality, because it is not necessarily that the visual quality is reduced equally in all aspects of the distortions . By exploiting this challenge, using HVS, the performance of FR metric can be improved significantly, as in PSNR-HVS [57] and SSIM [250].

The FR metrics are easy to compute and provides more accurate results and they are most appropriate for in-the-lab testing. Therefore, the FR metrics widely used to compare and analyze the performance of compression methods.

Need of original and test signal at the same physical location, the FR method may not be feasible in a networked application. Thus, in the context of multimedia communication environment (with the inclusion of communication network), these metrics are not applicable.

Reduced Reference objective quality evaluation
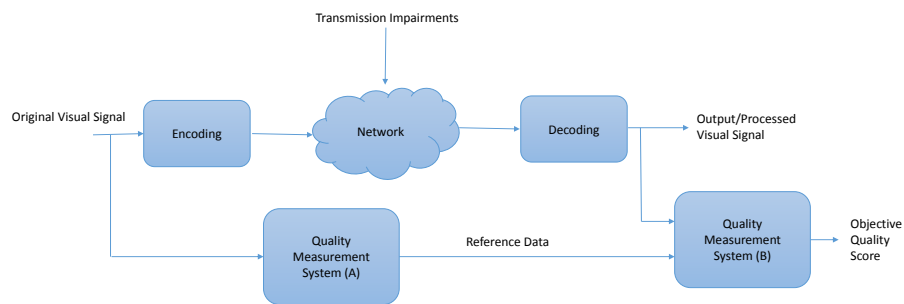
Figure 3.10: Reduced Reference objective quality assessment framework.

To solve of limitation of FR metric: requirement of the reference signal, the RR approach is introduced. In this technique, specific parameters are extracted from both the reference and the processed signals to compute the quality of the processed signal. As shown in Figure 3.10, quality measurement system (A) extracts some quality parameters of the original signal and it is used as a reference at the destination, quality measurement system (B), to estimate the quality score of test signal.

For an example, to evaluate the quality of processed video, the parameters blockiness, spatial and temporal signal information, and noise level, extracted from original signal can be used. Moreover, impairments such as freeze frames and loss of picture can also be detected at the output video.

This RR methodology offers potential for use where the decoder and encoder are physically separated for instance by a transmission chain, i.e. in the multimedia communication environment.

No Reference objective quality evaluation
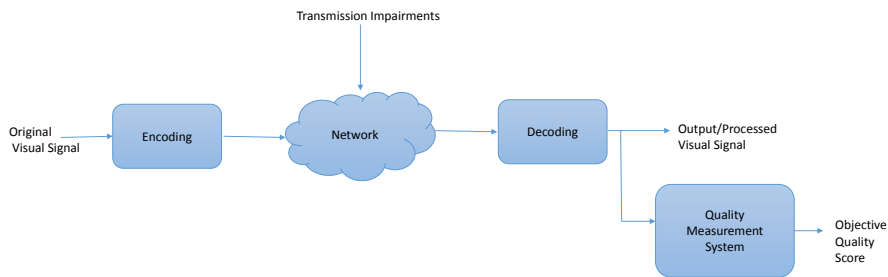
Figure 3.11: No Reference objective quality assessment framework.

In communication environment, it is not always possible to have a reference or even reduced reference information about the original signal. In this approach, the quality is estimated based on the extracted parameters of the processed signal, as shown in Figure 3.11. The lack of a reference means that the measurement may be subject to errors caused by picture content resembling by the specific impairments that are being detected. However, the common goal is to estimate the quality of processed visual signal without reference signal.

In the context of video communication, the most common way of estimating the quality is to use the detected impairments such as block distortion, freeze frames, and loss of pictures. The NR metrics can be categorized into pixel-based and bit stream-only metrics.

In NR pixel-based technique the quality is estimated by decoding the visual signal. Obviously, NR metric does not have the original and are therefore unable to exactly compute the distortion. Instead, they must infer the desired reference, distinguishing between desired signal and undesired distortion caused by compression and transmission artifacts [17]. In general, the performance of

the metric can be improved by limiting its scope, based on the target application, and by using the accurate models for reference and distortion. To the extent possible, it is always advantageous to incorporate the principles of HVS.

In NR bit stream-only technique the quality is estimated by counting the missing packets, called network-layer packet loss rate. In this technique, together with the complexity, usually the performance of the metric is also poor. The packet loss and its impact on the visual quality is depending on the adopted compression and transport techniques. For an example, in video communication environment the compression of video is compulsory due to the bandwidth limitation of the channel. To achieve a higher level of compression the video coders apply prediction between frames. The first frame is I-frame (intra-coded frame) and the subsequent frames are P-frames (predictively coded frames), if there is a single lost packet, it leads to an error in the decoded frame two and the error propagate significantly both in time and space to subsequent frames, ultimately reduce the quality of the video significantly. Moreover, the more accurate quality metric can be designed by combining all the information such as a bit stream, decoded pixels, and even from reference signal, if available.

Chapter 4

A dataset: to study the video quality of experience

This chapter presents the following contributions:

- a ReTRiEVED Video Quality Dataset: adopted protocols to create the dataset are briefly reported;

- as the use cases of the dataset, i) impact of the produced artifacts on the video QoE is presented, and ii) performance of the well known no-reference image/video quality metrics is evaluated on the dataset.

4.1 Introduction

As mentioned before, the recent advances in communication technology compression methods, storage devices, and high-speed networks, made it feasible to provide multimedia services over the internet. Among the multimedia services, video services (video streaming, video-on-demand, video broadcasting, etc.) are becoming the dominant services. Delivering the spatial and temporal sensitive video services over the noisy and bandwidth limited communication channel with a predefined level of QoE is challenging. This is because, the communication channel introduced impairments (delay, jitter, packet loss, and bandwidth limitation) degrade the quality of video.

Before moving into the focus of the study, video QoE, it is worthwhile to see the major source of artifacts in video communication system. The end-to-end- video communication chain can be divided into five components: capture, encode, network, decode, and display [17]; the overall QoE is influenced by the artifacts produced by each component.

For high QoE, original video should poses the high quality. For an example, if the original video is already afflicted by de-focus, blur, motion blur, noises due to low light, overexposure, camera

sensor noises or other undesirable effects, the overall QoE will be low, no matter the performance of the other components in the system. To communicate the video on a network, compression of the video is necessary. The compression method reduces the size of the video at the cost of introducing the compression artifacts, such as blocking, ringing, or blurring. Moreover, the compressed video is more vulnerable to network impairments such as packet loss. The severity of the artifacts depends induced network impairment and adopted compression and transport methods. Delivery network, often exhibit congestion and interference. The interference is even more in a wireless environment. The network produced impairments may include delay, jitter, and packet loss or bit errors either in isolation or in bursts and are varied over time based on the network conditions. The characteristic of display device used at the terminal, viewing conditions, and rendering methods also influence the video QoE.

The impact of the rendering process introduced artifact is more important for HD, 3D, multi-view/higher than 3D (LF) videos. The last but not least is the video content itself; different types of video content influence the quality at all points in the processing chain. For an example, coding parameters are significantly different from the video with high motion (running video) compared to the video with low motion (news reader). The effect of the content is not limited in the processing chain, also effect the QoE of the user, such as in general natural scenes provides better QoE to the viewer.

Video QoE is multidimensional and multidisciplinary field and is composed of the primary disciplines of image processing, computer vision, color, and computational and behavioral sciences. Among the top priority of the research in this field is the development of a computational model, referred as a quality metric, that is able to predicting the quality of video as perceived by the user. For evaluating the performance of the metric, an established practice is to use the dataset. In particular, the availability of video quality dataset, test video sequences with annotated subjective quality ratings, is important for assessing the video QoE, as well as testing new processing tools, or even assessing the effectiveness of objective quality metrics. Moreover, the collected quality ratings for the test videos are used to understand the response of HVS to visual signal, and the observed response is used for optimizing the video processing algorithms such as encoding and decoding.

The selection, as well as design, of the dataset is depending on the purpose of the study. Particularly, adopted protocols (selection of the SRC, HRC, quality assessment method, display device, subjects, environment, etc.) to create the dataset is depending on the purpose of the study. A scope of this chapter is to study the impact of transmission impairments and video content on QoE.

4.2 Literature survey

In the state-of-the-art, many ongoing works have devoted to devise the video quality datasets. The list of most popular and publicly available datasets known to us is reported in Table 4.1.

Table 4.1: Selected video datasets with corresponding distortion types.

Datasets	Date	Distortion Type
LIVE[205]	2009-10	MPEG-2, H.264/AVC, packet loss
LIVE Mobile[164]	2012	Compression, wireless packet-loss
EPFL-PoliMI [51]	2009	H.264/AVC, packet loss
MMSPG (SVD) [133]	2011	compression, spatial and temporal resolution
IRCCyNIVC[8]	2016	content, coding artifacts, transmission error (loss), and error concealment [183, 182]
ECVQ and EVVQ [240]	2012	coding artifacts: H.264/AVC and MPEG4
VQEG [81]	2000 and 2010	encoding (MPEG2, H.264), packet loss
Poly@NYU[145]	2009	H.264/AVC, SVC, packet loss
IT-IST[177]	2010	encoding: MPEG2 and H.264/AVC, packet loss
TUM Data Set [123, 122]	2012	encoding: H.264
MCL-V Database [140]	2015	Compression, scaling
IVP [263]	2011	encoding (MPEG2, H.264), packet loss
CSIQ [241]	2014	Encoding (H.264, HEVC, MJPEG, SNOW), packet loss, and Gaussian noise
AVC HD [218]	2013	H.264 encoding and slice losses
CVD Video Database [172]	2015	acquisition noises

As presented in Table 4.1, in most of the datasets, together with encoding artifacts packet loss is considered as a consequence of the noisy channel impairment. The impact of artifacts produced during the capturing process is deeply investigated in literature [172]. The most challenging component of the multimedia communication chain is a network. As mentioned before, as a consequence of the impairments produced by the network, Packet Loss Rate (PLR) is considered. However, the inclusion key network impairments, called QoS parameters (in the context of IP based network services): delay, jitter, packet loss rate, and bandwidth limitation, as HRCs in the dataset is crucial to study the impact of network impairments on video QoE. However, in literature, the key QoS parameters are not considered. Therefore, this chapter presents the description of the dataset designed by considering the QoS parameters as HRCs.

4.3 Adopted subjective quality assessment framework

In the following, the procedures adopted for creating the dataset are tailored.

4.3.1 Stimuli

In this experiment, eight heterogeneous uncompressed digital videos with different content, motion, texture, color temperature, and camera movement have been selected as SRCs. The heterogeneity of the selected sequences is demonstrated by their wide span in the spatial-temporal information plane (as shown in Figure 4.1).
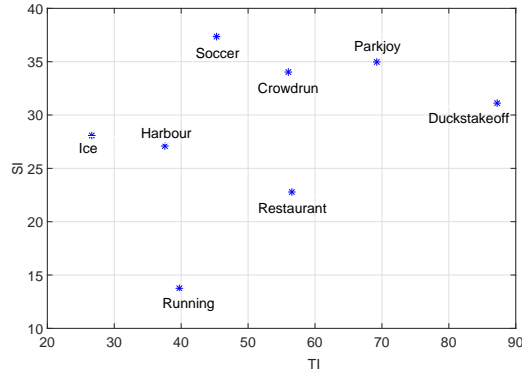
Figure 4.1: Source sequences features characterized with the help of Temporal perceptual Information (TI) and Spatial perceptual Information (SI) plane.

Next, Table 4.2 summarizes the features of the SRCs, while sample frames are shown in Figures 4.2. The sequences *Crowdrun*, *Duckstakeoff*, *Harbour*, *Ice*, *Parkjoy*, and *Soccer* are obtained from EPFL-PoliMI video dataset [51], while *Running* and *Restaurant* are extracted from the Consumer Digital Video Library (CVDL) [4].

Table 4.2: Details of source sequences including Size, Frame Rate (FR), and Length.

Videos	Size (pixel)	FR (fps)	Length (s)
Crowdrun	704×576	25	9
Duckstakeoff	704×576	25	9
Harbour	704×576	30	9
Ice	704×576	30	7
Parkjoy	704×576	25	8
Soccer	704×576	30	7
Running	720×576	25	9
Restaurant	720×576	25	8

4.3.2   Test dataset

As shown in Figure 4.3 PVSs are generated by streaming the original SRCs from a VideoLAN streaming server through a noisy channel simulated by NETwork EMulator (NETEM) [89]. The components of the adopted HRC (Hypothetical Reference Circuit) are described in the following subsections.
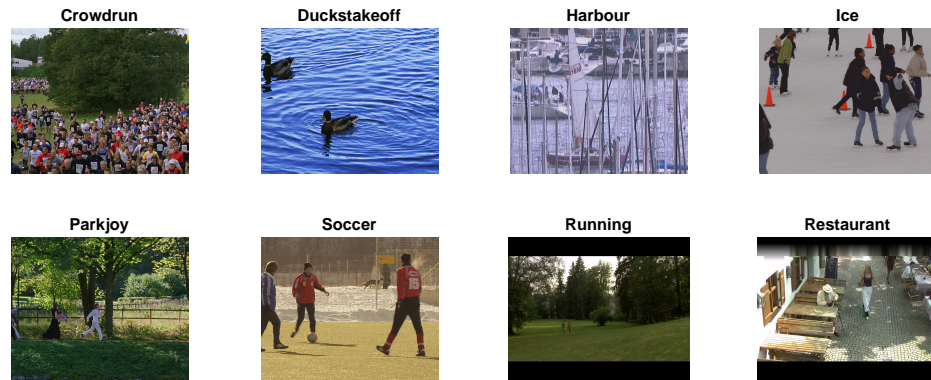
Figure 4.2: Sample frame of source videos.
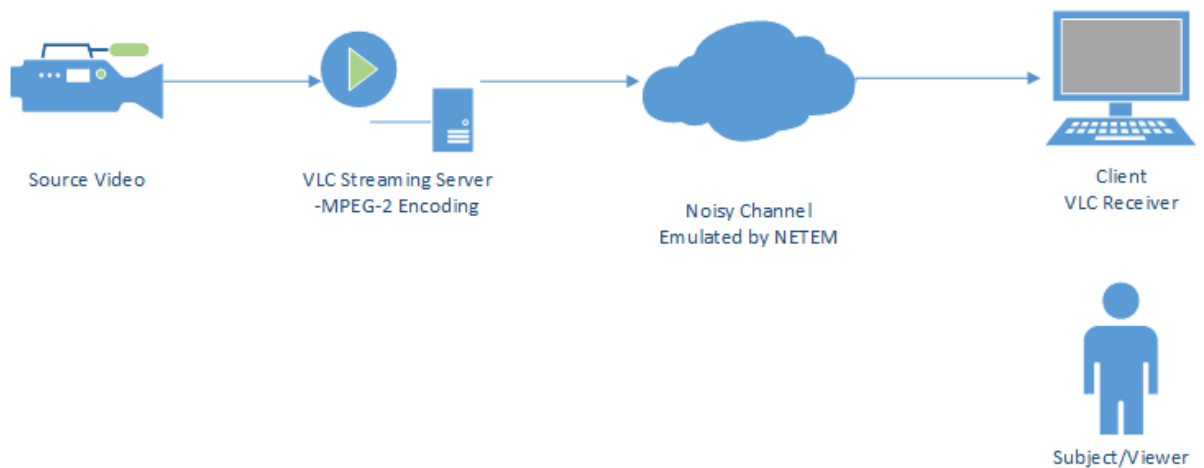
Figure 4.3: Block diagram of the experimental setup.

Streaming server

SRCs have been transmitted by using a VideoLAN (VLC Player: Version 2.1.3 with a caching size of 300 ms) streaming server. The server has been configured with the following parameters: MPEG2 (Moving Picture Experts Group) encoder, original frame rate as shown in Table 4.2, MPEG-TS (MPEG Transport Stream) encapsulation, coding rate of 9000Kbps (maximum value of the coding rate has been considered to minimize the visual impact of compression), and User Datagram Protocol (UDP) transfer protocol. Our focus is on the effect of transmission impairments on the perceived video quality. In this paper, MPEG2 compression scheme has been used. This choice relies on the fact that MPEG2 is still popular and it is used in a variety of applications including DVDs

(Digital Versatile Disc) and digital broadcast services. As a general rule, it should be always better the to use the newest coding standards. However, as a safe practice, providers must take into account the capabilities and the features of existing consumer electronics (i.e., receivers, decoders, etc.). Typically, the transition period from existing standard or technology to full adoption of some new standard or technology could vary over a large range. At the moment, most of terrestrial broadcasting systems are based on MPEG2. Moreover, the main focus of this work is not directed towards the study of the impact of the encoding artifacts. Therefore, we assume that the during analysis, the influence of the coding artifact can be compensated for by using hidden reference removal technique [109]. For this reason we are focusing this work on this standard.

Channel impairments

The noisy communication channel has been simulated by using NETEM. To obtain the PVS set, different values of impairments are considered. To have acceptable results, the standard values of transmission artifacts have been considered based on ITU and ETSI recommendations. Based on ITU recommendation on subjective VQAmethods for multimedia applications [108], framework and methodologies for the determination and application of quality of service parameters [111], and opinion model for video telephony applications [204], 1000 ms of delay, 10% of packet loss rate, and 5 Mbps of bandwidth have been considered as a maximum limit. Moreover, based on ETSI recommendation on speech and multimedia transmission quality [59], 5 ms of peak to peak one-way jitter (5 ms of jitter means maximum of 10 ms two-way jitter) has been considered as a maximum threshold. Five values of impairments have been considered. Furthermore, two more values (0.1% and 0.4%) of PLR have been simulated to be able to compare with other existing datasets [51].

The intermediate test points for each artifact have been selected by equally partitioning the space into 5 parts, and also keeping in mind that the PVSs should span the wide range of visual quality scores, as detailed in [205]. The effect of the delay has been simulated by introducing a set of five different delay amounts (100, 300, 500, 800, and 1000 ms) on each packet passing through the node. The effect of jitter has been added by introducing the delay of 100 ms plus or minus 5 variations (1, 2, 3, 4, and 5 ms). The effect of PLR has been introduced by randomly dropping the packets at a node with seven different PLR values (0.1, 0.4, 1, 3, 5, 7, and 10%). Finally, the channel bandwidth is controlled for five different values (512 Kbps, 1 Mbps, 2 Mbps, 3 Mbps, and 5 Mbps) by using the *TokenBucket* filter [89].

Client

The end point of our transmission framework is a device running a VideoLAN client.

A total of 184 videos were considered: 56 (8×7) videos with PLR, 40 (8×5) videos with jitter, 40 (8×5) videos with delay, 40 (8×5) videos with bandwidth and 8 MPEG2 compressed reference videos have been considered for the subjective evaluation. In order to remove the possible presence
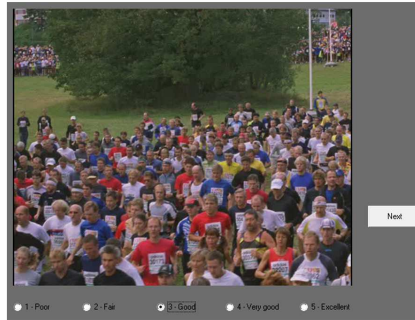
Figure 4.4: Screenshot of the GUI used for the subjective experiment.

of bias in the MOS, due to the effects of MPEG2 compression, the subjective tests included MPEG2 encoded videos, with and without simulated impairments. As a result of this choice, the subjective score of the reference video sequences can be used to remove the influence of the MPEG2 encoding and of the selected setup by using hidden reference removal techniques, thus obtaining the DMOS.

4.3.3   Test methodology

As mentioned before, the ACR method is a categorical judgment where the PVSs are presented one at a time and are rated independently on a categorical scale. This method is also called single stimulus method [108]. Single stimulus paradigm suit well to large numbers of emerging multimedia applications, such as quality monitoring for video on demand, internet streaming, mobile video services [226], etc. It reduces significantly the amount of time needed to conduct subjective studies compared to the double stimulus method in which SRCs and PVSs are shown at the same time [205]. Moreover, five scale ACR method is the most appropriate method in terms of stability and assessment time [119]. Therefore, the single stimulus discrete five scale ACR method (ACR-5) is taken as assessment method to obtain subjective quality scores of test video sequences.

4.3.4   Experiment setup and display

The experiment has been conducted in a controlled environment in order to produce reliable and reproducible results by avoiding involuntary influence of external factors [109]. The characteristics of the computer used in the experiment are briefly described in Table 10.4(a). Videos with their original size were displayed on a Panasonic TX-P42VT30E plasma screen TV (parameters are shown in Table 4.4) with an interactive Graphical User Interface (GUI) as shown in Figure 4.4.

    To compensate the effect of a potential bias based on order or position of simultaneous stimuli in the averaged result [126], stimuli have been shown in pseudo-random order for each subject

Table 4.3: System parameters.

Parameters	Values
Processor	Intel(R)Core(TM)i7-4770
Processor Speed	@3.40GHZ
RAM	8GB
System type	64-bit OS
Operating System	Windows 8.1
Video Player	VLC Media Player (version 2.1.3)

Table 4.4: Display parameters.

Parameters	Values
Display Device	Panasonic TX-P42VT30E TV
Screen Refresh Rate	60Hz
Screen Resolution	1920x1080 pixels
Image Mode	Normal
Contrast	30
Luminance	0
Color Temperature	Normal
Color Intensifier	On

according to their distortion type and intensity. Even though the videos were displayed in pseudo-random order, there were gaps of at least three videos between the video from the same reference in order to remove the memory and the contextual effect on the quality judgment. To minimize the effect of viewer fatigue on quality assessment, two experimental sessions were scheduled. Each session lasted 20 minutes including evaluation time and training session. Moreover, at least half-hour gap between each session was maintained to retain the attention of the subject.

4.3.5   Subject and their training

To collect reliable results, we exploited a sufficient number of subjects for quality evaluation as detailed in [121] [109]. There were 41 subjects, 25 males and 16 females of age between 21 and 51 who evaluated the effect of delay, jitter and packet loss rate and 30 subjects, 20 males and 10 females of the age between 21 and 51 years who evaluated the impact of bandwidth. The difference in the number of subjects is due to the fact that for the bandwidth impairment, a separate subjective experiment has been scheduled. Subjects were drawn from a pool of undergraduate to post-doctorate students from Università degli Studi Roma TRE. The students were relatively naive concerning video impairments and the associated terminology. They were asked to wear any vision correcting devices (glasses or contacts) that they normally wear to watch television.

In the first stage, the subject was verbally given instructions. In the training stage, lasting for two minutes, the subject was shown the original videos followed by video examples with the strongest impairments found in the experiment. In this phase, each subject got familiar with the assessment
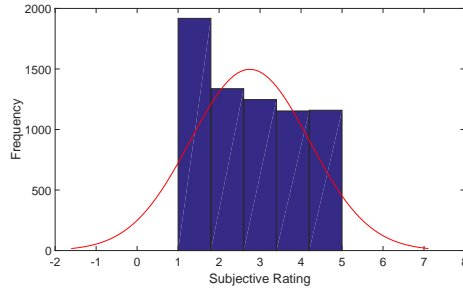
Figure 4.5: Distribution of subjective scores in the dataset.

procedure and established the annoyance value range. The videos used for the training session were different from the test videos in order to remove the memory and hidden reference effect.

4.4 Data Processing

From the experiment, subjective scores are recorded for test video sequences. The distribution, Figure 4.5, shows that the scores are distributed over the range 1 to 5.

As presented in Section 3.2.4, first outlier detection technique is applied; as a result of the outlier detection, scores given by 5 subjects are rejected for PLR, delay, and jitter and 7 subjects are rejected for bandwidth. After the outlier detection procedure, the MOS and DMOS scores are computed with associated confidence interval.

4.5 Use case of the dataset

As mentioned before, the dataset can be used for a wide range of applications. In the context of this dissertation, this is used particularly for two purposes: i) to understand the impact of producing artifacts on QoE, and ii) to benchmark the QoE assessment metrics.

4.5.1 Impact of the introduced artifacts on video quality of experience

To understand the impact of the introduced artifacts (transmission impairments) on video QoE, the computed MOS scores together with its confidence interval are plotted in Figure 4.6. It shows that the impact of the impairments on QoE is depending on the introduced artifacts: impairment types and levels of the artifact. It can be noticed that, the impact of delay on the QoE is not significant, whereas, the QoE is influenced by the different values of other artifacts: jitter, PLR, and bandwidth limitation.

The detailed analysis of the impact of transmission impairments is discussed in next chapter, Chapter 5.

(a) Delay.

(b) Jitter.

(c) PLR.

(d) Bandwidth limitation.
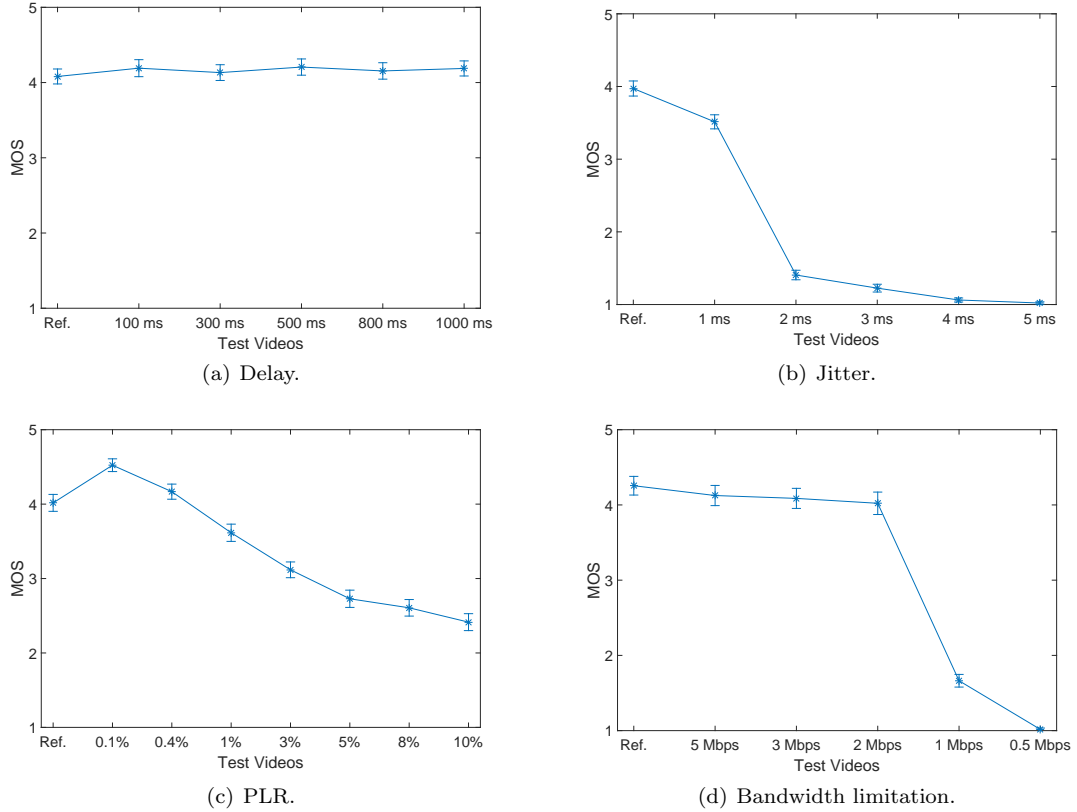
Figure 4.6: MOS with corresponding 95% confidence interval for the test videos distorted by different transmission artifacts.

4.5.2 Objective video quality metrics evaluation and comparison

As a use case of the proposed video quality dataset, the performance of most common NR quality metrics: Naturalness Image Quality Evaluator (NIQE) [160], Blind Image Quality Index (BIQI) [162], Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [159] and BLind Image Integrity Notator using DCT Statistics-II (BLIINDS-II) Index [201], is evaluated. For this purpose, correlation between the MOS and the estimated quality score has been studied by means of PLCC, SRCC, and PCA. The video quality has been expressed in terms of average frame quality, each frame is considered independently (each frame is an image).

From Figure 4.7 it can be noticed that the correlation between subjective scores and quality score estimated by the metrics for the test videos is very low and BLIND-II has higher values of the SRCC and PLCC if compared to other metrics. This result could be due to the fact that it was designed by using a natural scene statistics (NSS) model of discrete cosine transform (DCT) coefficients. However, the correlation is not significant. The same trend in the correlation is confirmed by the
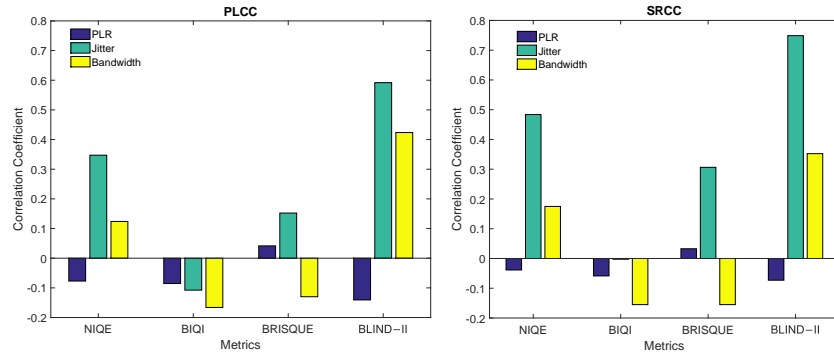
Figure 4.7: Performance comparison of objective metrics for videos which are distorted by PLR, jitter and bandwidth.
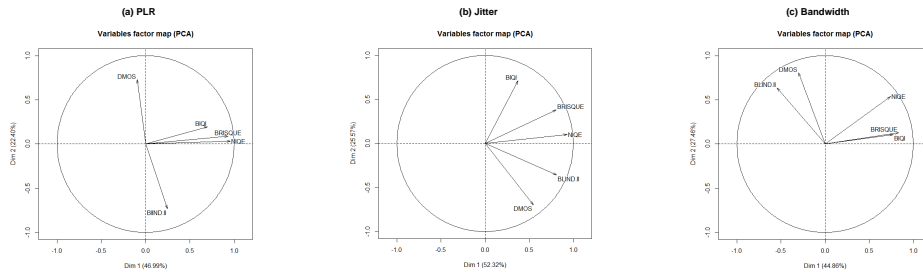
Figure 4.8: PCA results of the performance comparison of objective metrics for videos which are distorted by PLR, jitter and bandwidth.

result of PCA, Figure 4.8, the angles between the subjective score and metrics are noticeable.

Moreover, Figure 4.9 shows the PLCC is different for different videos, though the videos were distorted by the same level of the impairments, showing that the metrics perform differently depending on video content and impairments.

From these results it is clear that the video quality estimated from these metrics on the Re-TRiEVED video dataset is very far from the subjective assessment. So, it is important to devise a new video quality metric, and thus, a new video QoE metric is proposed in Chapter 6.3.2.

4.6   Conclusion

In this chapter, a video quality dataset is presented. Firstly, a detailed discussion about stimuli, test dataset, test methodology, experiment setup, subjects, and adopted data processing tools and techniques have been presented. The subjective procedures were designed based on the available standards, guidelines, and recommendations to realize a reliable video quality dataset. Test videos
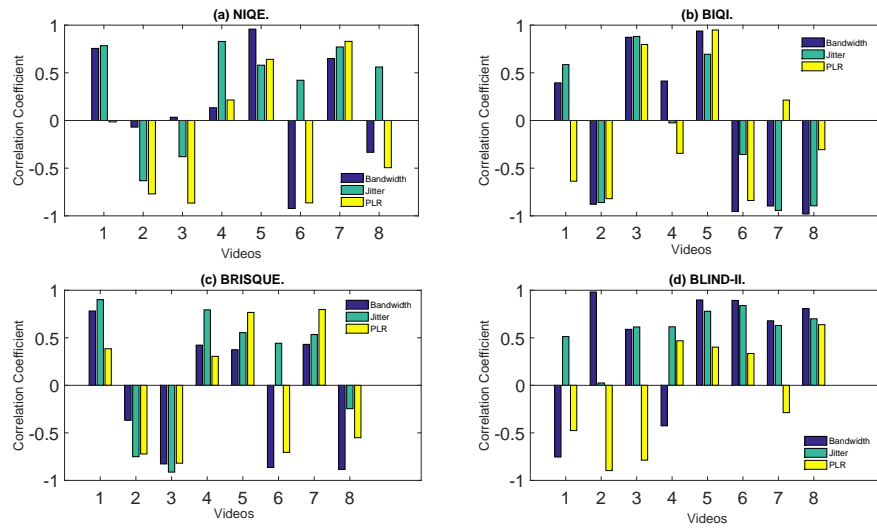
Figure 4.9: Performance evaluation of the metrics: PLCC for different video sequences and metrics.

were created from eight heterogeneous original videos by considering key impairments. Test materials, tools and associated results are freely available for the research community, and could be used for design, assessment and comparison of the video quality metrics.

Next, as a use case of the dataset, the impact of the HRCs on QoE is briefly presented. The detailed discussion on the impact of transmission impairments and video content is presented in the following chapter. Lastly, the performance of no-reference image/video quality metrics has been compared and the results show that the considered metrics poorly perform on the dataset, and the performance of these metrics varies for different impairments and SRCs, thus further evidencing the need of new no-reference video quality metrics.

Chapter 5

Impact of impairments on quality of experience

The contributions included in this chapter are:

- the analysis of the effect of the transmission impairments: delay, jitter, packet loss, and band-width limitation, on the video QoE;

- influence of video content on the QoE;

- effect of video content in the presence of the impairments;

- finally, the impact of video content related attributes on video QoE.

5.1 Effect of transmission impairments on video quality of experience

Related works

In the literature, many works have been carried out to understand the effect of impairments and artifacts on video QoE. A study of the joint impact of frame rate, frame size, coding bit rate, packet loss, video complexity, and MPEG2 compression have been presented in [212]. In [39] the influence of packet loss on perceived video quality is presented by exploiting packet loss distributions; this study is limited to two videos and two different values of PLR. In [98] the influence of packet loss, packet reordering and coding bit rates on the perceived quality of video streaming services is presented. The effect of packet loss and bandwidth reduction on perceived video quality has been analyzed and their relationship with perceived quality has been presented and discussed in [242]. In [96] the

61

Chapter 5

Impact of impairments on quality of experience

The contributions included in this chapter are:

- the analysis of the effect of the transmission impairments: delay, jitter, packet loss, and band-width limitation, on the video QoE;

- influence of video content on the QoE;

- effect of video content in the presence of the impairments;

- finally, the impact of video content related attributes on video QoE.

5.1 Effect of transmission impairments on video quality of experience

Related works

In the literature, many works have been carried out to understand the effect of impairments and artifacts on video QoE. A study of the joint impact of frame rate, frame size, coding bit rate, packet loss, video complexity, and MPEG2 compression have been presented in [212]. In [39] the influence of packet loss on perceived video quality is presented by exploiting packet loss distributions; this study is limited to two videos and two different values of PLR. In [98] the influence of packet loss, packet reordering and coding bit rates on the perceived quality of video streaming services is presented. The effect of packet loss and bandwidth reduction on perceived video quality has been analyzed and their relationship with perceived quality has been presented and discussed in [242]. In [96] the

authors conclude that the packet reorder, jitter, and packet loss are equally important parameters and the delay is the least important parameter in affecting perceived video quality, while the quality is more influenced by network impairments than video content. The effect of stalling events and video motion on QoE for the mobile streaming services is studied in [72] and the results show that there is no evident relation of video motion on QoE.

In [156] the impact of packet loss and jitter on perceived quality based on the subjective results is discussed. The results show that the mobile videos encoded with the H.264 baseline profile are very sensitive to the network impairments, and the perceived video quality drops quickly with nominal increase in packet loss and jitter. However, the validity of the study is limited, since for the subjective study only three reference videos and a limited number of subjects have been used. The effect of packet loss and jitter on perceptual quality is briefly discussed in [45] and the results show that the jitter degrades the perceptual quality as much as packet loss. The influence of network impairments on QoE in H.265/HEVC video streaming has been discussed in [170]. The authors conclude that the packet loss rate of 3% is the threshold value where most of the users found the quality annoying or worse.

Motivation

In the above mentioned works, the impact of the compression, encoding artifacts, and PLR on the perceived video quality has largely been discussed. However, these previous works have the shortcoming that only a limited number of subjects and/or test video have been used. Moreover, there are other important transmission impairments such as one-way delay (the delay), instantaneous packet delay variation (jitter), PLR, and bandwidth [213], that have to be considered conjointly. Therefore, in the following subsections, the impact of the transmission impairments on video QoE is presented.

5.1.1 Effect of delay

Box plot, Figure 5.1 (a), shows that median opinion scores at different levels of delay artifact are the same and that the box distribution is also uniform. As it can be noticed in Figure 5.1 (b), the delay does not have a significant impact on the perceived video quality. This is also confirmed by the trend of DMOS scores as shown in Figures 5.1 (c). These results indicate that there is no evident relationship between normalized DMOS scores and considered delay values. An additional evidence of this behaviour is provided by the ANOVA test. It has been performed against the null hypothesis to understand the effect of delay on the perceived quality. The result, $p_{value} = 0.977$, indicates that the MOS scores are not significantly affected by the variations on delay values.

The motivation behind this behavior of the delay could be that, in case of presence of transmission delay only (no packet loss or jitter), the video can be displayed smoothly with the help of a buffer [42]. The similar result, to the impact of delay artifact on the QoE, is presented in article [72], and argued
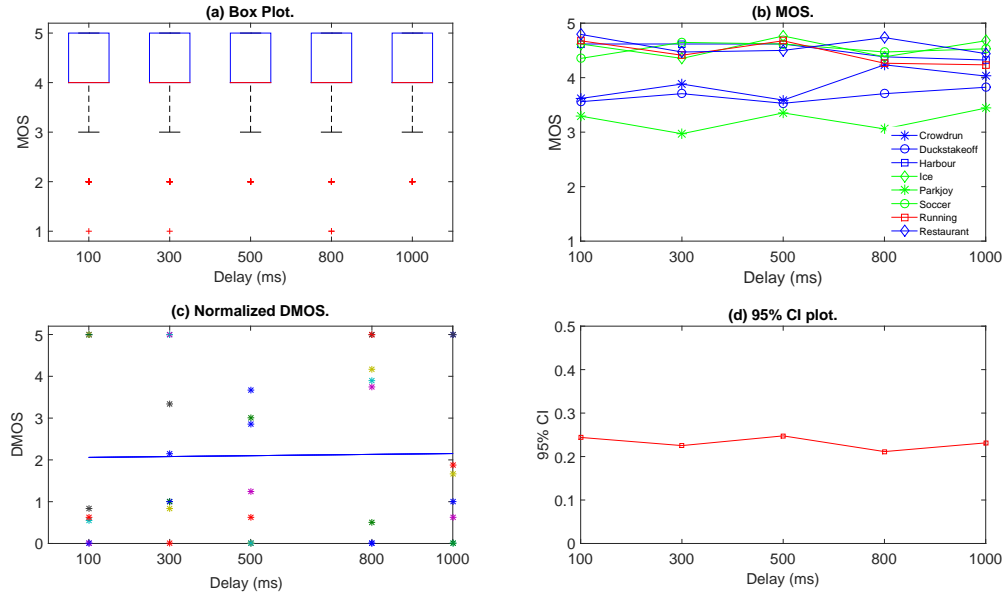
Figure 5.1: Perceived quality of the test videos in the presence of delay impairment.

that the subjects were not too frustrated with the initial delay. Moreover, the emphasis was given to a more focussed study with more videos having initial delays of several different lengths for better understanding.

Moreover, by analyzing CI of the opinion scores in Figure 5.1 (d), it can be noticed that the CI is significantly large for all the adopted delay values, indicating that the quality perceived by different subjects is different even for the same test video. This behaviour could be explained by the fact that different users have different perception even for the same video content, thus implying that the perceived quality is influenced by the video content rather than the adopted delay values.

5.1.2 Effect of jitter

As well known in literature [45], jitter significantly degrades the perceived video quality. In the performed test, when the jitter increases over 2 ms, the perceived quality goes towards its minimum value.

Figure 5.2 (a) shows the median opinion scores and distribution of opinion scores at different values of jitter and indicate that the median scores are significantly low for high values of jitter ($>$ 1 ms). The result is further confirmed by Figure 5.3. It shows a sample frame of a video affected by different values of jitter. Moreover, the Figures 5.2 (b) and 5.2 (c) show that perceptual quality score decreases significantly for high values of jitter, until a value of 2 ms is reached. When the jitter becomes larger than 2 ms, the decreasing rate is slower and becomes almost constant when jitter
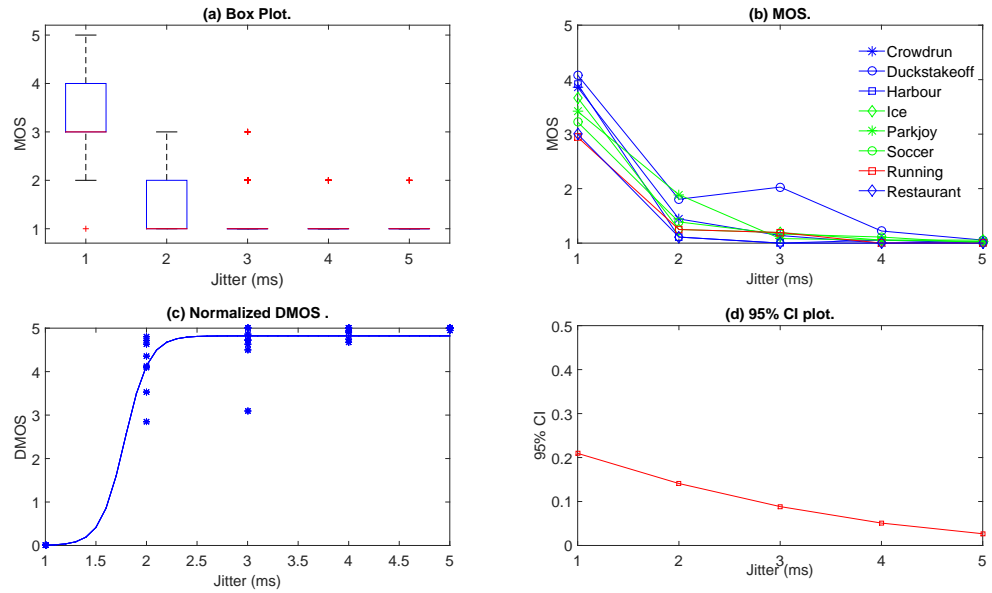
Figure 5.2: Perceived quality of the test videos in the presence of jitter impairment. For DMOS, in (c), the mapping function is $-5.059 * Jitter^{-2.374} + 5.062$; with RMSE= 0.404.

becomes larger than 3 ms.

Moreover, the results of the ANOVA test, shown in Table 5.1, $p_{value} = 0$, show that the MOS scores vary significantly for different values of jitter. However, if jitter values are larger than 2 ms, the p_{value} increases significantly. Therefore, the 2 ms of jitter can be considered as a JND value.

Figure 5.2 (b) shows that for videos $Duckstakeoff$ and $Parkjoy$ the MOS scores at 1 ms of jitter is 4.083 and 3.416, at 2 ms of jitter is 1.805 and 1.888, and at 3 ms of jitter is 1 and 1.083. These results indicate that for low values of jitter (1 ms) the perceived quality is significantly different for different videos. Moreover, as shown in Figure 5.2 (d), CI of opinion scores is high at low values of jitter and decrease for high values of jitter. These results indicate that for low values of jitter, the

Figure 5.3: Visual impact of different jitter. From left to right: original frame, 1 ms (PSNR 30.67 dB), 2 ms (PSNR 14.75 dB), and 5 ms (PSNR 13.07 dB).

Table 5.1: Summary of ANOVA test results.

Jitter (ms)	F	$F_{critical}$	p-value
1, 2, 3, 4 and 5	30.997	2.64	0.00
3, 4 and 5	0.435	3.47	0.653

Figure 5.4: Visual impact of PLR. From left to right: original frame, 0.1 % (PSNR 25.03 dB), 1% (PSNR 22.30 dB), and 10% (PSNR 13.90 dB).

perceived quality is significantly different for different video content. In other words, the perceived quality is significantly influenced by the video content for low levels of the jitter artifact.

5.1.3 Effect of packet loss

The impact of PLR values on the perceived quality is shown in Figure 5.4. The distribution of opinion scores for test videos at different levels of PLR is shown in Figure 5.5 (a). Figure 5.5 (a) shows that the median opinion scores decreases for high values of PLR. Figures 5.5 (b) and 5.5 (c) show the MOS and DMOS behaviors. In both cases, the trend is similar: MOS score decreases for every increase in PLR for all the videos and becomes almost constant when PLR becomes larger than 3%, while DMOS score increases almost linearly for higher values of PLR and when the PLR becomes larger than 3% the rate of increment becomes smaller. The results of the ANOVA test (shown in Table 5.2) suggest that the MOS scores are significantly different for different values of PLR artifact. However, when PLR becomes larger than 3%, the p_{value} increases. Based on the obtained results, 3% can be considered as JND threshold, as mentioned in [170].

Figure 5.5 (d) shows that at a low PLR value, the CI on the opinion scores is high i.e. perceived quality is differentiated for the users even at a same level of PLR. Moreover, Figures 5.5 (b) shows that different videos ($Crowdrun$ and $Parkjoy$) have different MOS scores even for the same values of PLR and the trend is confirmed by all the adopted videos.

These results indicate that, at a low value of PLR the perceptual quality is also influenced by the image content. However, at high values of PLR the perceptual quality of different videos is not significantly different. This result could be due to the fact that the at low level of PLR, the perceived quality is influenced by video content and user goal and expectations, as in the case in which the channel is affected by bandwidth and jitter.
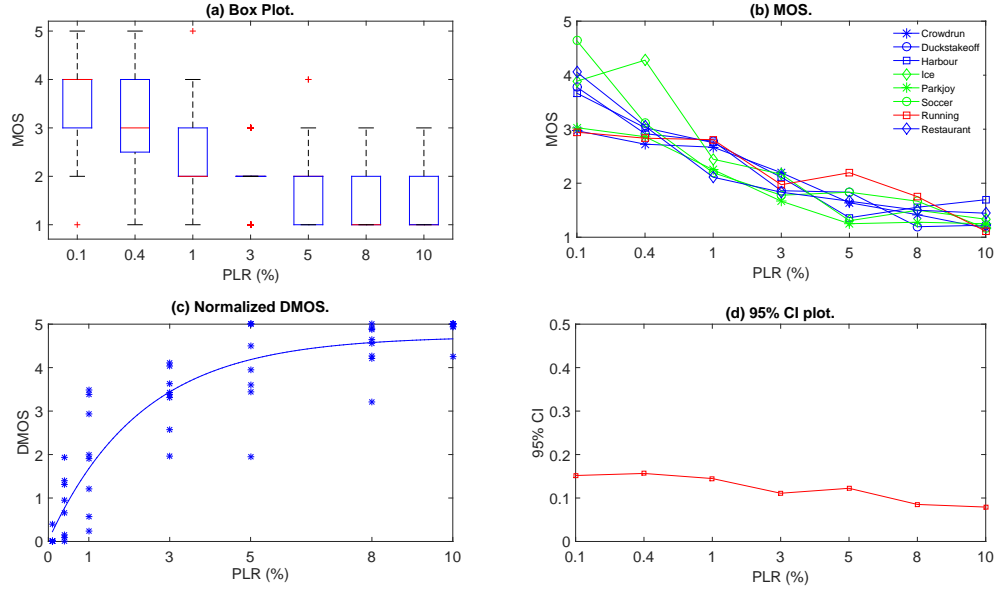
Figure 5.5: Perceived quality of the test videos in the presence of packet loss impairment. For DMOS, in (c), the mapping function is $3.261 * exp(0.04 * PLR) - 3.509 * exp(-0.86 * PLR)$; with RMSE= 0.746.

Table 5.2: Summary of ANOVA test results.

PLR (%)	F	$F_{critical}$	p-value
0.1, 0.4, 1, 3, 5, 8 and 10	29.803	2.29	0.00
5, 8 and 10	1.163	3.47	0.332

5.1.4 Effect of bandwidth

The distribution of opinion scores for different levels of bandwidth is shown in Box Plot, Figure 5.6 (a). Figure 5.6 (a) shows that the median score increases for high values of bandwidth. However, for high values (> 2 Mbps) of bandwidth the median scores are same.

Figures 5.6 (b) and (c) show the MOS and DMOS score variation in the presence of bandwidth. The analysis of the collected scores shows a common behavior for all analyzed videos: the perceived video quality increases for high values of bandwidth; however, when the bandwidth value exceeds 2 Mbps, the quality curve slowly increases and it becomes almost constant. Therefore, the 2 Mbps value of bandwidth can be considered as a JND threshold.

For better understanding the perceivable effects caused by bandwidth reduction, let us refer to Figure 5.7, where frames extracted from the videos transmitted over channels affected by bandwidth are shown. PSNR for each frame has been computed with respect to the original frame. As expected, it can be noticed that when the available bandwidth reduces, the quality decreases while when the
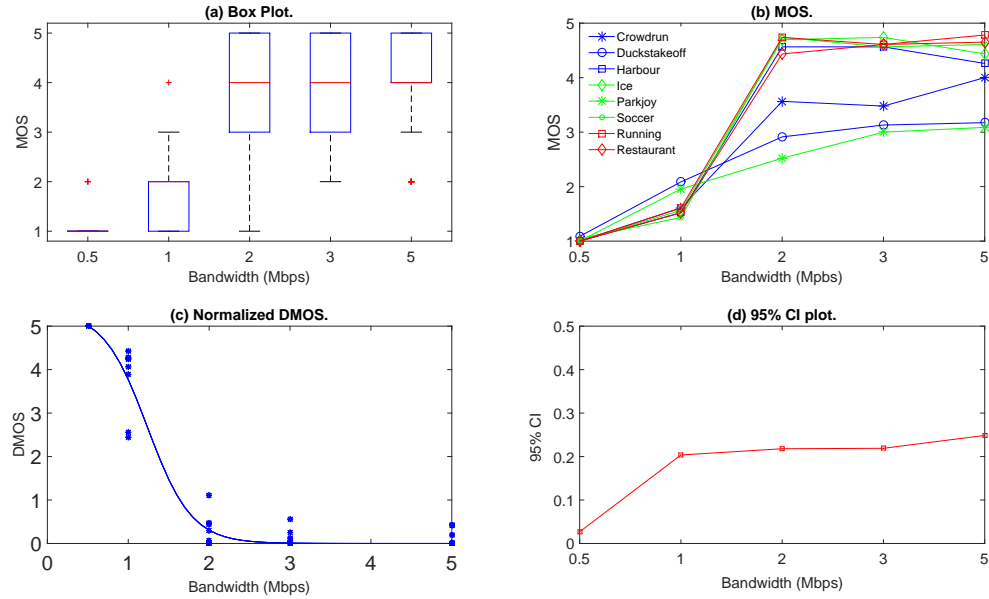
Figure 5.6: Perceived quality of the test videos in the presence of bandwidth impairment. For DMOS, in (c), the mapping function is $5.35/(1 + exp(3.671. * (Bandwidth - 1.237)))$; with RMSE=0.412.

bandwidth values outreaches 2 Mbps, reference and transmitted frames look almost identical and the PSNR value for the frame is high (Figure 5.7), thus confirming the trend highlighted by the analysis of the MOS and DMOS scores.

This result is further confirmed by ANOVA test. The test has been performed against the null hypothesis that the MOS scores are not significantly different for different values of bandwidth. The test result shown in Table 5.3, $p_{value} = 0$, shows that the bandwidth has significant influence on MOS. Moreover, beyond the 2 Mbps of bandwidth, the MOS scores are less influenced by high values of bandwidth.

Figure 5.6 (d) shows that at low values of bandwidth (0.5 Mbps) CI of opinion scores is minimized.

Figure 5.7: Visual impact of different bandwidth limitation. From left to right: original frame, 0.5 Mbps (PSNR 10.54 dB), 2 Mbps (PSNR 33.18dB), and 5 Mbps (PSNR 33.18 dB).

Table 5.3: ANOVA test results: bandwidth.

Bandwidth (Mbps)	F	$F_{critical}$	p-value
0.512, 1, 2, 3 and 5	73.437	2.64	0
2, 3 and 5	0.166	3.47	0.848

This result indicates that all the users perceive a similar quality for the videos. When the bandwidth becomes higher than 1 Mbps, the CI increases together with MOS scores. The high CI values indicate that, even at the same value of bandwidth and same video content, the quality perceived by different subjects is significantly different. Moreover, as can be also noticed from Figures 5.6 (b) at 0.5 Mbps the videos *Crowdrun* and *Soccer* show a similar MOS score while, at 2 Mbps of bandwidth, their scores are significantly different. It is likely that this behavior is due to the impact of the video content on perceived quality. Since, for the same encoder and impairments, the video quality appears different for different video content [15].

These results indicate that for low values of bandwidth the perceived quality is mainly influenced by the bandwidth. For high values of bandwidth (> 2 Mbps), the perceptual quality is also influenced by video content[131].

Finally, from the performed analysis, we can conclude that for low levels of transmission impairments (all the considered delay values, high values of bandwidth and low values of jitter and PLR) the perceived quality is significantly different from the test videos. Therefore, in Section 5.3 the influence of video content on the perceived quality is presented.

5.2 Study of the effect of video content on quality of experience

5.2.1 Effect of video content

To analyze the effect of video content on perceived quality, opinion scores collected from the subjects for the reference videos are considered. The distribution of opinion scores for the reference videos is shown in Figure 5.8 (a). The histogram, Figure 5.8 (a), indicates that most of the opinion scores given for the reference images are distributed over the range 3 to 5. Moreover, the summary of the opinion scores for eight different reference videos is plotted in Figure 5.8 (b) and it shows that the median score for each image is different and the box distribution is not uniform. This result indicates that the perceptual quality of reference videos is varying.

The result is further confirmed by the analysis of the perceived quality of the reference videos (without considering impairments) as shown in Figure 5.8 (c): the MOS scores are significantly different even in the reference videos. The level of significance is also tested by an ANOVA test: $p_{value} = 0$, indicates that the null hypothesis has to be rejected, thus confirming that the perceived
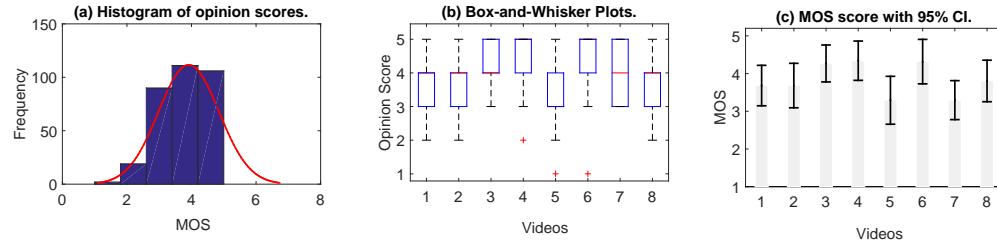
Figure 5.8: Perceived quality of reference videos.

quality is significantly influenced by the video content.

5.2.2 Effect of video content in the presence of impairments

The influence of the adapted values of impairments to the effect of content on QoE is presented in the following.

Delay

Figure 5.1(b) shows the MOS scores are not significantly changed for high values of delay. The confidence interval, shown in Figure 5.1 (d): computed based on the opinion scores given to the SRCs for the different values of belay artifact, is almost uniform. This result indicates that the introduced delay artifact does not influence the impact of content on QoE.

Jitter

As mentioned before, in Section 5.1.2, from Figure 5.2 (a) and (b) we can notice that the MOS scores are significantly different for the SRCs at 1ms of jitter, and the difference in the scores (also shown in Figure 5.2 (d)) is reducing for high values of jitter, and as jitter is 5ms, all the SRCs have the same MOS score. This result indicates that, at low values of jitter artifacts the QoE is different based on the presented video content (SRCs). However, at high values of jitter, the QoE is not much varied based on the content. From another viewpoint, at low levels of channel distortion the QoE is significantly influenced by the presented content, and the influencing level of content to the QoE reduces for highly distorted channel, i.e. if the channel is highly distorted, no matter with the content/SRCs, the QoE is low.

Packet loss

By following the results presented in Section 5.1.3, Figure 5.5 (b) show the difference in MOS for the SRCs is decreasing for high values of PLR. The same result is also shown in Figure 5.5 (d): difference in the scores measured in terms of the confidence interval is reduced for high values of

PLR. This result indicates the impact of video content on QoE is decreasing for high values of packet loss artifact.

Bandwidth limitation

As the impact of jitter and PLR artifact, the similar trend on the result: for high values of bandwidth (>2) the MOS scores are more separated for the SRCs, is shown in Figure 5.6 (b). The Figure 5.6 (d) also indicates the difference in opinion scores is reduced for the channel with low bandwidth values. This result indicates, if the channel has already been suffered by limited bandwidth artifact, the QoE is poor no matter with the video content. On the other hand, if the channel has a fair enough bandwidth (>1 Mbps) the QoE is also influenced by the content type (SRCs).

5.3 Analysis of the effect of video content related attributes on quality of experience

For this analysis, first content related low level attributes are selected, and the correlation analysis between the attributes and corresponding MOS scores of the SRC is performed.

Selected content related attributes

As mentioned before, in Section 3.2.1, it is not easy to map the extractable video features into semantic concepts, though the effective way for video content analysis is to use the attributes extractable from the sources [52]. Therefore, to describe the video content the SI, TI, CF, contrast, hue, saturation, color value, and brightness are used.

5.3.1 Study on ReTRiEVED video quality dataset

For the study, content related attributes are computed for eight reference videos. The opinion scores collected for the reference video sequences are used during the analysis. The correlation coefficient between the attribute and corresponding MOS score for the sequences is computed and plotted in Figure 5.9.

The Figure 5.9 shows that SI of the video is significantly correlated with the perceived quality. Moreover, the contrast, brightness and saturation are also strongly correlate with the perceived quality. However, PLCC and SRCC scores are not very high ($\simeq 1$) for all the features. These results indicate that none of the considered descriptors are sufficient to explain the video QoE.

Moreover, this result is further confirmed with the help of PCA, Figure 5.10. It shows that the content descriptors are well distributed over the space. The brightness has a small angle with MOS, and in the opposite plane SI, contrast and saturation also have a small angle with MOS.
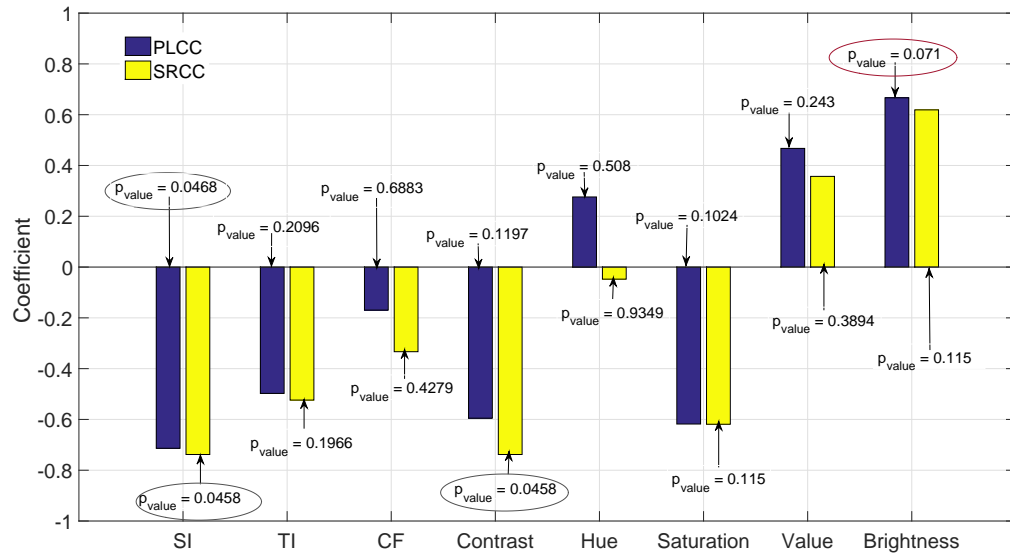
Figure 5.9: Correlation between the perceived video quality and content attributes.
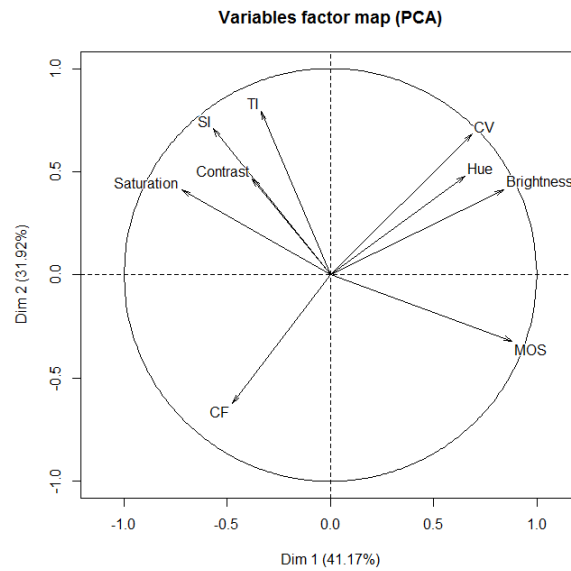
Figure 5.10: Results of PCA: analysis of video content attributes and their impact on perceived quality.

Furthermore, Figure 5.10 shows that first two principal components roughly explain about 73% of the total perceptual quality in the standardized rating.

Though, the number of SRCs considered for the subjective study is enough as specified in ISO 20462 standard [121]. However, the inclusion of a large number of SRCs has been always beneficial to overcome the over-and/or under-fitting problem. Therefore, to study the impact of video content on QoE, another subjective experiment is conducted. The subjective experiment and achieved results are briefly presented in the following.

5.3.2 Extended dataset: to study of the impact of video content related attributes on quality of experiance

In the following, video quality dataset and the correlation between video content descriptors and corresponding video QoE are presented.

Source sequences

In this experiment, 26 uncompressed SRCs have been selected based on the content varies (types, content, frame size, motion, rate, textures, color temperature, camera movements, etc.). The list of the selected videos is reported in Table 5.3.2, and sample frame of the videos is shown in Figure 5.11. The heterogeneity of the video content has been characterized by means of a wide span of spatial-temporal perceptual information, as shown in Figure 5.12. All the SRCs are taken from Consumer Digital Video Library (CVDL)[4].

Experiment

ACR method is selected as a subjective assessment protocol; and stimuli are rated from one to five (poor to excellent). The stimuli are shown to the subjects on a DELL U2413 digital monitor (resolution: 1920x1200 pixels) by using the VLC media player (version 2.1.3).

The experiment has been conducted in a controlled environment in order to produce reliable and reproducible results by avoiding involuntary influence of external factors. To compensate the effect of a potential bias based on the order of the test sequences in the averaged results, stimuli have been shown to the subjects in pseudo-random order according to their content and resolution. Effect of viewer fatigue is minimized by scheduling a short experiment session of 10 minutes.

For the experiment, 21 subjects were used for evaluating the QoE. Subjects were drawn from a pool of undergraduate to post-doctorate students from Università degli Studi Roma TRE. Most of the students were relatively naive concerning video processing related terminology and they were asked to wear any vision correcting devices (glasses or contacts) that they normally wear.

Verbal instruction was given to the subjects and it is followed by a training session to make them familiar with the assessment procedure and to establish an annoyance value range. The videos used for the training session were different from the test videos in order to remove the bias due to memory effect.

SRCs	Name	CIF	VGA	HDTV
SRC1	*Bennet Watt BeeClose*	CIF60fps	VGA60fps	1080i30fps
SRC2	*Bennet Watt BeeZoom*	CIF60fps	VGA60fps	1080i30fps
SRC3	*Bennet Watt CattleDogs*	CIF60fps	VGA60fps	1080i30fps
SRC4	*Bennet Watt DecantWine*	CIF60fps	VGA60fps	1080i30fps
SRC5	*Bennet Watt ElephantZoom*	CIF60fps	VGA60fps	1080i30fps
SRC6	*Bennet Watt FlockSunset*	CIF60fps	VGA60fps	1080i30fps
SRC7	*NTIA FlamencoShoes*	CIF60fps	VGA60fps	1080i30fps
SRC8	*NTIA TheFootPan*	CIF60fps	VGA60fps	1080i30fps
SRC9	*Bennet Watt CPSPIR003 original*			1080p30fps
SRC10	*broadcast BGRrevisedopen original*			1080i60fps
SRC11	*intel bostonwide original*			1080i60fps
SRC12	*mpr BusA3 original*			1080i60fps
SRC13	*mpr CellBlockB3 original*			1080i60fps
SRC14	*ntia cropduster original*			1080i60fps
SRC15	*ntia purple3e original*			1080p25fps
SRC16	*ntia TeaCupPolice original*			1080i60fps
SRC17	*ntia tulip1e original*			1080p25fps
SRC18	*ntia tulip2e original*			1080p25fps
SRC19	*ntia tulip3e original*			1080p25fps
SRC20	*pscr burnB-0-25-58-BC-TargetPerson original*			1080i60fps
SRC21	*pscr stands36h-ver1 original*			1080i60fps
SRC22	*vqeghd1 src02 hrc06*			1080p30fps
SRC23	*vqeghd1 src03 hrc09*			1080p30fps
SRC24	*vqeghd1 src06 hrc01*			1080p30fps
SRC25	*vqeghd1 src09 original*			1080p30fps
SRC26	*vqeghd2 csrc12 original*			1080i60fps

Table 5.4: SRCs includes 8 CIF videos (352 × 288, progressive and 60fps), 8 VGA (640 × 480, progressive, and 60 fps) and 24 HDTV (1920× 1080, interlaced or progressive, and 25 or 30 or 60 fps, and expressed as a 1080i60fps: 1080 is the size, i indicates interlaced, and 60 represent frame rate).

Data processing and results

Once opinion scores were collected, an outlier detection procedure (as presented in Section 3.2.4) was applied to detect and remove the scores given by subjects whose scores deviate strongly from other observers or mean score. In this experiment, scores given by all the subjects were accepted for the analysis.

The computed MOS scores with confidence intervals is plotted in Figure 5.13. It shows that the MOS is different for test video sequences; even for the sequences with the same resolution and frame rate. Level of significance on the difference in the scores, is tested with the help of ANOVA at a significance level of 5%, and null hypothesis (H0): group of means (MOS scores) is equal. The result of the test ($p_{value} = 0$) indicates that the null hypothesis cannot be accepted. This result indicates the MOS scores are significantly different for the SRCs. From this result, we can conclude that the

| (a) SRC1 | (b) SRC2 | (c) SRC3 | (d) SRC4 | (e) SRC5 |

| (f) SRC6 | (g) SRC7 | (h) SRC8 | (i) SRC9 | (j) SRC10 |

| (k) SRC11 | (l) SRC12 | (m) SRC13 | (n) SRC14 | (o) SRC15 |

| (p) SRC16 | (q) SRC17 | (r) SRC18 | (s) SRC19 | (t) SRC20 |

| (u) SRC21 | (v) SRC22 | (w) SRC23 | (x) SRC24 | (y) SRC25 |

| (z) SRC26 |
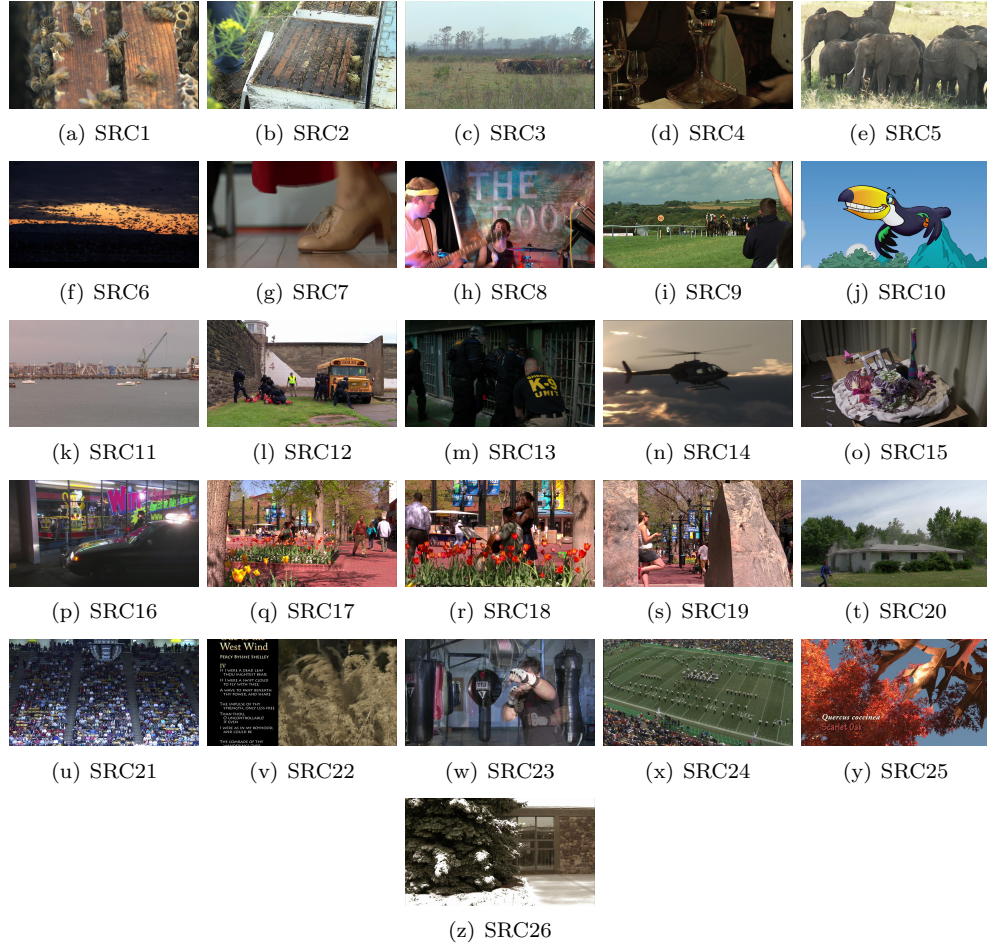
Figure 5.11: Sample frame of the SRCs.

QoE is significantly influenced by the video content, itself, indifferently with the frame size.

Content attributes

To analyze the impact of video content on QoE, in addition to the content attributes: SI, TI, CF, hue, saturation, value, and brightness, the global motion coefficient– to describe the video motion, are computed. Where, the global motion coefficient is computed by using the three-step-search algorithm [202].The global motion characterization measure is defined as (5.1):

$$G = \frac{|E - M|_{ave}}{1 + M_{ave}}, \tag{5.1}$$

where $|E - M|_{ave}$ and M_{ave} can be obtained by averaging $|E - M|$ and M over the frames of video sequence. Furthermore, M and E are the mode and mean of the motion vector magnitudes
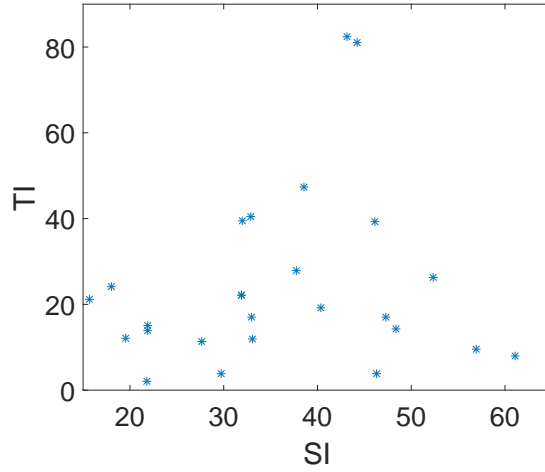
Figure 5.12: Distribution of SRCs over spatial-temporal plane. The SRCs are well distributed over the spatial-temporal plane. Meanwhile, two videos with very high temporal information are considered.
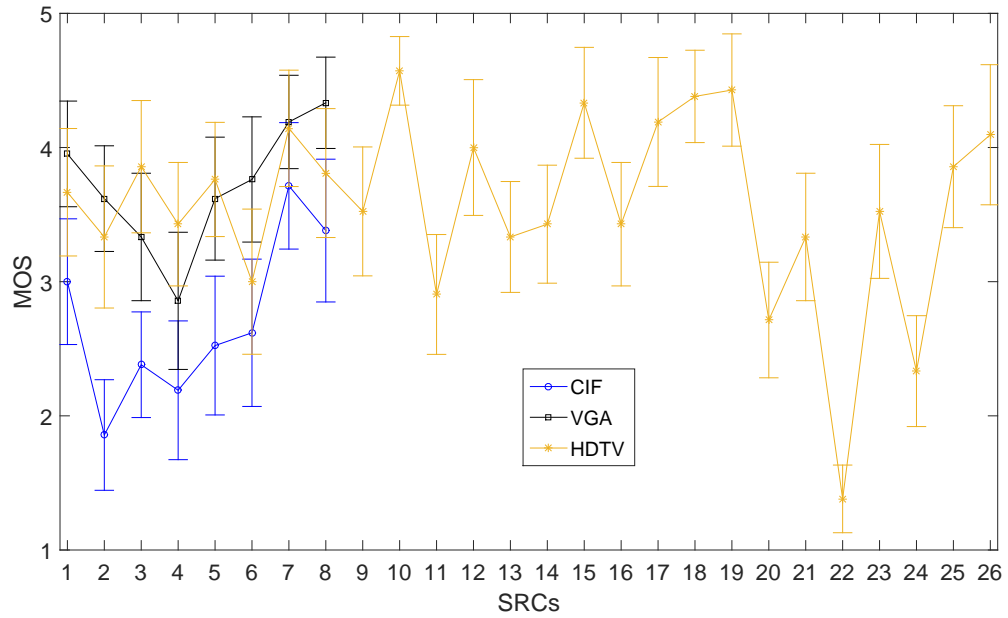
Figure 5.13: Subjective quality scores with 95% of confidence interval.

corresponding to two consecutive frames and can be computed by equation (5.2) and (5.3):

$$M = mode_{(i=1,2...m)}\{(M_{x(i)})^2 + (M_{y(i)})^2\}^{\frac{1}{2}}, \tag{5.2}$$
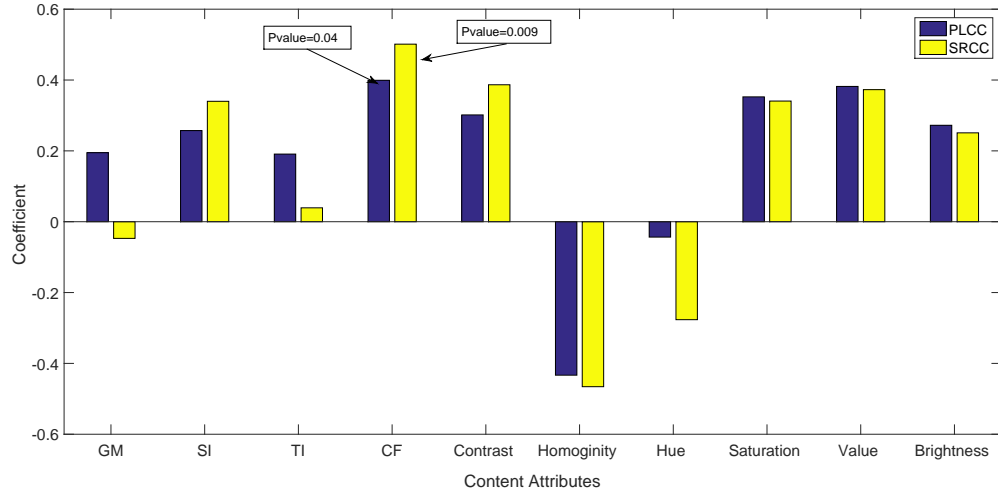
Figure 5.14: Correlation of the content attributes with video QoE.

$$E = \frac{1}{m} \sum_{i=1}^{m} \{(M_{x(i)})^2 + (M_{y(i)})^2\}^{\frac{1}{2}}, \tag{5.3}$$

where, $M_{x(i)}$ and $M_{y(i)}$ are the horizontal and vertical motion vector component of motion vector i and m is the number of motion vector per frame.

Results

The results of the correlation analysis (performed between the video content attributes and video QoE) are as follows:

- *Frame size:* Figure 5.13 shows that the QoE of VGA video is significantly higher than for CIF. Similar trends in the result is also shown for HDTV and VGA. However, in some cases the subjects are not able to find the difference, as a result for some videos the MOS score is higher for VGA than for HDTV. This result could be the effect of buffering time (the delay) taken by VLC player for displaying the HDTV video.

- *Spatial information:* The result of the correlation analysis, the correlation coefficient computed between SI of the videos and MOS values, shown in Figure 5.14, indicates there is no strong evidence to conclude that the SI is highly correlated with the video QoE. Similar trends in the result is also confirmed by the results presented in Figure 5.15: the angle of SI in respect of MOS is significant for all the videos (CIF, VGA and HDTV).

(a) CIF (8 SRCs)

(b) VGA (8 SRCs)

(c) HDTV (24SRCs)
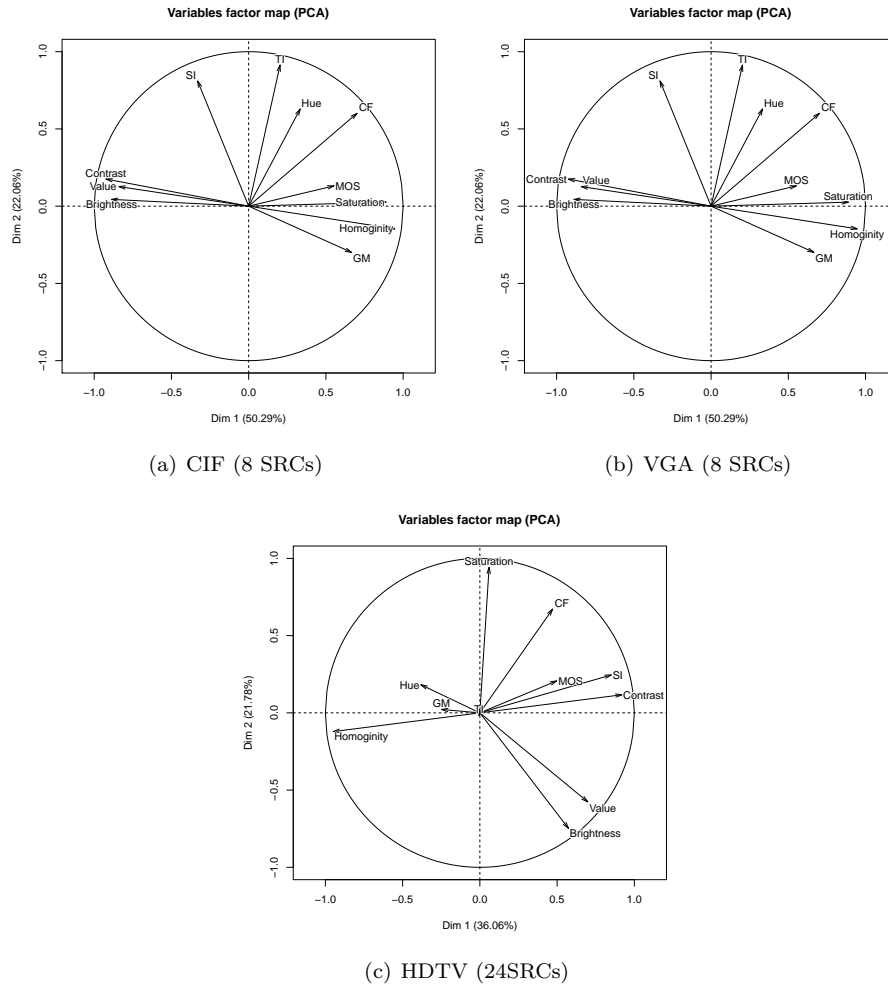
Figure 5.15: Result of PCA.

- *Temporal information:* Similar to the SI, the Figures 5.14 and 5.15 show that there is no evident relationship between the video QoE and temporal perceptual information of the video. Similar results was concluded for the TI in Section 5.3.1.

- *Motion:* As expected, the result of TI is followed by the content attribute motion, which is computed as global motion coefficient to express motion of the video. Our conclusion, about the impact of motion on QoE: motion does not have a direct correlation with QoE, is not far from the results presented in [72].

- *Homogeneity and Contrast:* The correlation between scene homogeneity and contrast with the perceived quality of video is presented in Figure 5.14; it shows that there is a noticeable correlation wit QoE, and in comparison with contrast the homogeneity has a higher correlation.

The similar trend in the result is also shown in Figure 5.15, the angles with respect to MOS are not much bigger for the videos with different frame size.

- *Hue, saturation, and value:* The impact of *hvs* indicators to QoE is analyzed. The achieved results, presented in Figures 5.14, show that there is no high correlation with the QoE for all the indicators. However, the saturation and value (color information) have a relatively noticeable correlation with the video QoE compared to hue. The result is confirmed by the result of PCA, shown in Figure 5.15.

- *Brightness:* Similar to the *hvs* indicators, the brightness is also not directly correlated with the video QoE. This result of brightness is also presented in Figures 5.14 and 5.15.

- *Colorfulness (CF):* The correlation between the QoE and colorfulness of the video is shown in Figure 5.14: it indicates the correlation coefficient is not significant. The similar result (noticeable angle between CF and MOS) is also shown in Figure 5.15. However, the CF has a smaller angle to the MOS compared to most of the other attributes, and this result is uniform for the videos with different frame size. This result indicates that there is some correlation between CF and QoE, but the amount of correlation is not significantly high.

  From another viewpoint, the provided medium level of correlation has a meaning, because the chroma component of the image has a Gaussian distribution like relationship with the naturalness of the image, and the image naturalness has a strong correlation with image QoE. Moreover, the relationship between image quality and colorfulness is also like a Gaussian distribution curve, i.e. the visual signal with medium colorfulness poses high QoE compared to very low and very high colorful signal [62].

5.4   Conclusion

In this chapter, the influence of the impairments and video content on the perceived video quality is analyzed.

In brief, the results show that:

- the adopted initial delay does not have a significant impact on the perceived video quality, while jitter, PLR, and bandwidth do. On the performed experiments, the values 2 Mbps of bandwidth, 2 ms of jitter, and 3% of PLR can be considered as JND limits since above and below these thresholds the perceived video quality changes significantly;

- if the communication channel is influenced significantly by impairments (high values of PLR and jitter, and low values of available bandwidth) the perceived quality is mainly dependent on channel conditions. Therefore, the perceived quality is more dependent on the impairments rather than the video content and human factors;

- if the channel is less influenced by impairments (low values of PLR and jitter, and high value of available bandwidth), the perceived video quality is also dependent on video content and human factors.

The key video quality influencing factors or content attributes - spatial perceptual information, temporal perceptual information, motion, colorfulness, contrast, hue, saturation, color value (value), and brightness - are used to analyze the impact of the video content on the perceived video quality. The results show that, the considered attributes do not have a direct and linear correlation with the video QoE. From another viewpoint, the single attribute cannot characterize the overall video QoE.

Chapter 6

Video quality of experience metric

Major contributions of this chapter are two folds:

- the performance evaluation of QoS to QoE models: QoE estimation by exploiting the QoS parameters, is performed; and

- a blind VQM is proposed, and performance of the metric is analyzed and compared with the state-of-the-art metrics.

6.1   Introduction

As mentioned before, the QoE can be assessed by two approaches: subjective and objective. The only most appropriate method for assessing perceived video quality is the subjective one. However, the subjective method have many shortcomings such as expensive, time consuming, not suitable for automatic, in-loop/service, and on-line real-time processing, and problem of reproducibility of the results, because of it is depending on the viewers physical conditions, emotional states, personal experience, context, etc. Therefore, a common goal is to devise an objective measure to emulate the subjective score i.e. MOS.

As a result of the failure of the traditional quality measures, in this visual quality assessment field, many ongoing efforts have been given to devise the metric. The basic reasons for the failure are: not every change in an image is noticeable; not every pixel/region in an image receives the same attention level; not every change leads to distortion (otherwise, many edge sharpening and post-processing algorithms would have not been developed); not every change yield a same extent of perceptual effect with a same magnitude of change (due to spatial/temporal/chrominance /masking) [141].

In general, according to the ITU, objective quality assessment models can be categorized into five types: media-layer models, parametric packet-layer models, parametric planning models, bitstream layer models, and hybrid models (as shown in Table 6.1) [84].

Models	I/P Information	Applications	Standards	Limitations
Media-layer model	media signals (video) by exploiting HVS	quality benchmark	ITU-T J.144 [104] ITU-T J.148 [105]	obtaining media signals at the network mid-point is difficult but payload of packets can be decoded
Parametric packet-layer model	packet header information	non-intrusive monitoring (network probe)	ITU-T P.1201 [106]	difficult for evaluating the content dependence of QoE
Parametric planning model	quality design parameters	network planning, terminal/application designing	ITU-T G.1070 [102]	requires a priori information about the system under testing
Bitstream layer model	Packet header and payload information	in-service non-intrusive monitoring (terminal-embedded operation)	ITU-T P.1202 [107]	high complexity, since also used content characteristics from the coded bit-stream
Hybrid model	combination of any	in-service non-intrusive monitoring	ITU-T J.343 [103]	complexcity

Table 6.1: Objective QoE assessment models.

In literature, in general, mostly two techniques have been used to estimate the QoE: by using the concept of fidelity measure and HVS system, and by exploiting the quantitative relationships between the user QoE and network QoS. In comparison to first approach, the latter approach is simple and effective from the network service view point: it uses the system recorded QoS parameters.

6.2 Quality of experience estimation by using the quality of service parameters

6.2.1 Introduction

Qualitative relationship between QoE and QoS

The qualitative semantic relationship between the QoS parameters and user QoE is shown in Figure 6.1. User QoE as a function of QoS disturbance is divided in several areas [63]:

- Constant sub-optimal QoE (no effect): For a vanishing QoS disturbance the user considers the QoE equivalent to that of the reference. This result indicates, a slight growth in the QoS disturbance does not affect the user QoE. For an example, the QoE does not reduce significantly, if we reduce the channel bandwidth from 5 Mbps to 3 Mbps [175].

- Sinking QoE (user distorted): When the QoS disturbance exceeds a certain threshold, lets say JND1, the former quasi-optimal QoE level cannot be maintained anymore, and thus, user

satisfaction sinks for increasing values of QoS distortion.

- Unacceptable QoE (user gives up): As soon as the QoS disturbance reaches threshold, JND2, the outcome of the communication networks becomes unacceptably bad and the QoE is poor.
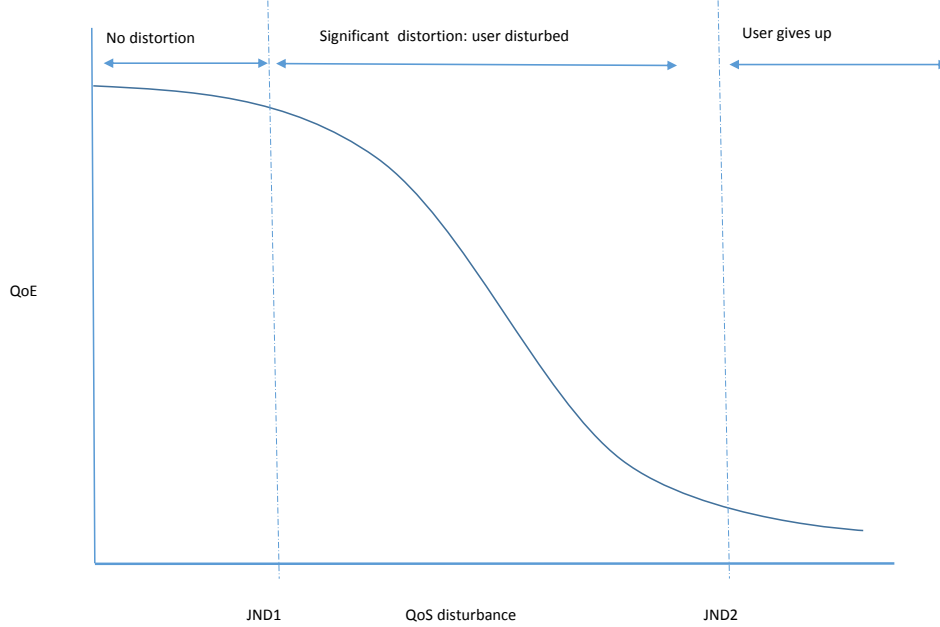
Figure 6.1: General shape of the mapping: qualitative relationship between QoS and QoE.

A basic idea of this relationship is introduced in 1834 by German physiologist Ernst Heinrich Weber; the idea was the operation of the human sensory system traced back to Just Noticeable Differences (JND). This fundamental principle relating human perception to the relative change of stimulus has later been extended by Gustav Fechner, what is known today as Weber-Fechner law [197]. Mathematically, the differential perception (dP) is directly proportional to the relative change (dS/S) of the physical stimulus:

$$dp = k.\frac{dS}{S}, \tag{6.1}$$

by taking integration both sides, Equation 6.1 becomes

$$p = k.ln\frac{S}{S_0}, \tag{6.2}$$

where, p identifies the magnitude of perception, and the constant of integration S_0 can be interpreted as stimulus threshold. This principle is a breakthrough for developing the mapping model

between QoS (as physical stimulus) and QoE (as perceived intensity). In literature, many mapping models are devised to estimate the perceptual quality (QoE) from the physical stimulus such as QoS parameters and signal distortion measures.

QoS to QoE mapping models

QoE can be estimated by using the measure of network QoS related parameters: delay, jitter, PLR, throughput, etc. [38, 35]. Measuring QoS parameters are quite easy for the operators; however, it could not represent the actual perceived quality by the user [184]. So it is a fundamental importance to develop a mapping model to predict QoE from QoS parameters.

In literature many mapping models have been presented. The most common mapping model is the linear model; Equation 6.3, where x is the physical stimulus (QoS parameter). However, the performance of this model is very far from the reality, and thus non-linear models are used.

$$QoE = a + bx, \tag{6.3}$$

The commonly used non-linear mapping functions are: cubic function (Equation 6.4), logistic functions (a: Equation 6.5, b: Equation 6.6, c: Equation 6.7, and five parameter function: Equation 6.8), exponential function (Equation 6.9), power function (Equation 6.10), and logarithmic function (Equation 6.11) [77, 80].

$$QoE = a + bx + cx^2 + dx^3, \tag{6.4}$$

$$QoE = a + \frac{b}{1 + c(x + d)^e}, \tag{6.5}$$

$$QoE = a + \frac{(b - a))}{1 + \exp[-c(x + d)]}, \tag{6.6}$$

$$QoE = \frac{a}{1 + exp[-b(x - c)]}, \tag{6.7}$$

$$QoE = a + \frac{d}{(1 + (\frac{x}{c})^b)^e}, \tag{6.8}$$

$$QoE = a.exp(b.x) + c.exp(d.x), \tag{6.9}$$

$$QoE = a.x^b + c, \tag{6.10}$$

$$QoE = -a.log(x) + b, \tag{6.11}$$

Moreover, other many correlation approaches are presented in literature [14]. Among them a most popular single parameter model is an IQX Hypothesis (Equation 6.12), presented in [63].

$$QoE = a.exp(-b.x) + c, \tag{6.12}$$

where, a, b, and c are the positive parameters.

Another popular multi-parameter model, called VQM(VQM) (Equation 6.13), is presented in articles [181, 100, 244].

$$QoE = f(x_1, x_2, x_3, ...x_n), \tag{6.13}$$

where, $x_{1,2,......n}$ are the QoS parameters.

For better performance, more complex models are presented by using statistical analysis and machine learning [58, 155]. However, in this dissertation, the performance of the single parametric models is evaluated on the dataset.

6.2.2   Performance evaluation of the quality of service to quality of experience models

The QoS parameters (packet loss, delay, jitter, throughput, etc.) result quality problems like glitches, artifacts, excessive waiting times, etc. For this reason, the parameters are considered as the Key Performance Indicators (KPIs). It is necessary to investigate generic relationships between QoE and KPIs, and the relationship is expressed in terms of mapping models.

In [126, 14] the authors discuss and compare the performance of widely used mapping functions. A logarithmic function to predict QoE from network related parameters, bit rate, packet loss rate, jitter, etc., is presented in [197]. In [189], a perceived VQA technique exploiting video coding rate and packet loss rate is presented. QoE mapping and adjustment model form QoS parameters for cloud based multimedia services by considering delaying, packet loss rate, jitter and throughput has been discussed in [94]. Moreover article [124] proposes a QoE assessment model for video streaming service based on the different weight assigned to PLR, jitter, delay and bandwidth. The performance of IQX hypothesis is analyzed in [92].

The above mentioned works have the following shortcomings: i) interactions between key QoS parameters and their effects on QoE is poorly defined; ii) study is performed to the less number of videos and subjective scores–results over or unbefitting problems; and iii) more efforts are need to benchmark the models. Thus, the scope of this dissertation is to benchmark the widely used models on a dataset by using key QoS parameters.

Performance analysis tool

Performance analysis and selection of the models is based on their fitting capability with subjective scores expressed as MOS. Fitting capability is measured in terms of the following statistical measures: Sum of Squares due to Error (SSE), R-Square (r^2), Adjusted R-Square ($adjusted\_r^2$) and Root Mean Squared Error ($RMSE$). Based on these statistical measures single decision variable (μ) is defined in Equation 6.14. The high value of μ indicates better fitting for the given data.

$$\mu = \frac{r^2 \times adjusted\_r^2}{SSE \times RMSE} \tag{6.14}$$

Table 6.2: Performance evaluation of the mapping models.

Mapping Functions	PLR (μ)	jitter (μ)	bandwidth (μ)
1. Linear Mapping (Equation 6.3)	0.0164	0.0141	0.0080
2. Cubic Polynomial (Equation 6.4)	0.0353	0.0774	0.0271
3. Logistic (a) (Equation 6.5)	0.039	0.0859	0.0332
4. Logistic (b) (Equation 6.6)	0.0095	0.0024	-4E-06
5. Logistic (c) (Equation 6.7)	0.0031	0.0263	0.0301
6. Exponential (Equation 6.9)	0.0227	0.0305	0.0069
7. Power (Equation 6.10)	0.0388	0.031	0.0184
8. Logarithmic (Equation 6.11)	0.0384	0.0320	0.0185
9. IQX Hypothesis (Equation 6.12)	0.0361	0.0828	0.0077
10. Five Parameter Logistic (Equation 6.8)	0.0323	0.0825	0.0273

Results

The result of the performed analysis is summarized in Table 6.2. From the analysis, we can conclude the following results:

- for PLR artifact, power function, logarithmic function, IQX hypothesis and logistic function gives quite similar results. For this dataset, among them, logistic function (a) has a higher value of the μ.

- for jitter artifact, logistic function (a), IQX hypothesis, and five parameter logistics function (a) give a close and accurate approximation. However, according to value of decision variable μ logistic function (a) gives optimal solution.

- for bandwidth limitation artifact, logistic function (a) , logistic function(c) and cubical function give the best approximation compared to other functions. Among them the logistic function (a) gives best approximation, since it has highest value of μ.

Finally, based on a decision variable logistic function (a) is a optimal solution for all the three parameters and, by considering the computational complexity, the IQX hypothesis is also the best

(a) Bandwidth limitation.   (b) Jitter.   (c) Packet loss
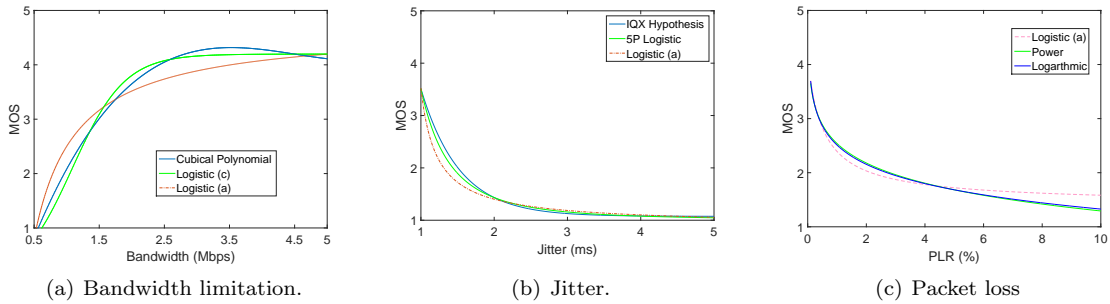
Figure 6.2: Mapping models for estimating the QoE by using QoS parameters.

alternative for PLR and jitter. This result is also confirmed by the mapping models plotted for the artifacts, shown in Figure 6.2. It shows that the performance of the mapping models are very close to each other, and thus, we can select any of them based on the complexity and considered artifacts.

6.3   Distortion based quality of experience estimation

6.3.1   Introduction

As specified before, the other way of estimating the QoE is by using the media signal (media-layer model). Accordingly, in this section, a QoE metric is proposed for video streaming services by using the received video: introduced distortions in the video.

Related works

In the state-of-the-art, many efforts have been devoted to estimating the video QoE. Most of them are investigating the effects of common network parameters on the received video. Authors in [258] studied the impacts of temporal jerkiness on video quality, which is caused by packet loss or late arrival of packets and, based on their findings, they proposed a neural network based VQM. Authors in [207] present an analytical model for no-reference VQM by taking into account both video play-out rate and network throughput. A VQM based on the spatio-temporal natural scene statistics and motion coherency in the video scenes has been proposed in [202]. A VQM based on the statistical estimation of PSNR of the coded transform coefficients for H.264/AVC encoded sequences is presented in [32] . By considering the features of H.264/AVC encoding, such as blocking, blurring, and spatial activity, a VQM is presented in [173] . In [20] a machine learning approach is recommended to estimate the QoE. The QoE is represented as an engagement, and expressed as a function of the quality metrics. The engagement could be the video play time, number of visits and the quality metric represent observed indicates such as buffering ratio, average bit rate, etc. In [206] a video quality estimation method is proposed by considering the policies applied for packet processing by

routers and the level of total network utilization. In [219] the authors propose a novel rate adaptation algorithm called QoE enhanced adaptation algorithm over Dynamic Adaptive Streaming over HTTP (DASH). The adaptation algorithm preserves the minimum buffer length to avoid interruption and minimizes the video quality changes during the playback. In [144] a multi-factor QoE evaluation model based on the content classification by spatial and temporal information for H.264/AVC encoded video has been proposed. Authors in [216] propose an acceptability-based QoE estimation model by considering encoding parameters, bit rate, video content characteristics, and mobile device display resolution. A QoE metric for HDTV by using PSNR metric and PLR artifact is presented in [99]. The authors in [63] present user perceived QoE prediction model from network related QoS parameters. Finally, [24] presents the performance comparison of QoS to QoE mapping models for wired and wireless communications.

The above mentioned state-of-the-art metrics have the following limitations: i) they are not specifically designed for QoE estimation for video streaming services, ii) they are complex and time consuming, iii) their performances are far from the subjective scores in a real scenario, and iv) they generally assume the presence of one main network artifact: in practical streaming services, this approach can lead to a wrong QoE estimation, since only part of data can be affected by one particular impairment.

6.3.2 Proposed No-Reference video quality of experience metric

The goal of this section is to propose a generic blind video QoE metric, called V_{QoE}, for streaming services. Since, the metric is designed for real time services, the frame by frame QoE estimation approach has been used. The frame quality is expressed by considering the portion of frames that, due to transmission errors, cannot be correctly decoded. These will be in the following addressed to as *broken blocks*. The overall video QoE is expressed as the average of the QoE estimated for all frames. The results show that the proposed method is computationally less complex and faster than the other considered blind video quality metrics and its performance is superior. The details of the proposed metric and achieved results are presented in the following.

The video QoE is computed in three steps: i) the total number of broken blocks for each frame is computed, ii) the QoE for each frame is estimated from the total number of broken blocks by using a mapping function between broken blocks and QoE, iii) the overall video quality is estimated as the average of frame quality.

Broken blocks estimation

The basic steps for the estimation of the number of broken blocks are detection and verification [23]. In brief, let F_k is the kth generic frame of the video sequence. It can be partitioned in $N_r \times N_c$ blocks $\mathbf{B}_k^{(i,j)}$ of $r \times c$ pixels with top-left corner located in (i,j). Moreover, $\Delta\mathbf{B}_k^{(i,j)} = \mathbf{B}_k^{(i,j)} - \bar{B}_k^{(i,j)}$

denotes the deviation of the luminance in k^{th} frame of the block $\mathbf{B}_k^{(i,j)}$ from the corresponding mean values.

The inter-block correlation $\rho_k^{B(i,j)}$ can be computed as:

$$\rho_k^{\mathbf{B}(i,j)} = \frac{\left\langle \Delta\mathbf{B}_k^{(i,j)}, \Delta\mathbf{B}_{k-1}^{(i,j)} \right\rangle}{\left\| \Delta\mathbf{B}_k^{(i,j)} \right\|_{L_2} \left\| \Delta\mathbf{B}_{k-1}^{(i,j)} \right\|_{L_2}}. \tag{6.15}$$

where $< \bullet, \bullet >$ denotes the inner product and $\| \bullet \|_{L_2}$ the L_2-norm.

To identify the distorted block, the blocks have been classified into three groups; low, medium and high content variation. The content variations are evaluated based on their temporal inter-block correlation $\rho_k^{B(i,j)}$. Moreover, based on the content variation groups, the corresponding variability map $\mathbf{\Gamma}_k^V = \{\Gamma_k^{VB(i,j)}\}$ has been defined by comparing the inter-frame correlation of each block as:

$$\Gamma_k^{VB(i,j)} = \begin{cases} 1, & if \ \rho_k^{B(i,j)} < \theta_l \\ 0, & if \ \theta_l \leq \rho_k^{B(i,j)} \leq \theta_h \\ 2, & if \ \rho_k^{B(i,j)} > \theta_h \end{cases} \tag{6.16}$$

where the two thresholds, θ_l and θ_h have been selected in order to grant $|P_{fa} - P_{md}| < \varepsilon_1$. Here P_{fa} is the probability of false alarm, P_{md} is the probability of missed detection, and ε_1 is a significantly small value. Value of the ε_1 is experimentally determined from the training session. In this study the values of θ_l and θ_h are set to 0.2 and 0.9.

The blocks with medium content variation, $(\Gamma_k^{VB(i,j)} = 0)$ are less likely to be broken. The blocks with low and high content variation should be further analysed, to find out whether they should be considered as broken blocks or not. For each block, if $\Gamma_k^{VB(i,j)} = 2$, it has to be further checked. If at least v blocks among the surrounding blocks present a strong temporal correlation, where the parameter v has been identified by experimental tests, then the block is classified as belonging to a static region. Then its potential distortion index $\Gamma_k^{CB(i,j)}$ is set to zero, in other words the block is not considered as a broken block. That is:

$$\Gamma_k^{CB(i,j)} = \begin{cases} 0 & if \ |\varsigma| > \nu \\ \Gamma_k^{VB(i,j)} & otherwise \end{cases} \tag{6.17}$$

where ς equals to how many surrounding blocks of $B_k^{(i,j)}$ with $\Gamma_k^{VB(p,q)} = 2$.

Finally, if $\Gamma_k^{CB(i,j)} \neq 0$, then further test is necessary. Let E_l and E_r be the L_1 norms of the vertical edges respectively on the left and on the right boundary of the block, and with A_c, A_l and A_r the average values of the L_1 norms of the vertical edges inside the current block and of the left

and right adjacent blocks. A block with $\Gamma_k^{CB^{(i,j)}} \neq 0$ is classified as affected by visible distortion if:

$$\left| E_l - \frac{(A_c + A_l)}{2} \right| > \theta \quad or \quad \left| E_r - \frac{(A_c + A_r)}{2} \right| > \theta \tag{6.18}$$

where the threshold θ has been defined on the basis of experimental trials (in this article $\theta = 100$). In particular it corresponds to JND (Just Noticeable Difference) collected and evaluated for 90% of subjects.

The same procedure is applied to the horizontal direction. If the block edges are consistent (i.e. no visible distortion has been detected along horizontal and vertical directions) $\Gamma_k^{CB^{(i,j)}}$ is reset to 0. As a result, total number of broken block (TOT_k) is computed as the total amount of blocks where $\Gamma_k^{CB^{(i,j)}} \neq 0$.

QoE estimation

After the total number of broken blocks in a frame is computed, the QoE has to be determined from the total number of broken blocks by using a mapping function. Based on its mapping capability and simplicity, IQX Hypothesis is selected to estimate the QoE from the total number of broken blocks. Hence, the QoE for k^{th} frame is estimated as:

$$QoE_k = a.exp(-b.(TOT_k)) + c \tag{6.19}$$

where a, b, and c are the regression parameters which can be achieved from the training session and TOT_k is the total number of broken blocks for k^{th} frame. The overall video QoE is considered as the average of the frame quality.

6.3.3 Results and discussion

To evaluate the performances of the proposed algorithm, the availability of a video quality database is important. In the state-of-the-art, many video quality databases have been recommended [65], among them for this study the EPFL-PoliMI VQAdatabase [51] and LIVE Video Quality Database [205] are selected. In literature, it is also noticed that the performance of the metrics is also varied on the datasets. Because, some of the metrics are tuned for a particular dataset, i.e. for the specific artifacts. Therefore, our metric is validated to the datasets: which are largely been used to benchmark the state-of-the-art VQA metrics.

For analysis, from the EPFL-PoliMI database 156 video streams (78 video sequences at CIF and 78 sequences at 4CIF spatial resolution), encoded with H.264/AVC and corrupted by simulating the packet loss due to transmission over an error-prone network and their corresponding subjective scores MOS are considered. In this database, in order to simulate burst errors, six different PLR (0.1%, 0.4%, 1%, 3%, 5%, 10%) patterns was used and every reference sequence has a two impaired

(a) *Crowdrun* Cb component

(b) *Ice* Cb component

(c) *Foreman* Cb component

(d) *Mobile* Cb component
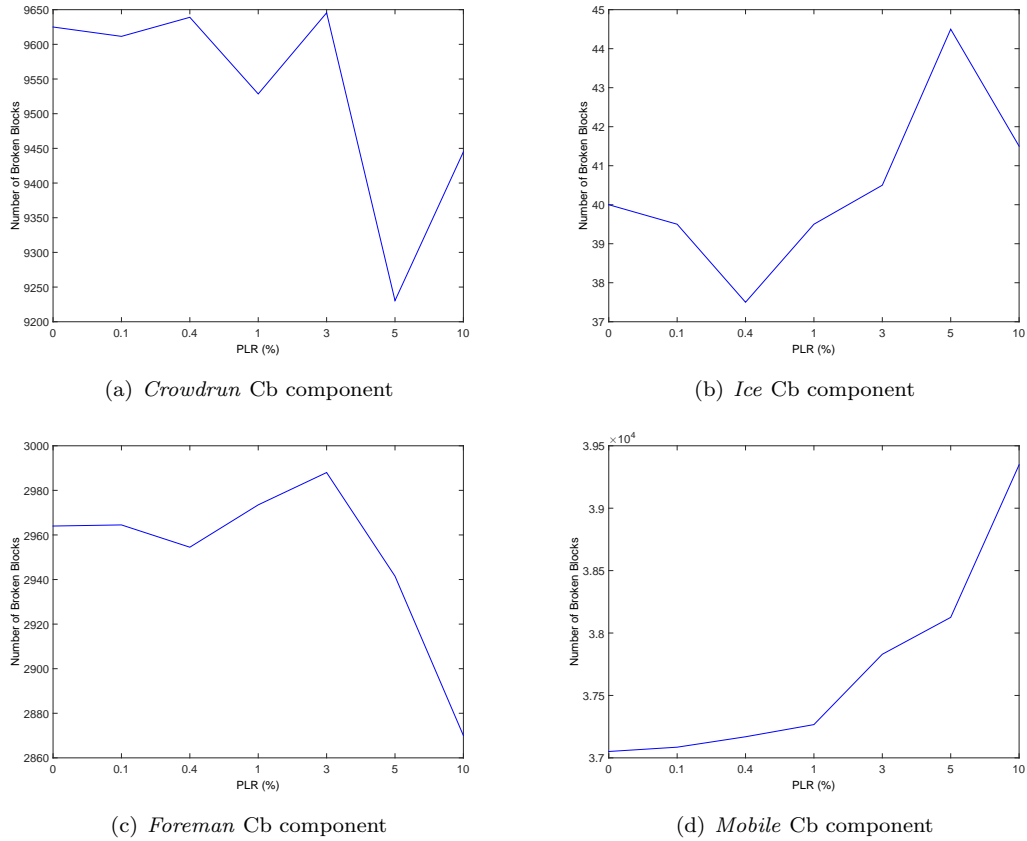
Figure 6.3: Total number of broken blocks of the Cb component for the 4CIF (*Crowdrun* and *Ice*) and CIF (*Foreman*, and *Mobile*) videos.

sequence for the same level of PLR. From the LIVE Video Quality Database, 80 videos (with a resolution of 768X432 pixels and H.264/AVC encoded) and their corresponding subjective scores (DMOS) are considered. Among all videos, 10 are the reference sequences, 40 were affected by the wireless distortions (four test videos per reference) and the remaining 30 were affected by IP distortions (three test videos per reference). The details about the databases are presented in [51] and [205] .

From the analysis of the achieved results, it can be noticed that mainly the luminance component of the video has a direct and significant impact on the video quality. For the sake of compactness in the following only the results of a subset of all considered videos will be included: the 4CIF video sequences *Crowdrun* and *Ice* and the CIF sequences *Foreman* and *Mobile* extracted from the EPFL-PoliMI database. The analysis has been performed on 52 impaired video sequences: the original *Crowdrun*, *Ice*, *Foreman* and *Mobile* sequences and 12 impaired ones obtained by considering six PLR values. For each video sequence, the relation between the PLR and total number of broken

(a) *Crowdrun* Cr component

(b) *Ice* Cr component

(c) *Foreman* Cr component

(d) *Mobile* Cr component
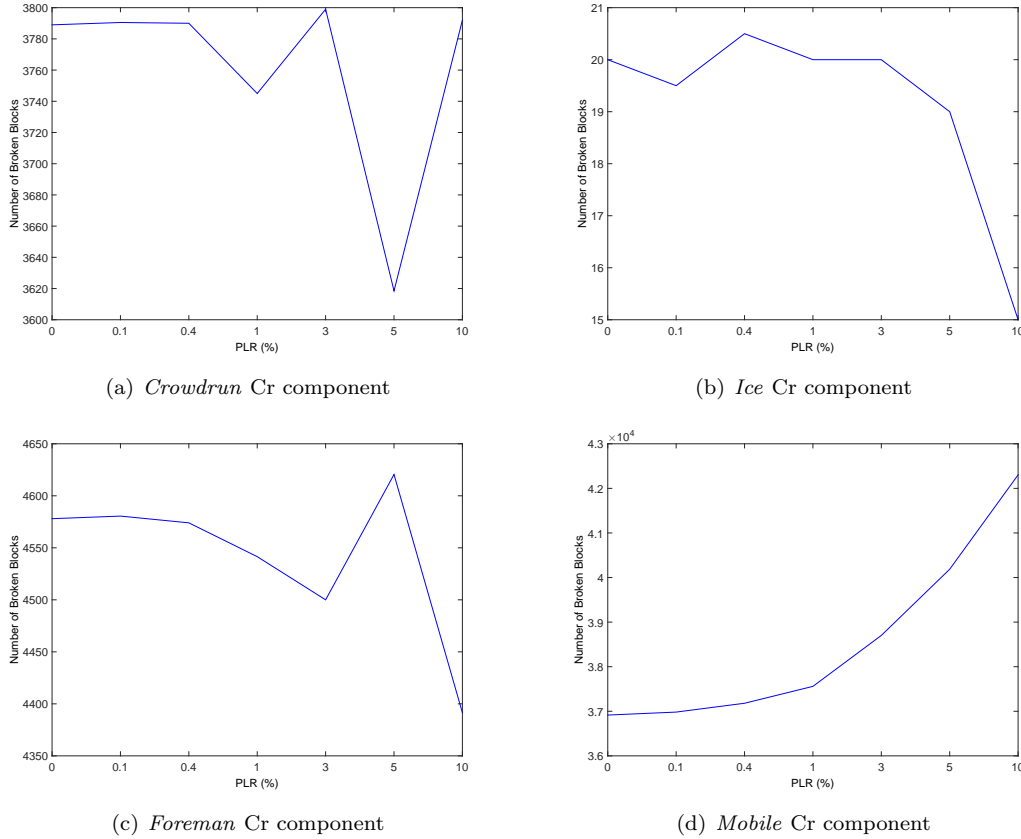
Figure 6.4: Total number of broken blocks of the Cr component for the 4CIF (*Crowdrun* and *Ice*) and CIF (*Foreman* and *Mobile*) videos.

blocks of each video chrominance component (Cb and Cr) is shown in Figures 6.3 and 6.4.

It can be noticed that the total number of broken blocks does not increase for higher values of PLR. However, for the luminance component Y, from Figure 6.5 it can be noticed that for high values of PLR correspond to the higher number of broken blocks, and this trend is confirmed for all the considered video sequences. From these results we can conclude that mainly the luminance component shows a closer and synclastic relationship between PLR and broken blocks. As demonstrated in [24] [63] this behavior has an impact on QoE. Based on these considerations, it is possible to reduce the computational complexity by considering only the luminance component for quality assessment.

In the following, the QoE computed through the proposed metric, V_{QoE} is compared with other widely discussed NR quality metrics: NIQE [160] , BIQI [162] , BRISQUE [159] , BLIINDS-II [201] and blind prediction of natural video quality [202] . Among them, the first four metrics have been initially designed for image quality estimation and lately used for video quality estimation by

(a) *Crowdrun* Y component

(b) *Ice* Y component

(c) *Foreman* Y component

(d) *Mobile* Y component
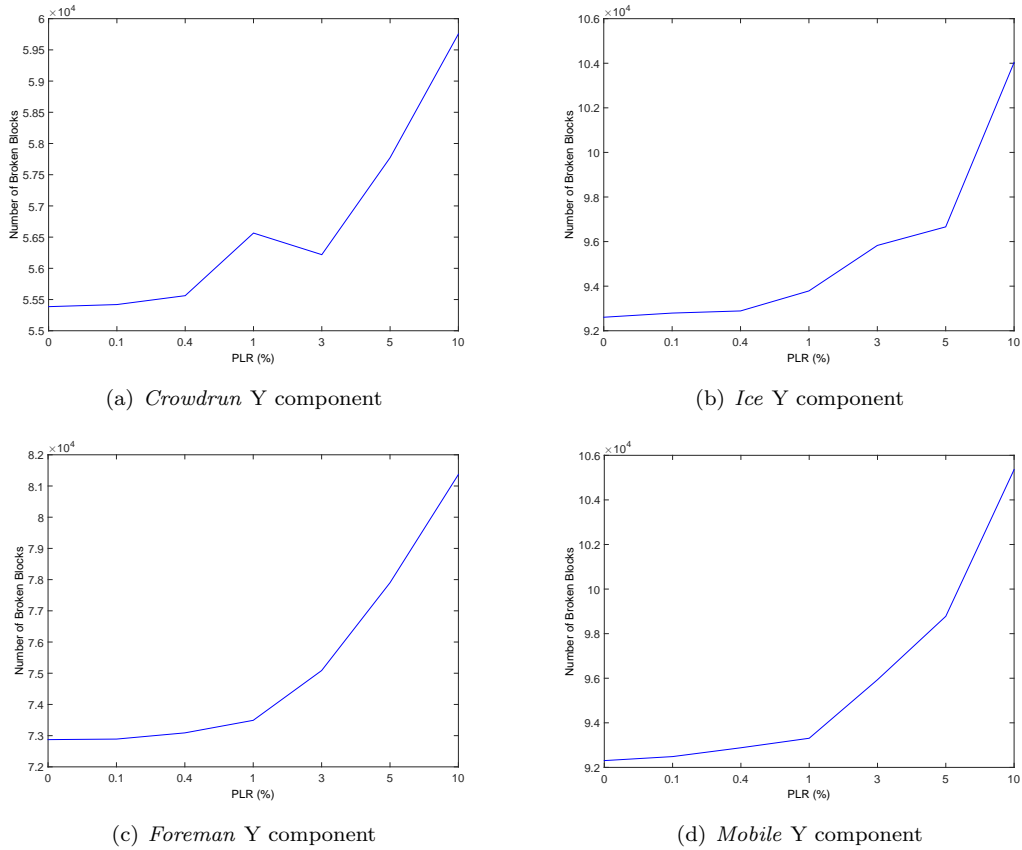
Figure 6.5: Total number of broken blocks of the Y component for the 4CIF (*Crowdrun* and *Ice*) and CIF (*Foreman* and *Mobile*) videos.

averaging the single frame pair quality score for all the video. More recently, these tools have also been used for the performance comparison of newly proposed video quality metrics [202].

To compare the performance of the metrics, SRCC and PLCC between the collected subjective scores and predicted scores have been computed. During the experiment 80% of the videos were selected randomly for the training and rest of the videos were used for testing.

The performance of the metrics for CIF and 4CIF videos from EPFL-PoliMI database are presented in Figure 6.6. The results show that the proposed method (V_{QoE}) has higher values of PLCC and SROCC compared to the considered metrics for both CIF and 4CIF videos. Besides that, the proposed method also shows higher PLCC and SRCC values than Video BLIND ($PLCC = 0.75$ and $SRCC = 0.807$ as mentioned in [202]) for the same database.

Moreover, the performances of the proposed method are also evaluated for the LIVE Video Quality Database. The results show that the metric V_{QoE} has $PLCC = 0.7909$ and $SRCC = 0.8571$. As presented in [202] for the same database but for all the video categories (MPEG-2, H.264, wireless,

(a) For CIF videos   (b) For 4CIF videos
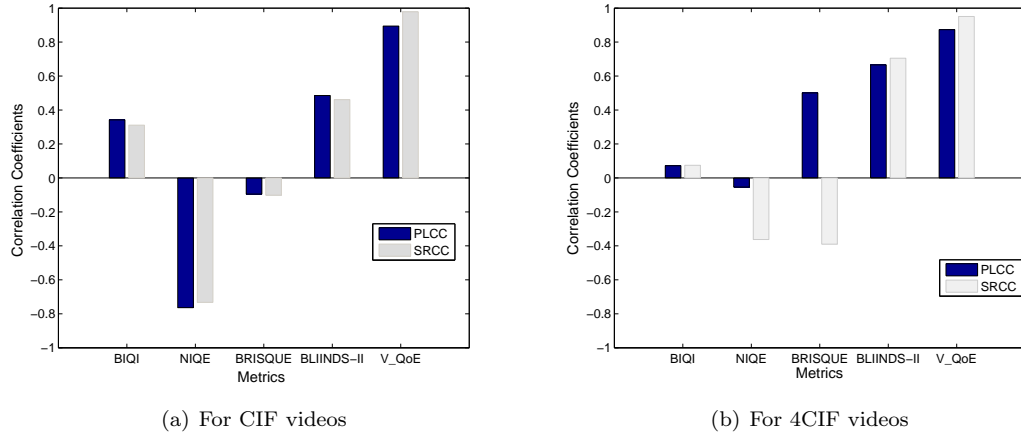
Figure 6.6: Performance comparison of the proposed metric(V_{QoE}) with considered state-of-art metrics.

and IP) the performance of Video BLIND is $PLCC = 0.881$ and $SRCC = 0.759$.

For real time video streaming services, computational complexity and processing time of QoE metric is as important as its QoE prediction capability. The considered state-of-the-art metrics, including Video BLIND is more complex, and need more processing power and computational time. This is mainly due to the fact that the metrics have been designed by considering computationally heavy and complex techniques [41] , as generalized Gaussian density parameter estimation techniques and motion coherency computation. In the proposed metric V_{QoE}, only luminance component is considered and thus, the overall processing time is depending mainly on the block-wise correlation operation only.

6.4 Conclusion

In this chapter, first, performance of QoE estimation (from QoS parameters) models is evaluated to the ReTRiEVED video quality dataset. The achieved results indicate among the considered models only few models performs better, and the performance of the models is depending on introducing artifacts. Followed by the discussion, sub-optimal models are recommended based on the artifacts. By considering the performance for all the considered impairments and complexity, the IQX hypothesis is selected as an optimal model, in general.

Next, a NR and real-time video QoE metric V_{QoE} is proposed. The video quality is computed based on the video distortion, which is measured as the number of broken blocks. The frame QoE has been derived from a number of broken blocks by means of IQX hypothesis. The overall video quality is expressed as an average of frame QoE. The performance of the proposed method is compared with widely discussed blind image/video metrics. The result shows that the proposed metric outperforms

the considered state of art metrics, and it is also faster.

Chapter 7

Light field imaging

In this chapter, a brief introduction of LF imaging is presented. Next, major steps involving in the LF image communication system followed by the quality distortion model are reported. Finally, issues related to QoE assessment of LF image are presented.

7.1 Introduction

Space is filled with a dense array of light rays of various intensities. The set of rays passing through any point in space is termed *pencil* and Leonardo da Vinci refers to this set of rays as a *radiant pyramid* [11]. To represent all the information available to an observer at any point in space and time, a plenoptic function is defined. The plenoptic function, $P(\theta, \gamma, \lambda, t, V_x, V_y, V_z)$, measures the intensity of light seen from any viewpoint, the 3D spatial position with respect to camera center (V_x, V_y, V_z), any angular viewing direction (θ, γ), over time (t), and for each wavelength (λ). In another viewpoint, as an illustration, a stationary person can see:

- *a grayscale snapshot*, as shown in Figure 7.1 (a), is the intensity of light $P(\theta, \gamma)$ seen from a single viewpoint, at a single time, and averaged over the wavelengths of the visible spectrum. This can be also represented as $P(x, y)$, however for simplicity the spherical coordinates are used;

- *a color snapshot*, as shown in Figure 7.1 (b), is the intensity of light $P(\theta, \gamma, \lambda)$ seen from a single viewpoint, at a single time, and as a function of wavelength;

- *a movie*, as shown in Figure 7.1 (b) (a movie frame), is the intensity of light $P(\theta, \gamma, \lambda, t)$ seen from a single viewpoint, over the time, and as a function of wavelength;

- *a holographic movie*, as shown in Figure 7.1 (c) (a movie frame), is the intensity of light $P(\theta, \gamma, \lambda, t, V_x, V_y, V_z)$ seen from any viewpoint, over the time, and as a function of wavelength;

(a) Grayscale snapshot.　　　　(b) Color snapshot and a movie.　　　　(c) Holographic movie.
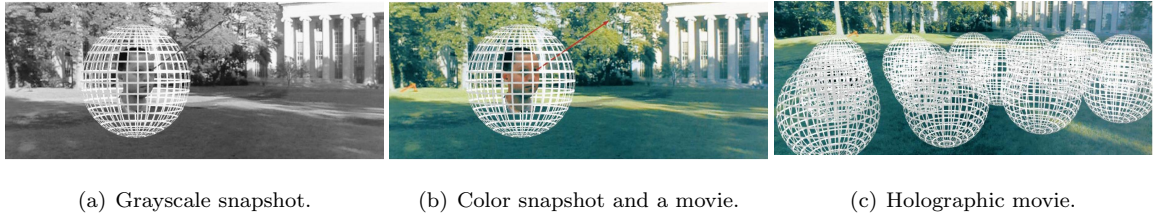
Figure 7.1: An example: modeling of the light as a plenoptic function (Figure by Leonard McMillan [198]).

- *plenoptic function*, as shown in Figure 7.1 (c), is the intensity of light $P(\theta, \gamma, \lambda, t, V_x, V_y, V_z)$. We can reconstruct every possible view, at every moment, from every position, and at every wavelength. Actually, it contains every photograph, every movie, and everything that anyone has ever seen.

Due to the complexity associated with the higher dimensionality of the plenoptic function, for the practical realization of the imaging system, some restrictions apply to the plenoptic function: radiance of a light ray remains constant along its path through empty space (one spatial dimension reduction), the time is fixed (static scene), and a specific wavelength is selected. As a result, the 7D plenoptic function is reduced to 4D LF. The LF is a plenoptic representation describing the amount of light faring through every point in space (V_x, V_y) in every direction (θ, γ). In brief,

- Surface plenoptic function (SPF): the radiance of a light ray remains constant along its path through empty space. Given that radiance along a light ray does not change unless blocked, SPF reduces the dimensions of the original plenoptic function to 6D including time (1D), wavelength (1D), a point on the surface S (2D) and azimuth and elevation angles (2D) [261, 53];

- 5D plenoptic function: McMillan and Bishop [154] present 5D representation of plenoptic as a set of panoramic images of different 3D locations. The reduction of the dimension is achieved for signal with a specific wavelength. The wavelength can be restricted to three channels (RGB) and each of the channels represent the range of wavelength captured by the camera sensor and by considering the static scene (fixed time), the time dimension can be eliminated. This assumption is still valid for video, by considering each image as a video frame [135, 75]. As a result of the assumption, the 7D plenoptic function becomes 5D function: $P(\theta, \gamma, V_x, V_y, V_z)$;

- LF or Lumigraph: in free space, the 5D plenoptic function becomes a 4D function, "the (scalar) LF" by considering SPF on 5D plenoptic function.

Levoy and Hanrahan, [135] defined the 4D LF as: the radiance as a function of position and direction, in regions of space free of occluders (free space). The proposed parameterizations technique is "light slab concept": by using the prospective geometry. From another viewpoint, it represents the ray by using the intersections of two planes in arbitrary position defined by $L(u, v, s, t)$, as shown
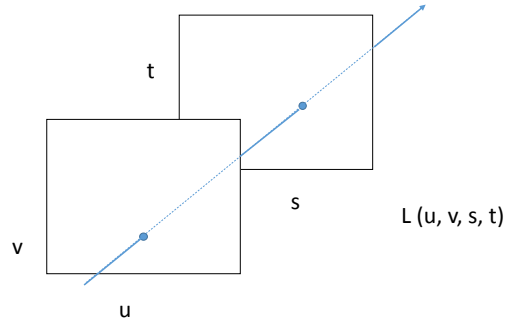
Figure 7.2: LF parametrization using "light slab concept".

in Figure 7.2. By convention, the coordinate system on the first plane is (u, v) and on the second plane is (s, t), (camera and image) and the oriented line is defined by connecting a point on the uv plane to a point on the st plane.

The latest advancement on the LF imaging is leaded by [169] and [150]: the LF camera with an additional microlens array inserted between the camera sensor and the main lens. The camera can be considered as a relay system, where the main lens creates a main image in the air, then this main image is re-mapped to the sensor by microlens array, and thus it is able to provide multiple views of the image in a single shot [70].

7.2   Light field image processing chain

The basic steps involving in the LF image communication process are shown in Figure 7.3. First radiated light is captured by using the appropriate acquisition device. The recorded data with associate metadata should be represented for further processing such as encoding and/or rendering. For the communication purpose, the compression is needed to reduce the big size size of LF data. Finally, at the receiver, the received compressed data should be decoded and rendering for display. These steps are briefly presented in the following subsections.

7.2.1   Acquisition

LF capturing is the sampling of 4D plenoptic function and it can be performed by exploiting different techniques such as camera arrays [254], moving camera [221], coded aperture [236], and using a microlens array in a camera. In the following, most popular techniques are briefly reported.
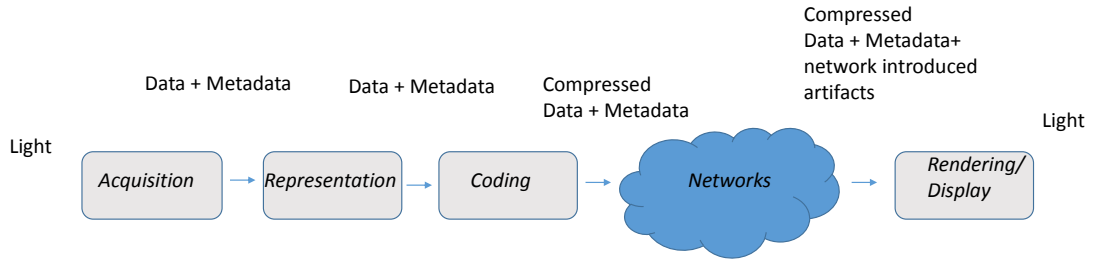
Figure 7.3: Basic block diagram of LF image communication.

(a)   (b)   (c)   (d)
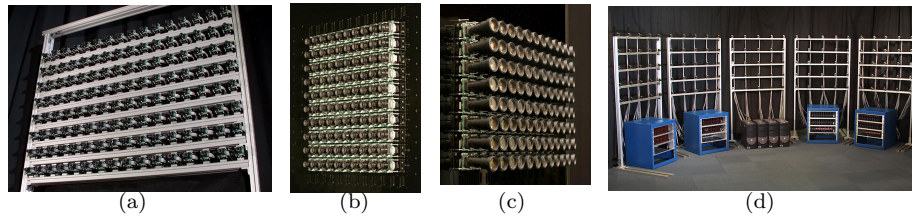
Figure 7.4: Stanford multi camera array with different configurations.

Camera array

Due to the low setup cost and computationally efficient mapping between the captured array of images and LF rays, the LF acquisition by using the multi camera array (camera grid) is the widely used and most popular technique. Moreover, the two-parallel-plane LF representation can be easily employed in the camera array acquisition system. A most popular LF acquisition system is The Stanford Multi-Camera Array [254], where 128 video camera's array is arranged and used in a variety of ways, shown in Figure 7.4. The camera array captures the visual content, @ 640 × 480 pixels × 30 fps × 128 cameras, with the synchronized timing, continuous streaming, and with flexible arrangements. Figure 7.4 (d) shows a widely spaced camera array with the aim of capturing LF. The tightly packed camera array is shown in Figure 7.4 (b) designed for high performance imaging. The intermediate spacing arrangement is shown in Figure 7.4 (a) designed for synthetic aperture photography.

Besides, the Stanford multi camera array, other many works has been performed for LF acquisition. A grid of 16 (4 × 4) cameras is used in [166] to capture 16 video sequences simultaneously. In [262] 48 cameras, grid size 8 × 6, are used to capture the images up to 640 × 480 pixel images at maximally 30 fps. An array of 64 cameras, 8 × 8 grid, is used for capturing the images at an average of 18 fps, with 320 by 240 videos coming from the cameras in [259]. Moreover, other techniques are also used to capture the LF. In [221] the LF is capturing by using the moving camera. The LF images with different grid size and resolution is recorded by Lego Granty in [233]. The LF

(a) Conventional camera   (b) LF camera
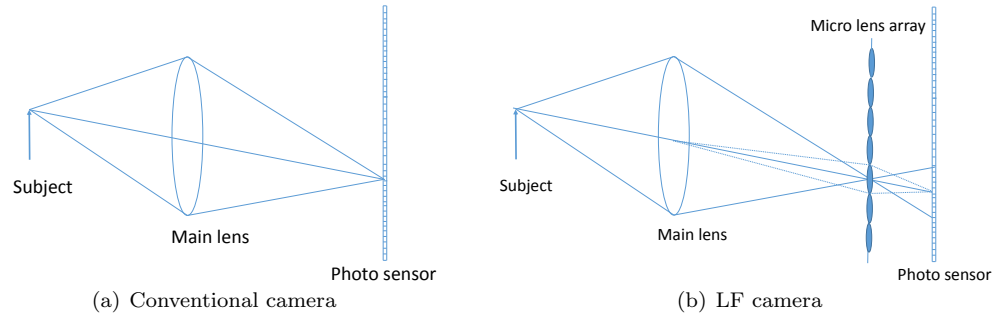
Figure 7.5: Conceptual schematic of LF camera with respect to conventional camera.

microscope is used in [136], where the microscope eyepiece has been removed and replaced with a microlens array; the array contains 288 ×192 microlenses.

Hand held light field camera

By using the microlens array, the LF capturing can be performed with a plenoptic camera (plenoptic 1.0 cameras) and focused plenoptic camera (plenoptic 2.0 cameras). The first commercial handheld plenoptic camera (Lytro) is made available by Ng [169]. Whereas, the focused plenoptic camera has also already made available in the market by Raytrix [2].

The difference between the conventional camera with LF camera is shown in Fihure 7.5. It shows that the conceptual schematic of the LF camera consist a main lens, microlens array and a photosensor. The main lens focuses the subject onto the micro lens array. The micro lens array separates the converging rays into an image on the photo sensor behind it [169]. Accordingly, the plenoptic camera captures the distribution of light rays as described by the 4D LF function by putting a microlens array in front of the image sensor. In particular, the focal plane of the microlens is on the camera image sensor plane, and the camera only captures angular information in each microlens image for a single point in the 3D space, and it results in a low spatial resolution of the final rendered images. To overcome this drawback (a trade off the spatial resolution with angular resolution), the focused plenoptic camera is proposed in [71]. As shown in Figure 7.6, focused plenoptic camera captures both angular and spatial information in each microlens image by putting the focal plane of microlenses far from the image sensor plane.

7.2.2   Preprocessing and representation

Selection of the representation model is based on the acquisition device, encoding formats and/or possible rendering method. To arrange the captured LF images in a standard 4D format (two plane representation) some preprocessing operations may be needed. For multi-camera acquisition system, the rectification operation may be needed for coordinating matching and timing synchronization.

(a) Standard plenoptic camera   (b) Focused plenoptic camera
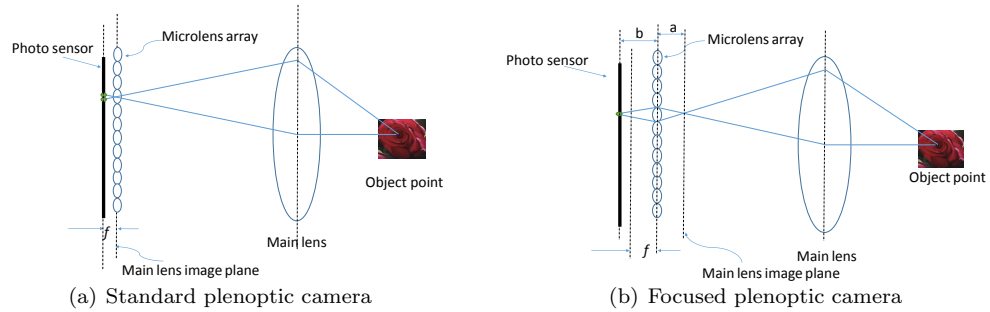
Figure 7.6: Standard and focused plenoptic camera. In focused plenoptic camera, position of microlens array satisfy the lens equation $1/a + 1/b = 1/f$. Where, a, b, and f are the distance from micro lens to main lens, distance from microlens to sensor, and focal length of microlens respectively.

The Granty recorded LF image, taken from the dataset [252], is shown in Figure 7.7.
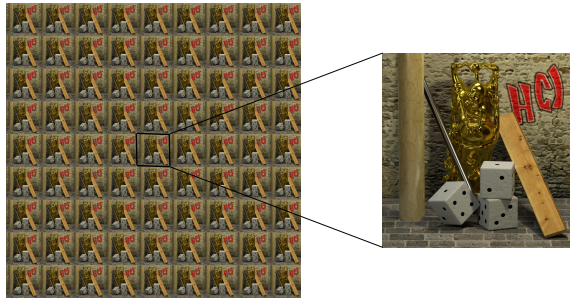
Figure 7.7: LF image (grid 8×8) captured by using the Granty.
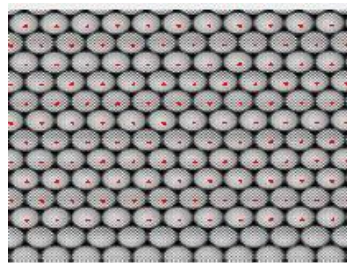
Figure 7.8: Examples of a white image showing estimated lenslet centers as red dots.

In this dissertation, the LF images taken by a hand held LF camera, called Lytro Illum, are also used for the study. For Lytro Illum recorded LF image, the decoding of the microlens recorded raw sensor data is needed. The decoding process creates 4D LFs by using the appropriate camera lenslet arrangement (Figure 7.8).

(a) Lytro Illum image

(b) Focused plenoptic image
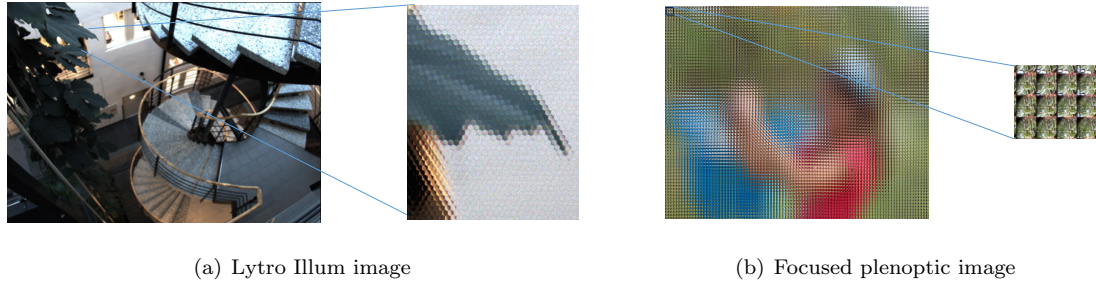
Figure 7.9: Lenslet LF images: (a) Lytro Illum camera recorded image and (b) focused plenoptic camera recorded image.

(a) LF data structure

(b) Sub-aperture views
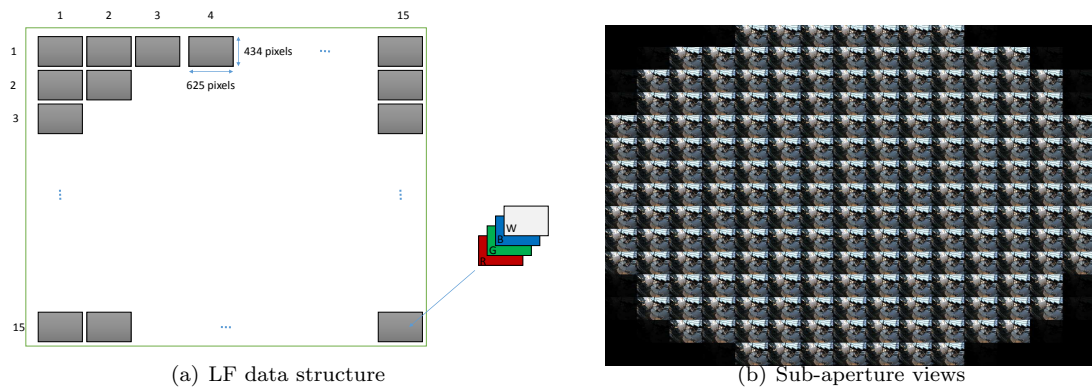
Figure 7.10: Data structure of LF Toolbox decoded Lytro Illum image.

To decode the Lytro Illum recorded raw sensor data LF Toolbox v0.4 [54] is used. During the decoding, followed by demosaicing and devignetting applied to the lenslet image (Figure 7.9 (a)), and color (gamma) correction, the lenslet image is converted into a 4D LF data structure (as shown in Figure 7.10 (a)). The LF data structure is a stack of 2D low-resolution RGB channel images in addition to a weighting image. The weighting image carries the confidence associated with each pixel, which can be useful in filtering applications that accept a weighting term [54]. The resulting dimension of the LF data structure is $15 \times 15 \times 434 \times 625 \times 4$, where 15×15 represents the number of views, 434×625 represents the resolution of each view and 4 corresponds to the RGB and weighting image components. For the processing, the weighting component is discarded, i.e. only RGB color channels are considered. In Figure 7.10 (b), views around the corers are black, this is due to the hexagonal geometry of the microlens.

7.2.3 Encoding

The LF image can be considered as a grid of elemental images, in which the content of each elementary image is similar to its neighbors. Due to the big size of LF camera recorded information (data and metadata), the compression of LF content is needed for storage and transmission.

In literature, some ongoing efforts have been given to encode the LF image. The encoding methods can be divided in two categories: lenslet image encoding and sub-aperture image encoding, based on the input (LF image representation) given to the encoder.

- *Lenslet image encoding*: In this technique, the lenslet image (shown in Figure 7.9) is directly given to the encoder as an input for encoding. In the state-of-the-art, most of the proposed LF image encoding methods are based on this technique: using HEVC the intra profile to encode the lenslet image. In [48] performance of the HEVC-intra is improved by using the self-similarity compensated prediction and estimation technique. Self-similarity compensated prediction is integrated in the HEVC intra profile to improve the performance of the encoding technique in [161]. In [139] intra prediction is improved by allowing the predictor to use only blocks from its reconstructed neighbors and using the advanced motion vector prediction.

- *Sub-aperture image encoding*: In this method, sub-aperture images (multiple views) are given to the encoder as an input for encoding. In [143], the sub-aperture images are used as the frames of the pseudo-video: coding order of the views accounts the similarities between adjacent views, and the coding is performed by using HM and JEM encoder [5, 6].

7.2.4 Rendering and display

As indicated before, many applications can benefit from LF imaging technology such as photography, astronomy, robotics, medical imaging, and microscopy. The most appealing applications of light imaging are: interactive rendering (where the focus, exposure, and depth of field can be adjusted after the picture is acquired), parallax, 3D visualization, multi-view applications, etc. Moreover, many ongoing efforts have been given to devise the applications of LF content. In particular, this technology is developed as a solution for Virtual Reality (VR) [95].

The selection of a LF rendering method is based on the targeted application of the LF image and, ultimately, on the display device. LF image can be rendered for 3D display by using 3D rendering methods: many ongoing efforts have been given for developing 3D rendering methods [116]. For example, a set of rendering techniques for 360 degree LF display is presented in [118] with the following features: autostereoscopic display (requires no special viewing glasses), omnidirectional (generates simultaneous views accommodating large numbers of viewers), and interactive (can update content at 200Hz). Currently, Holografika also announces full angle 3D LF display with perfect 3D experience with the feeling of reality [91].

The scope of the dissertation is not for advanced display devices. Whereas, our focus is for normal 2D display devices. Therefore, in this section, very basic rendering techniques are presented.

Digital Refocusing

LF imaging allows changing the focus of output photographs after the picture is taken for extracting a striking amount of detail that would have been irretrievably lost in a conventional photograph. In literature many refocusing algorithms are proposed for LF imaging. The basic digital refocusing algorithm is implemented by shifting and adding the sub-aperture images of the LF [168]. In brief, for every sub-aperture image $I(x, y)$:

- computes the (u, v) corresponding to that image;

- shifts the sub-aperture image by $\delta(x, y) = C * (u, v)$;

- average the shifted image into an output image;

where, larger C means refocusing further from the physical focus and the sign of C affects whether focusing closer or further. Moreover, for non-integral $\delta(x, y)$, the bilinear interpolation is used to blend into 4 nearest pixels of the output image. Moreover, the (x, y) coordinates are the microlens location in the image field of view (lenslet image) and (u, v) coordinate is the pixel location in the microlens image.

Digitally extended depth of field

The depth of field refers to the range of distance that appears acceptably sharp i.e. over a range every object in the scene is in-focus. An extended depth of field indicates that the extension of the depth of field without sacrificing resolution or brightness. The similar effect can be achieved by using the conventional camera, if we reduce the size of the lens aperture. A crucial advantage of the digitally extended depth of field photograph is: it uses the light coming from the larger lens aperture. From another viewpoint, it captures light more efficiently and allows the less grainy images with higher SNR [168].

Digitally stopping-down

A simplest way to compute the image with large depth of field is, by extracting a sub-aperture image from the LF. However, the problem with digital stopping down (as in its physical counterpart) is the waste of majority lights that passes through the full aperture: the result is grainier image with low SNR. An example of the digitally stopping down is all-in-focused view, which is resulted by summing only the central portion of each microlens.

Pseudo-video

The sub-aperture images are used as video frames, and it can be displayed as a pseudo video, with different viewing trajectory: horizontal, vertical, circular, etc. For an example, all possible views of the LF image content can be displayed by using circular animation at a frame rate of 60 fps with the help of *LFDispVidCirc* function of the LF Toolbox [50]. In this way, the user can visualize the spatial distribution as well as the angular information of the scene.

Rendering for focused plenoptic image

For generating a 2D view from focusing LF camera recorded content, basic rendering algorithm and weighted blending algorithm are presented by Georgiev and Aumsdaine [71].

- *Basic rendering algorithm:* For focused LF image (shown in Figure 7.9 (b)), each micro image can be considered as a low resolution view of the scene. The basic rendering technique selects the suitable portions of the micro image to stitch and then compose a 2D view image, shown in Figure 7.11. As shown, the input for the algorithm is LF image, an array $(N_x \times N_y)$ of micro images with the resolution of $MI_x \times MI_y$, and portion of a micro image (in the following referred as a patch) of the size $Ps \times Ps$ pixels are extracted from each micro image, and stitching them (patches) together: the result is a 2D view image with the resolution of $N_x \times Ps$ and $N_y \times Ps$ pixels. By varying the size and position of the patch in micro image different 2D images can be generated. The plane of the focus in the generated 2-D view image can be controlled by choosing a suitable patch size. As a result of varying the relative position of the patch in the micro image, the 2D views with different horizontal and vertical viewing angles are produced.

- *Rendering with weighted blending algorithm:* The basic rendering method may introduce some blocky artifacts due to a non perfect match between the patches with fixed size. To overcome this problem, rendering with weighted blending algorithm is proposed. The main idea of this algorithm is to smooth these artifacts with a weighted blending method: averaging together all these overlapping regions across different micro images by weighting the pixels differently in the summation process. As a result of this algorithm, objects in the scene which are outside the plane of focus has a more natural blurred look [46].

7.3 Light Field image quality issues

LF imaging demands high computational power and presents image resolution and quality issues. The LF image is also subject to several distortions during acquisition, processing, decoding, compression, storage, transmission and reproduction phases, and each of these stage results in a degradation
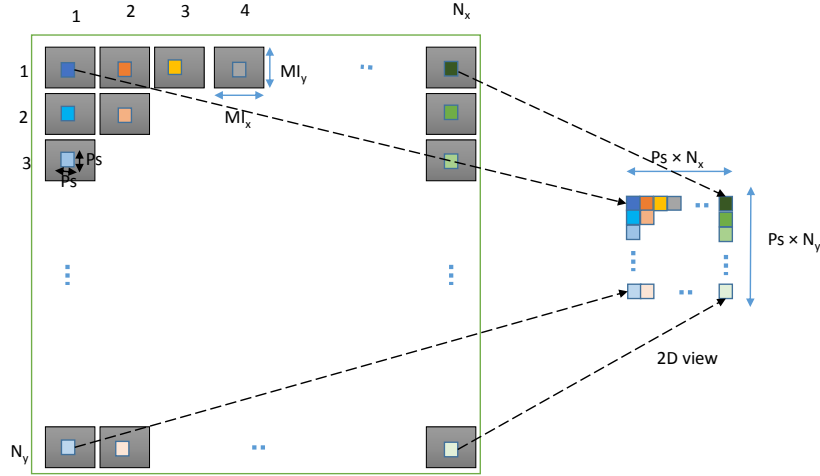
Figure 7.11: Basic rendering method.

of the visual quality. In addition to conventional 2D/3D processing steps, the LF imaging follows the computationally complex additional steps; these steps also introduce the artifacts/distortions, and ultimately, degrade the quality of LF image. In particular, in preprocessing (raw data decoding and representation) step lots of information (such as color) get lost and/or distorted. Due to the big size of LF camera recorded information, aggressive compression of LF content is needed. Next, in most of the applications, the end user of the LF data is human, and LF content is shown to the final user after post-processing, such as rendering. The rendering methods also introduce artifacts [70]. Meanwhile, the knowledge of degraded quality or quality level is important in the design and optimization of the LF imaging system.

7.3.1 Generic light field image quality distortion model

A generic quality distortion model could be:

$$AA_x + R_x + CA_x + TA + RA_x + DD_x, \tag{7.1}$$

where,

- AA_x is the acquisition artifacts, produced during the acquisition process. Moreover, this is also depending on the demosaicing, devigneting, and color/gama correction techniques applied to the raw sensor data;

- R_x is the artifact introduced during the representation phase. In LF imaging, the artifact is introduced during the conversion of camera sensor data to the 4D LF. Moreover, the selection of the representation is also depending on the probable encoding techniques;

- CA_x is resulted from the encoding and decoding procedure. The coding artifacts are depending on the content to be encoded and adopted encoding methods;

- TA is the artifact introduced by the transmission channels: wired or wireless. The transmission impairments result a significant degradation in the LF image quality;

- RA_x is the distortion resulted by the rendering methods. The selection of rendering method is depending on the probable visualization technique and device, and the produced artifacts are also depend on the rendering methods;

- DD_x is the artifact introduced by the display devices. Different display device has different parameters such as contrast and color, and the QoE is also influenced by the target display device and its parameters.

7.3.2 Quality assessment issues

Usually, the performance of an IQM is evaluated by comparing the estimated score with the corresponding subjective score (MOS) of test images. The basic steps of LF image processing are presented in Figure 7.12, and it shows that the quality of LF image can be evaluated at two points: *P1'-after post-processing* or/and *P2'-before post-processing*. Due to the complex nature of LF content and visualization possibilities, there are many issues related to perceptual quality evaluation of LF image.

- *Subjective quality evaluation:* In general, before the LF content shown to the user, post-processing operation (such as rendering) is performed based on target display device and/or visualization technique. The quality of the LF content could be evaluated after post-processing, and thus the HVS characteristics are more relevant at the point *P1'* than *P2'*.

- *Objective quality evaluation:* On the other hand, the result of post-processing operation (such as refocused view) may not cover all the information available in the LF content. In particular, a refocused view can be considered as a single view among the many views available in LF image, and the refocused view may not cover all information about LF content. For the quality evaluation, it is important to consider the complete information, if possible. Thus, for objective quality assessment, quality evaluation at the point *P2'* is more relevant than *P1'*. Accordingly, in literature [48, 237, 139] to compare the performance of compression methods, PSNR is computed at points *P2'*.

Furthermore, the issues related to subjective and objective quality assessment of LF image:

1. Can we compare the objective quality score computed at *P2'* with the subjective score estimated at the point *P1'*?
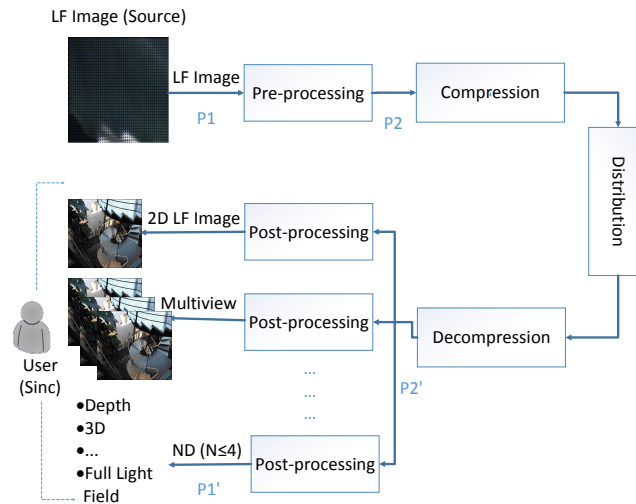
Figure 7.12: LF image processing chain: quality evaluation prospective.

2. What is the better way for visualizing?

3. What would be the subjective quality assessment method?

In this dissertation, our assumption is the accumulated distortion in LF content will be evaluated by objective metrics at point P2', and the distortion of the LF content will be replicated in rendered view, and it is ultimately evaluated by subjects at point P1'.

For subjective quality evaluation, the selection of a visualization technique is necessary. As mentioned before, there are different ways for the visualization and each of them have own pros and cons. For example:

- refocused view/s– many refocused views are possible by focusing on a different plane/depths, and for time consuming and expensive subjective quality evaluation considering all the possible views is not practical; quality of refocused view is also depending on the refocusing/rendering methods, and artifacts depends on the selected rendering methods; and in general, in the refocused views in-focus region is clear and out-of-focus region is blurry, and considering a single refocused view may not be sufficient to evaluate the quality of entire LF image;

- all-in-focused view– validated 2D image quality protocols can be adopted. However, all-in-focused view can be considered as only one view of the LF image; quality measured on a single view may not replicate the quality of complete LF image; and this technique may not be a good option to evaluate the distortion on depth information;

- pseudo-video– there will not be a clear temporal/motion information in views as in natural video, and there is no validated subjective quality assessment protocol for this type of visualization.

To the based on these issues, different visualization techniques and subjective assessment methods are used to study the LF image QoE in the next chapters.

Chapter 8

Light field image watermarking

Securing multimedia data from undesired manipulation is a widely investigated topic in literature. A large numbers of techniques have been developed for protecting images, videos, and also audio from malicious attacks. Nowadays, new imaging systems, LF imaging, pose new challenges from the data protection point of view. Thus, in this chapter, a novel embedding scheme tuned to LF image is presented. The major contributions of this chapter are:

- a watermarking system that does not affect the depth map estimation procedure and therefore the refocusing procedure and the 3D scene reconstruction are presented;

- the impact of LF data modification with respect to the perceived quality has been investigated;

- experimental tests, both objective and subjective, have been carried out on the LF images (captured by using the Granty device) for assessing the performances of the proposed algorithm;

- finally, robustness of proposed watermarking scheme are also evaluated to the focused LF images for different compression methods.

8.1   Introduction

The increasing interest towards LF media calls for methods for protecting these data from manipulations and unauthorized reproduction or diffusion. To this aim watermarking techniques have been designed for copyright protection: before data (cover data) distribution an invisible signature (the watermark) is hidden in the host source by using a secret key. This key is used later for verifying the presence of the signature during the detection phase.

The design of an effective watermarking scheme consists in finding a trade-off between three factors:

1. *quality*: the watermarked data should be as close as possible to the original one;

2. *robustness*: it should be possible to detect or extract the watermark even if wanted or un-wanted modifications, noise, compression, errors, or attacks, affect the watermarked data;

3. *capacity*: the amount of bit embedded in the data that is depending on the particular application.

As widely demonstrated from the many efforts performed in designing image or video water-marking schemes, such constraints are competing. The use of LF image as cover data poses multiple choices starting from the selection of the valuable data to be protected. In fact, the cover data to be watermarked can be

- the rendered image (focusing, rendering model,...)

- the depth information;

- the raw data (the acquired LF).

When dealing with the LF watermarking, the number of views results in a variety of possibilities for embedding: exploiting all existing views or only a subset of them, exploiting the intrinsic redundancy give by the acquisition setup, using a distributed system, etc.

Evaluating the *quality* issue, that is the imperceptibility of the watermark or its impact on the cover data, is a challenging task. According to the particular scenario, the perceived quality of the rendered 3D scene of the 2D single views, the BER between original and watermarked data, or the impact in the reconstructed depth has to be considered. Even the *Robustness* evaluation depends on applications: the watermarked could be desirable to be detected in the rendered scene, in the generic view or in the extracted depth. Each one of these possibilities pose different challenges in designing a robust scheme and in designing reliable robustness evaluation schemes.

As can be noticed from these comments, coupling watermarking techniques to LF data it is not an easy task. Also for this reason, in literature very few works are dealing with the watermarking of LFs: most of them are considering free view point data or investigating a watermarking system trying to adapt a scheme designed for 2D images to this scenario.

In [127] and more recently in [128] a free-view TV scenario is considered. In these systems, the viewers select freely the viewing position and angle. The watermark is embedded into every frame of multiple views by exploiting the spatial masking properties of the HVS. An interesting analysis is performed on the distortions caused by different interpolation algorithms. In [16] a watermarking scheme in the spatial domain is proposed. The invisibility of the watermark is achieved by modu-lating the strength of the embedding according to the presence of high frequencies. The robustness of the system is considered, and a watermarking selection is made by verifying the possibility of detecting them in the rendered scene. In [224], a spread spectrum multibit watermarking scheme

for free-view video is presented. The watermark is embedded into every frame of multiple views, while the watermark detection is carried out exploiting the DCT of virtual frame generated for an arbitrary view.

There is, however the need for understanding the impact of LF modification due to embedding both with respect to the perceived quality of rendered scene as well as to the quality of the reconstructed depth, the need of designing reliable robustness evaluation tools, the realization of the data set of original data with related ground truth, and, finally, a better classification of the requirements of each scenario vs. a watermarking system. This works wants to contribute towards the definition of the general requirements of the watermarking scheme and to investigate the impact of embedding data in the LF in the perceived quality of the rendered scene. In more details, in this work the designed watermarking scheme is tailored to embedding the watermarking in the raw data and to verify its visibility from the views generated by using the LF rendering.

8.2   Proposed watermarking scheme

In the proposed watermarking scheme, all views have been used for embedding the watermark, thus increasing the embedding capacity as well as the potential impact on the watermark visibility.

Let us consider all views placed on a rectangular lattice of size $N \times M$, and let us denote each view as $V(i, j)$ where $i = 1, ..., N$ and $j = 1, ..., M$. Each view is a matrix of size $h \times k$ pixels. The size of the watermark is w bit. The embedding algorithm, shown in Figure 8.1, can be summarized in the following steps:

- All the views are merged in order to obtain a matrix I. This is performed by rearranging in lexicographical order the generic view $V(i, j)$ to obtain a matrix of size $h \times (N \times M \times k)$ pixels;

- The first level of the Haar wavelet transform of I is computed;

- The subbands containing the horizontal (HL) and vertical (LH) details are selected for hosting the watermark W;

- The watermark W is embedded, in each color component, by inserting its odd rows in the HL subband and its even rows in the LH subband by using the additive scheme:

$$HL_W = HL + alpha * W_{odd}; \tag{8.1}$$

$$LH_W = LH + alpha * W_{even}; \tag{8.2}$$

where $alpha$ is the watermark strength.

- The watermarked LF array is reconstructed by performing the inverse first level Haar Transform and rearranging the single views in the original rectangular shape.
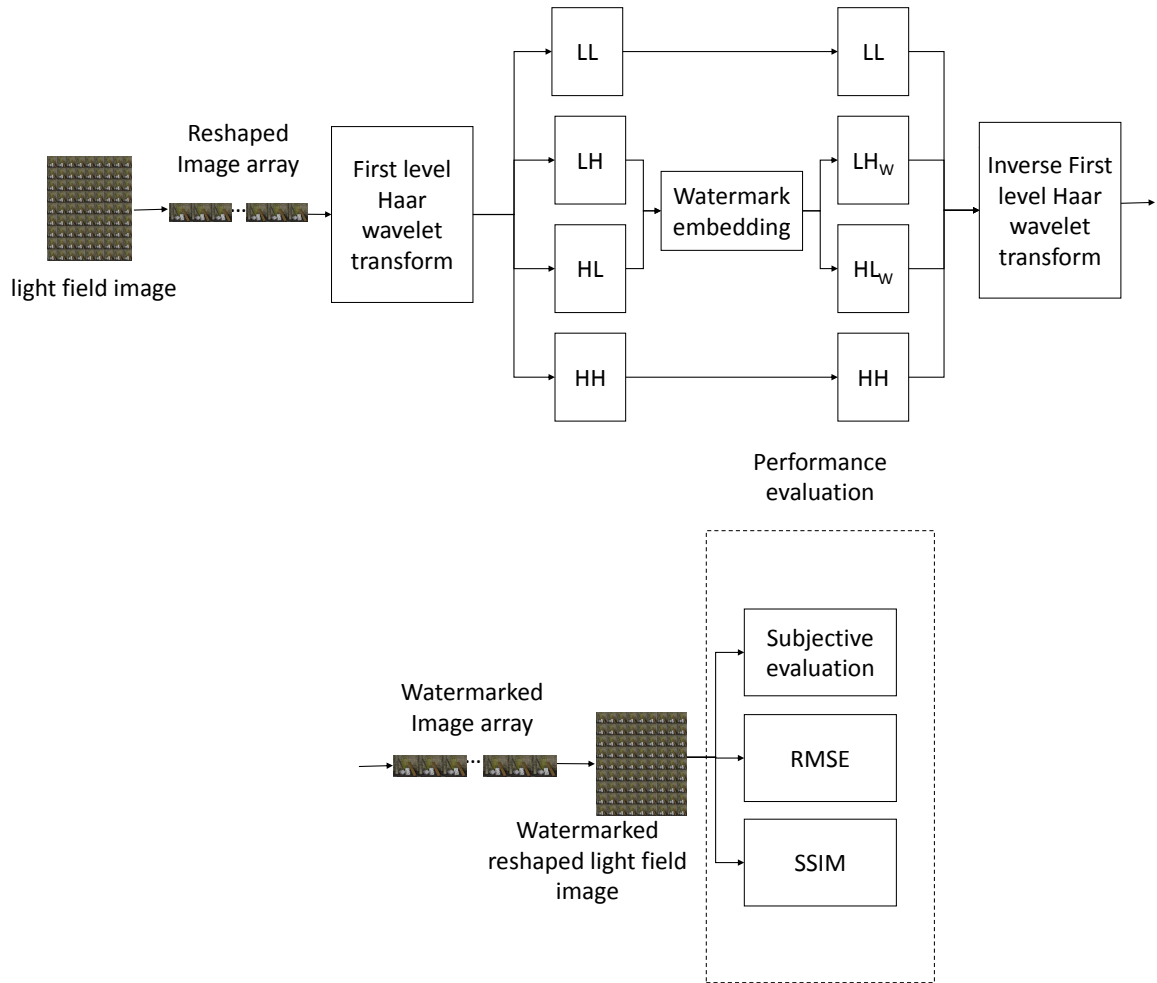
Figure 8.1: Block diagram of the proposed watermarking scheme.

Since this is a non-blind watermarking scheme, the watermark can be extracted by performing the reverse process with respect to the one used for the embedding.

8.3   Experimental validation on gantry light field images

In order to verify the effectiveness of the proposed method, since at the best of our knowledge it does not exist a method for assessing the quality of LFs, we performed both a subjective and an objective evaluation as detailed in the following subsections. All tests have been carried out on the LF database that is made available by [252]. Each LF is composed of 81 images placed on a grid of size 9×9.
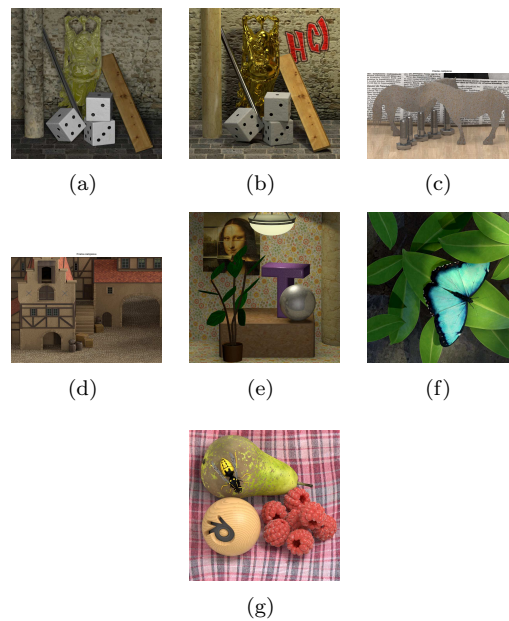
Figure 8.2: Central views of the "Blender" images in the database [252]: (a) Buddha, (b) Buddha2, (c) Horses, (d) Medieval, (e) Mona, (f) Papillon, (g) Stilllife.

This database contains thirteen high quality densely sampled LFs. Among these, 7 are computer graphic generated with the program Blender and for for all the considered views, the corresponding ground truth maps are available, the remaining 6 datasets are real scenes that have been recorded by using a camera placed on a gantry. For the last category of images the ground truth is available only for the central view. In the performed experiments we considered only the computer generated datasets and the details about the database are reported in Table 8.1 while their central views are shown in Figure 8.2. The selected watermark is a gray level image of size 118×195.

Dataset name	Resolution (pxl)
Buddha	768 * 768
Buddha2	768 * 768
Horses	576 * 1024
Medieval	720 * 1024
Mona	768 * 768
Papillon	768 * 768
Stilllife	768 * 768

Table 8.1: Dataset parameters.

(a)   (b) PSNR= 40.42 dB   (c) PSNR= 31.71 dB

(d) PSNR= 27.93 dB   (e) PSNR= 25.59 dB   (f) PSNR= 23.75 dB
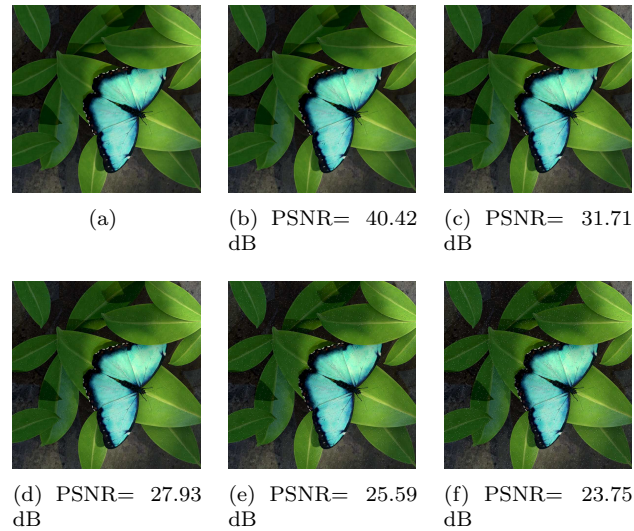
Figure 8.3: Central view of the dataset *Papillon* (a) and it watermarked versions with (b) $alpha = 0.1$, (c) $alpha = 0.3$, (d) $alpha = 0.5$, (e) $alpha = 0.7$, (f) $alpha = 1$.

8.3.1 Subjective tests

In order to verify the visual impact of the insertion of the watermark we performed a session of subjective tests. A total number of 20 subjects (5 females and 15 males) participated to the test. The age of the participants is between 22 and 52. The LF images were rendered through a 4D LF circular animation at a frame rate of 60 fps [50]. Each dataset was watermarked with 5 different values of *alpha*: 0.1, 0.3, 0.5, 0.7, and 1. The values of *alpha* have been empirically selected based on their impact on the watermark visibility. The central view of the image *Papillon* watermarked with the five values of *alpha* is shown in Figure 8.3.

The subjective tests have been carried out according to [109]; using a ACR method. After a training phase in which the subjects were made confident with the content and the evaluation scheme, the content was randomly displayed and each subject has been asked to rate the quality of the rendered LFs on a scale from 1 to 5 with 1 corresponding to worst quality and 5 to best quality.

After the collection of the opinion scores, an outlier detection procedure is applied to detect the subjects whose score strongly deviates from the mean behavior, and that show a significant bias compared to the average behavior. In this study, the scores given by only one subject were discarded and the scores given by 19 subjects were retained. After the performance of the outlier detection procedure, the perceived video quality was measured in terms of MOS [108].

The results are shown in Figure 8.4 from which, as expected, it is possible to notice, for all considered LFs, a decrease in the MOS score with the increase of the value of *alpha*. Moreover, for values of *alpha* larger than 0.3 the watermark becomes more visible and consequently the MOS score drops.
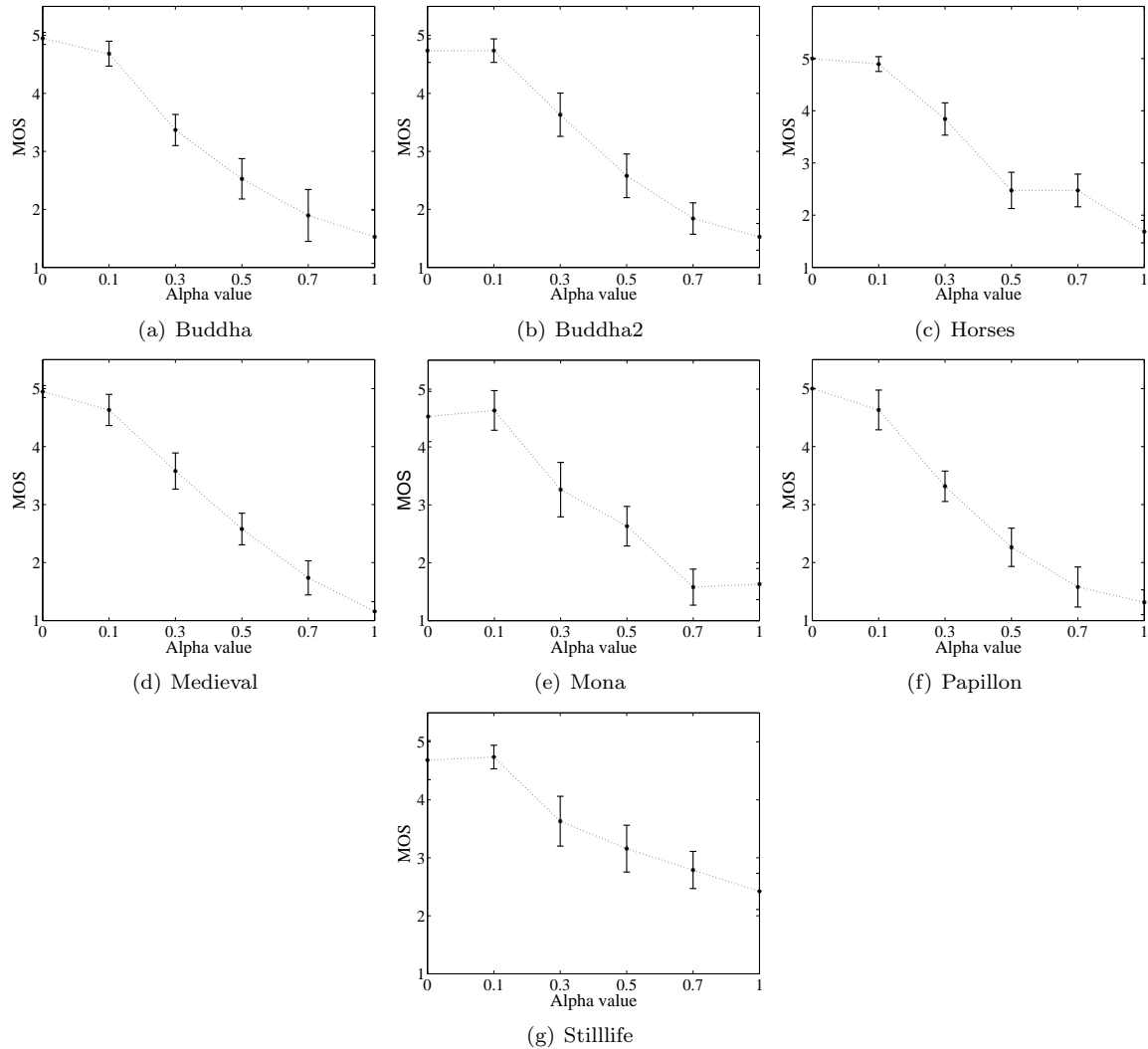
Figure 8.4: MOS for the 7 LFs under test for the different values of chosen alpha.

Moreover, the collected results have been analyzed with the ANOVA test. First, an *inter LF image* analysis has been performed to understand the impact of the different content on perceptual visual quality when watermarked with the same level of *alpha*. The obtained result (P_{value}=0.9924) suggests that the perceived quality does not significantly change with the image scene. Then, we performed an *intra LF image* analysis to understand the impact of the variation of *alpha* on the perceived quality. The result (P_{value}=0) shows, as already seen by the MOS analysis, that the perceived quality is significantly affected by the increase of the strength of the watermark.
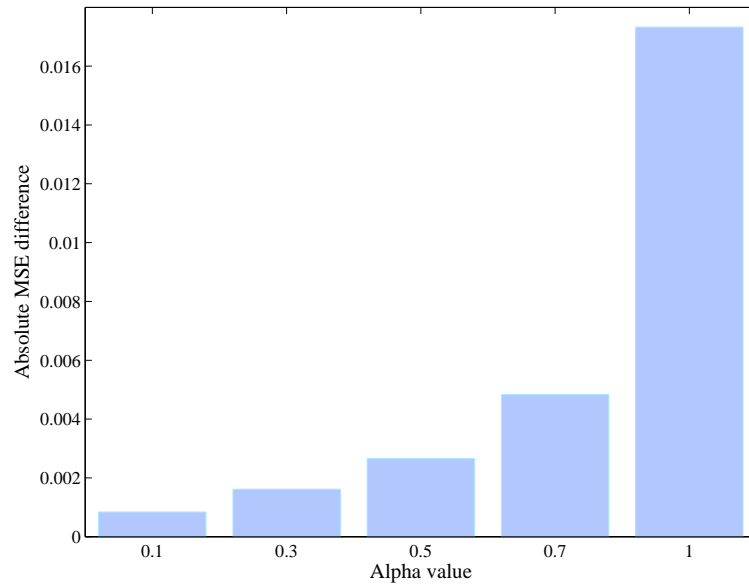
Figure 8.5: Analysis of depth map degradation when the watermark is inserted.

8.3.2   Objective evaluation

As mentioned in the introduction, one crucial aspect in the LF imaging is the refocusing step. In order to perform a correct refocus, it is important that the depth map is not affected by errors that lead to a wrong refocus since this will result in non natural images. To this aim, after performing the watermark insertion, we estimated the depth map corresponding to the central view and we computed the MSE among this map and the corresponding ground truth. Similarly, we estimated the depth map corresponding to the central view when no watermark is added and also in this case we computed the MSE with the ground truth. The depth map estimation was performed by using the algorithm proposed in [167] in which a Maximum Likelihood estimation is used. The achieved results are reported in Figure 8.5. In this figure we report, for each value of alpha considered, the mean of the absolute difference among the MSE computed between the estimated depth map when no watermark is added and when it in embedded. It can be noticed that the variation in MSE is very small, in fact the value of the difference is lower than 0.018. This means that the watermark insertion does not affect the quality of the depth map and thus it does not create alterations in the refocusing procedure.

The performed tests show that for *alpha* values larger than 0.3 the watermark starts being visible, thus reducing the subjective perceived quality. Anyways, the watermark insertion, even for high levels of *alpha* , does not create artifacts that affect the depth map, thus preserving the refocusing capability.

(a) DemichelisCutFrame1   (b) planeToyFrame23   (c) planeToyFrame123   (d) Laura
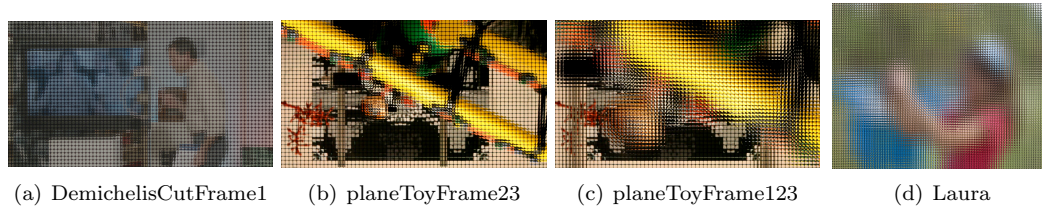
Figure 8.6: Focused LF images.

8.4   Robustness check for focused light field images

8.4.1   Focused light field images

For the study, a set of four LF images (shown in Figure 8.6) were considered; the basic features of the images are summarized in Table 8.2.

S. No.	Image Name	Image Size (Pixel)	EI size	Total EI
1	DemichelisCutFrame1	2850×1558	38×38	3075
2	planeToyFrame23	1904×1064	28×28	2584
3	planeToyFrame123	1904×106	28×28	2584
4	Laura	7104×5328	74×74	6912

Table 8.2: Focused light field images basic features: name, image size, Elementary Images (EI) size, and total number of elementary images, of the LF images.

8.4.2   Considered compression methods

The robustness of the proposed watermarking scheme was evaluated for the compression methods: JPEG, JPEG2000, HEVC intra, and LF image compression method HEVC SS [47].

8.4.3   Adopted performance evaluation technique

To evaluate the performance of the proposed watermarking scheme, the watermark strength (α) is set to 0.3. Because, consideration of all the possible values of the strength is not possible: time needed for the experiment (particularly for HEVC SS encoding) was very long. The selection of α value depends on the results presented in Section 8.3.1: the result shows that, the α is equal to 0.3 can be considered as a JND threshold value from the point of human visual perceptual quality.

The grid of Elementary Images (EI) array is treated as a single image and the red color channel of the image is used as a cover data. The watermark image of the size of $32 \times 32 \times 3$ bits is embedded in the cover data (as described in the Figure 8.7). All EIs are used for embedding the watermark, thus increasing the embedding capacity as well as the potential impact on the watermark visibility.

To evaluate the robustness of the proposed watermarking scheme, the following steps are considered: i) firstly, the watermark is embedded in the plenoptic images, ii) the embedded image is
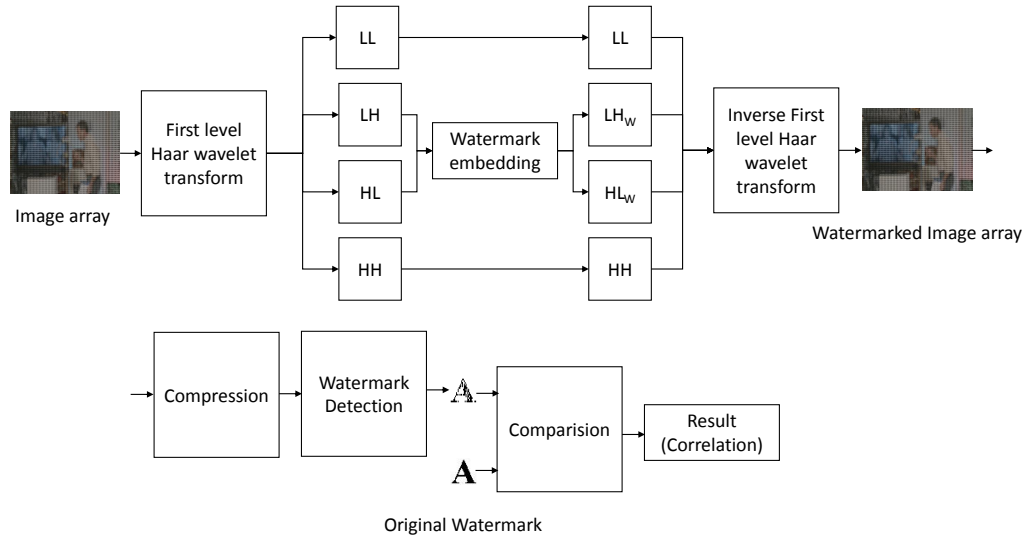
Figure 8.7: An example: watermarking on LF image.

compressed for different values of compression ratio by using the compression methods, and iii) the watermark is extracted from the compressed images and the correlation between the extracted and the original watermark is computed.

8.4.4 Results and discussions

In the following subsections, impact of watermarking on the quality of the LF image, the performance of the considered compression techniques, and the result of the robustness evaluation are presented.

Impact of the watermark on LF image quality

One important factor to consider in the design of a watermarking scheme, is the quality of the watermarked data while considering the amount of embedded data (capacity) and the robustness. It means the watermarked image should be as close as possible to the original one. However, measuring the quality of the LF images is still an open issue. As usual for classical image watermarking, it is expected that high watermark strength (α) results in poor image quality. Also in the performed experiment this trend has been confirmed. Figure 8.8 shows that the quality of the LF images, measured in terms of PSNR , decreases for high values of watermark strength. The PSNR is computed between the original and the watermarked rendered image. In more details, the LF image is rendered by using a basic rendering and weighted blending rending method (as discussed in Section 7.2.4) with the patch size of four and the patch was at the center of the EI. Moreover, the
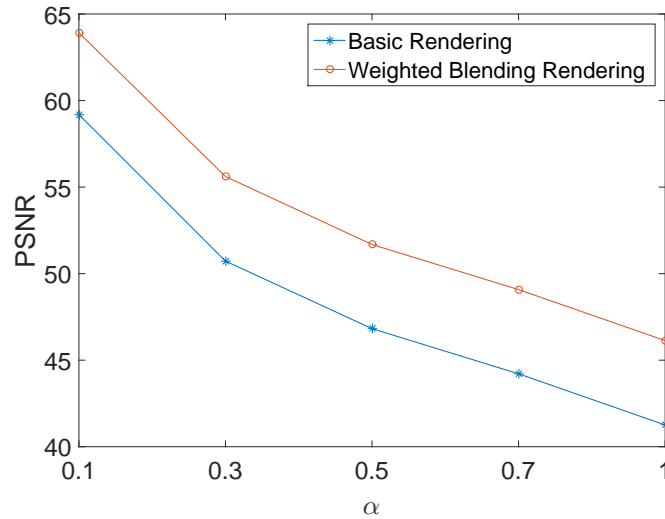
Figure 8.8: PSNR of the rendered view of the image *planeToyFrame23* for different level of watermark strengths (α). Moreover, the perceptual quality (measured in terms PNSR) of weighted blending rendering method is always higher than for basic rendering to all the values of α.

Figure 8.8 shows that the quality of the rendered image by weighted blending rendering is always higher than the basic rendering method for all the set of α values.

Performance comparison of the encoding methods

During the robustness check, the watermarked LF images were compressed by using different compression methods. The achieved results, plotted in Figure 8.9, show that the recently proposed compression techniques HEVC SS results in higher values of PSNR for all the considered compression ratios. This result indicates that the HEVC SS has a better compression capability for the given quality over JPEG, JPEG2000, and HEVC intra.

Robustness check for encoding

As mentioned before, to evaluate the robustness of the proposed watermarking scheme against the compression methods, the watermark strength (α) is set to 0.3. Figure 8.10 shows that the relationship between, the correlation coefficient between the extracted watermark and the original watermark at different compression levels for all the considered compression techniques. Figure 8.10 (b), (c) and (d) demonstrate that, for the images *planeToyFrame23*, *planeToyFrame123*, and *laura* the correlation decreases sharply for the small increment on the compression factor for all the considered compression techniques. Figure 8.10 (a) shows that the correlation coefficient decreases rapidly for high compression rates. JPEG2000 shows the slightly higher correlation for high values of

(a) DemichelisCutFrame1

(b) planeToyFrame23

(c) planeToyFrame123

(d) laura
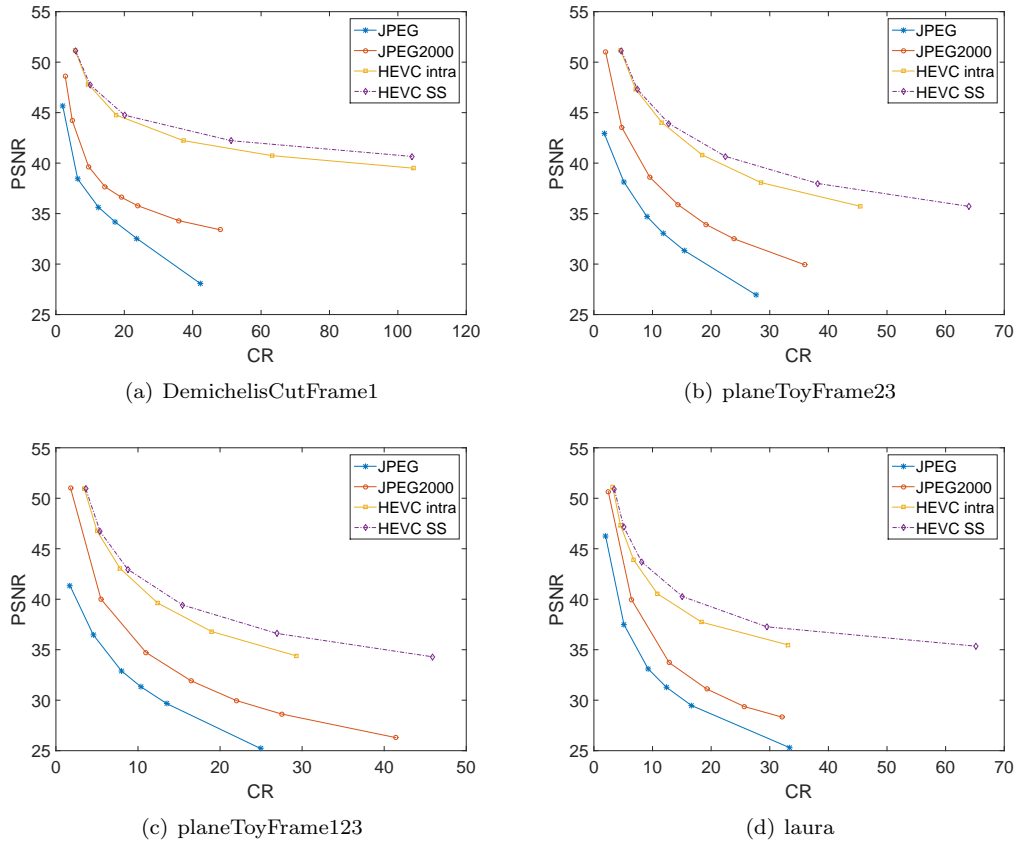
Figure 8.9: Performance of the considered compression techniques for different images. The quality of the decoded image is measured in terms of PSNR and the compression level is expressed in terms of Compression Ratio (CR).

compression factor than other compression techniques. Moreover, for all the images the correlation is slightly higher for JPEG2000 compared to others compression methods.

From these results, we can summarize that the proposed watermarking scheme is not robust against the high values of compression rate for all the considered compression methods when applied to LF images. The results indicate that the proposed watermarking scheme is more robust against JPEG2000 compared to other considered compression techniques. However, as presented in the above (Figure 8.9) the HEVC SS has the higher PSNR compared to other compression methods for a given compression ratio. This is because the proposed watermarking schemes depend on the DWT, as JPEG2000 compression method.

Figure 8.11 shows that the relationship between, the correlation coefficient (computed between the extracted and original watermark) and quality (expressed in terms of PSNR) of the corresponding compressed image from which the watermark is extracted. From the Figure 8.11, it can be noticed
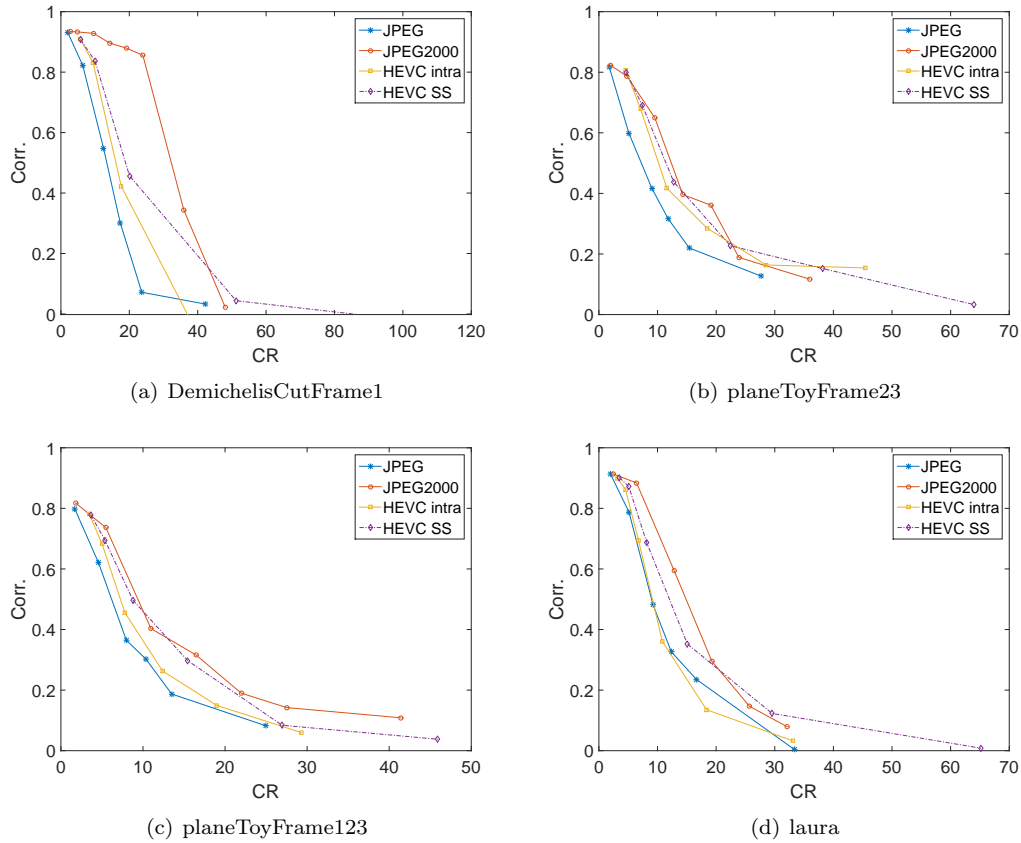
Figure 8.10: Watermark robustness against compression ratio.

that the PSNR is always high for HEVC intra and HEVC SS compared to JPEG and JPEG2000, and the result is expected. Moreover, it shows that the correlation always increases for high values of PSNR for all the considered images and compression methods. It has also been noticed that the correlation is high for JPEG and JPEG 200 compared to HEVC intra and HEVC SS for a given value of the PSNR. This result indicates that the proposed watermarking scheme is not robust against the newer compression techniques.

8.5 Conclusions

In this chapter a watermarking system for LF image is presented. The designed method exploits the redundancy of the data to be watermarked for increasing the strength of the watermark. In particular, the goal of this work, was to study the perceived impact of modification in the LF array, and its impact on the 3D scene reconstruction. The subjective results define a visibility threshold (α = 0.3) for the embedded data, while the objective test demonstrates that the reconstructed depth

(a) DemichelisCutFrame1

(b) planeToyFrame23

(c) planeToyFrame123

(d) laura
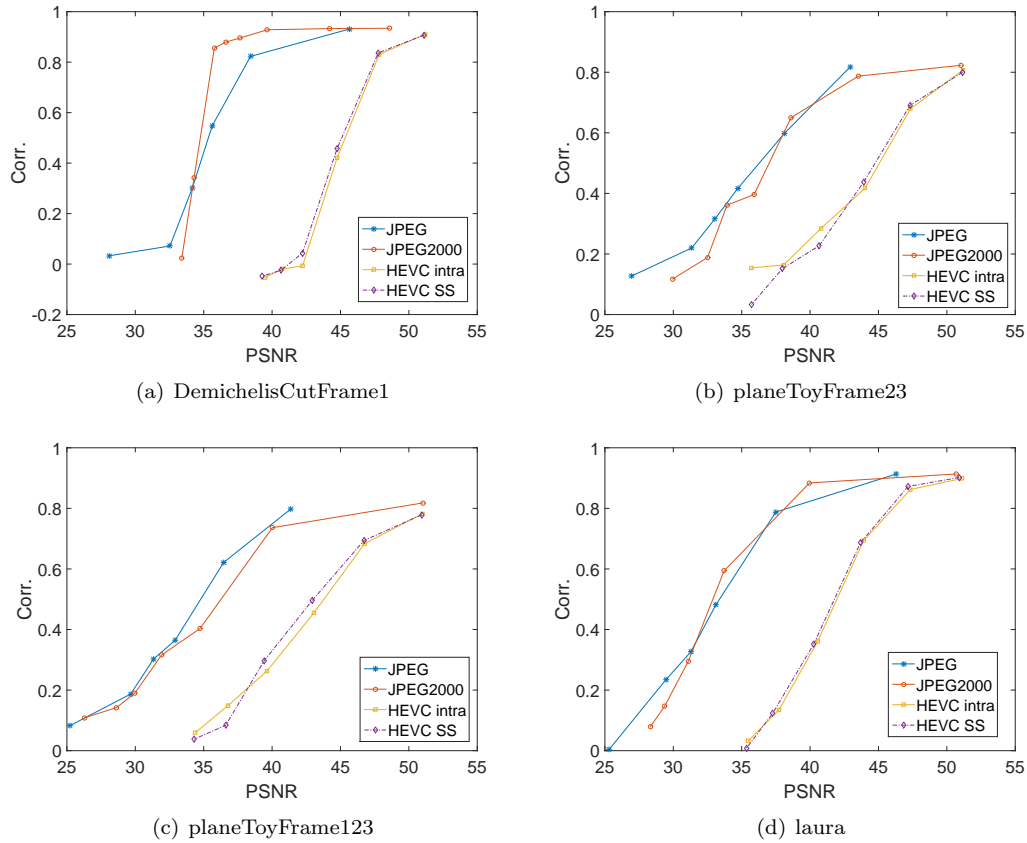
Figure 8.11: The relationship between the LF image quality (PSNR) and watermark robustness (measured as Corr.–correlation coefficient).

map is almost not affected by the watermarking system thus allowing a perfect rendering of the scene.

Next, the robustness of proposed watermarking method is evaluated on the focused LF images for different compression techniques. The achieved results show that the watermarking scheme is robust only for very low level of compression ratios for JPEG, JPEG2000, HEVC intra and HEVC SS. The correlation between the original and extracted watermark increases for the compressed embedded images with high values of PNSR. Moreover, the performance of the watermarking scheme is more robust against JPEG2000 compared to other three methods. However, the HEVC SS has higher PSNR compared to other compression methods for a given compression ratio. This result is due to the fact that watermarking scheme is also exploiting DWT features as for JPEG2000.

Chapter 9

Quality evaluation of light field images

The major contributions of this chapter are as follows:

- a Reduced Reference (RR) LF IQA framework is presented. In brief, original and distorted depth maps are estimated from original and distorted LF images. To evaluate the proposed framework, three depth map estimation methods were used. Distortion of the depth map is used for estimating the QoE of LF image;

- a LF image dataset is provided. It is composed of test images and annotated subjective quality ratings. To the best of our knowledge, in literature, no data set provides the information about test LF images and subjective quality ratings. A subjective experiment has been performed to collect the opinion scores for test LF images;

- an analysis of the performances of standard image compression methods when applied to LF images has been presented;

- an analysis of the performances of widely used 2D IQMs, when applied to LF images, has been performed by using the dataset.

9.1 Introduction

As mentioned before, in Section 7.3.1, the LF image also suffers form many distortions and thus, the QoE assessment of LF image is crucial. Based on availability of reference information, quality metrics can be classified in three different categories: FR, RR, and NR. The availability of FR information allows better prediction of quality; several well established metrics such as PSNR and

SSIM, are commonly used. On the other hand, FR metrics are rarely applicable in real-world image communication environments, and its applicability is further limited in LF image, due to the size of the reference LF image (data and metadata). Therefore, in this dissertation a RR quality assessment framework is proposed for LF image.

9.2 Related works

To the best of our knowledge, no work is published in this particular direction, i.e. there is no objective IQM designed for LF image. However, similar works are already performed in 2D and 3D image, and few of them are briefly reported in the following. The basic ideas used in these works can be used as a background information for designing a new quality metric for LF image.

9.2.1 Image quality metrics

In evaluating image processing algorithms, many FR metrics such as SSIM [250], PSNR, and PSNR-HVS [57], are widely used. Performances of the quality metrics [19, 247, 142] have been evaluated for a wide range of images from different image quality datasets [185, 211]. The results presented in [247, 36] show many challenges and issues regarding the applicability and performances of the metrics even for 2D images. In LF image processing, in particular, for evaluating the performance of LF image encoding methods, the PSNR metric has been used [48, 237, 139]. The usability of widely used metrics for LF images is still to be assessed.

9.2.2 Depth map quality and overall visual quality of experience

In literature, no work is performed to understand the relationship between depth map quality and overall QoE for LF image.

In the last two decades, huge efforts have been given towards the development of 3D technology. Meanwhile, some works have been carried out to devise quality metrics for 3D images [209, 264]. Since, in 3D visualization, depth information is an important information, few works [21, 90, 125, 260] have been devoted to understanding the influence of depth map quality in quality perception of 3D video, and results show that 3D visual quality highly depends on the depth map quality.

In [21] effect of depth map quality on the perceptual quality of synthesized view is studied, and presented results show that the 3D video quality is depending on the depth map quality. The significant correlation between the perceived overall image quality and perceived depth is shown in [90]. The results presented in [125] show that the distortion in depth map results degradation in the perceptual quality of synthesized view. In [260] impact of disparity in stereoscopic image quality is presented. The achieved results show that the strong correlation between quality of disparate images and overall perceived quality of the image. The disparity information is used to improve the

performance of existing 2D metrics. Similarly, the depth map information has been used to improve the performance of 3D VQM in [25, 151, 82, 146].

9.2.3   Light field image quality dataset

To train, test, and benchmark the objective IQM an availability of test LF images and corresponding subjective quality rating, MOS, is crucial. In literature, few efforts have been given to create LF image datasets. The list and features of recently proposed LF image dataset are presented in [176]. However, publicly available datasets do not include information about test LF images and annotated subjective quality scores. Therefore, in Section 9.4, a recently created LF image quality dataset is presented, and it has been used to benchmark the IQMs for LF image.

9.3   Proposed light field image quality assessment framework

The quality of depth map is important in LF imaging, since it is used for a wide range of applications such as encoding and 2D/3D rendering [71] [116]. In particular, errors in depth values at a given pixel position, affect the quality of the rendered view where this pixel will be used for rendering. Even small errors in depth can lead to significant errors in the rendered view, and thus the proposed approach relies on the exploitation of the depth map quality to estimate overall quality of LF image.

A reduced information about LF image, i.e. depth map, has been used to predict the quality of distorted LF image. Quality evaluation framework of the proposed metric is shown in Figure 9.1. The followed steps are:

1. reference depth map is estimated from the reference LF image. In the following, the reference depth map is referred as DM_{ref}. The DM_{ref} has been used to estimate the level of distortion on the depth map;

2. distorted depth map, referred as DM_{dis}, is computed from the distorted LF image;

3. level of distortion in depth map is computed based on DM_{ref} and DM_{dis} as:

$$Dis = f(DM_{ref}, DM_{dis}), \tag{9.1}$$

   where Dis is the measure of distortion on the depth map and $f(.)$ is the function to represent the FR IQMs such as SSIM;

4. finally, a mapping model is used to estimate the perceptual quality of test LF image from the Dis.
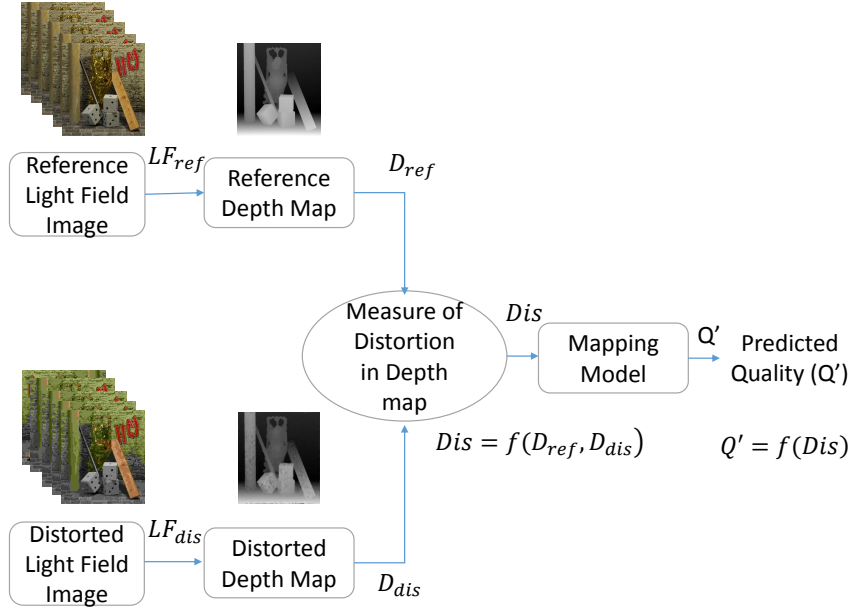
Figure 9.1: Proposed Reduced Reference (RR) light filed IQM framework.

9.3.1   Depth map estimation

To evaluate the proposed framework, DM_{ref} and DM_{dis} have been computed by using three LF image depth map estimation algorithms: Multi Resolution Depth Map (MRDM) [167], Stereo-Like Taxonomy Depth Map (SLTDM) [34], and Accurate Depth Map (ADM) [115]. The brief description of the algorithms is presented in the following:

- MRDM: In this technique, multiple views of a scene are used for estimating the depth map. First, local estimates of depth is computed based on the optimization of a log-likelihood function, defined as a difference of sub-aperture images: center view versus the other views. To enhance the performance of depth estimate in flat regions, additional constraints for smoothness and occlusions are incorporated in the optimization function. Next, a multi-resolution approach is adopted. Basically, at each resolution level, the depth map is locally estimated for reducing the complexity compared to global optimization. This approach helps to face the potential accuracy losses in depth map resulted due to the presence of flat zones and to preserve the edges. To improve the quality of noisy depth maps resulting from local optimization, de-noising is performed by using 2D weighted median filters.

- SLTDM: In this method, the taxonomy of stereo algorithms is used. It is the method of

(a) Light filed image

(b) MRDM   (c) ADM   (d) SLTDM
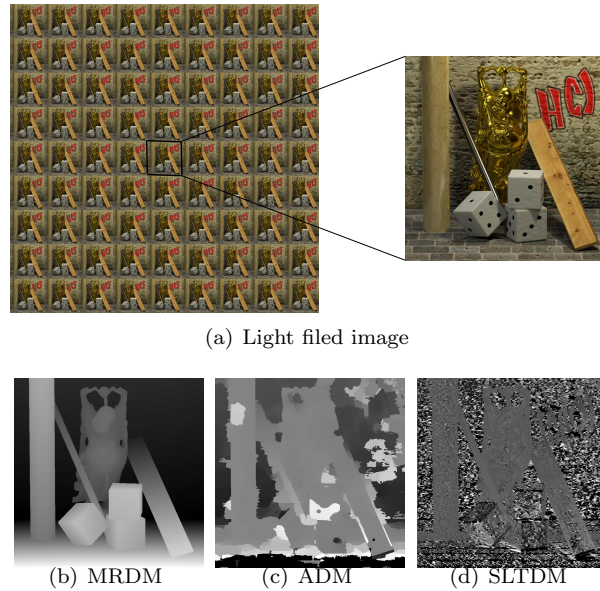
Figure 9.2: LF image, Buddha2, and estimated depth maps using three different methods: MRDM, ADM, and SLTDM.

determining the distance to a point seen by a pair of stereos cameras, i.e. finding the disparity between the images of two reflected cameras. In brief, a cost volume is computed by comparing each pixel in a sub-aperture image with the pixel of all other sub-aperture images. Next, cost aggregation is performed, and the disparity is selected for minimum cost per pixel.

- ADM: In this method, stereo matching between the sub-aperture images is presented. In brief, the phase shift theorem in Fourier domain is exploited to estimate the pixel shifts of sub-aperture images. A cost volume is computed to evaluate the matching cost of disparity levels with the help of sub-aperture images and central view sub-aperture images shifted at different sub-pixel locations. The gradient matching costs are adaptively aggregated. Next, a weighted median filter is adopted to remove the noise in cost volume, and multi-label optimization is performed for reliable disparity at weak texture region. Finally, iterative polynomial interpolation is performed to enhance the estimated depth map.

The considered LF image format and the estimated depth maps are shown in Figure 9.2. It shows that the quality of estimated depth maps varies from high to low for the considered depth map estimation methods. In the context of performance evaluation of the proposed quality assessment framework, consideration of the best to worst performing depth map estimation methods is important.

The influence of compression and noise artifacts in the estimated depth map is shown in Figure 9.3. It shows that the artifacts in LF image degrade the quality of the estimated depth map.

(a) JP, bpp= 0.99 (b) JP2, bpp= 1.43 (c) S and P, ND=0.5

(d) S and P, ND=0.5 (e) GN, var=0.001 (f) GN, var=0.0001
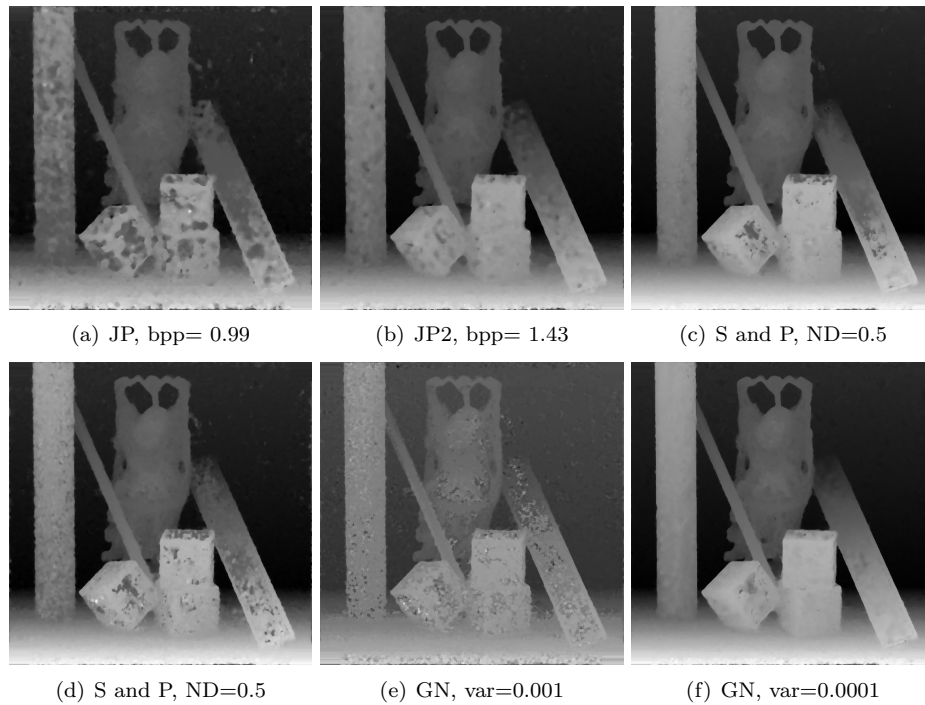
Figure 9.3: Depth map (estimated by MRDM) of distorted LF images. Distortion on the image is a result of JPGG (JP) and JPEG 2000 (JP2) compression and Gaussian Nosise (GN) and Salt & Pepper Noise (S and P). Where, var is variance and ND is noise density.

The result indicates that the distortion in the light filed image is replicated in the estimated depth maps.

9.3.2 Measure of distortion in depth map

Measure of distortion (Dis) on the distorted depth map DM_{dis} is estimated by using the reference depth map DM_{ref}. Our goal is to compare the two depth maps, DM_{ref} and DM_{dis}, and estimate the degree of similarity between them. In the following, DM_{ref} is assumed pristine original and DM_{dis} is the distorted/contaminated by noises. To analyze the impact of LF image distortion on depth map, the depth maps of distorted LF images have been computed. As a measure of distortion, SSIM is estimated for the distorted depth maps with the help of a reference depth map. Assuming, the structural information is more important than color information in depth map, SSIM has been used, as suggested in [249, 250]. In brief, SSIM assumes that the natural images are highly structured, and that the HVS is more sensitive to structural distortion. In particular, it defines the function for the luminance (l) comparison of the signals, the contrast (c) comparison of the signals, and the structure (s) comparison of the signals, respectively, as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + c1}{\mu_x^2 + \mu_y^2 + c1}, \tag{9.2}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c2}{\sigma_x^2 + \sigma_y^2 + c2}, \tag{9.3}$$

$$s(x, y) = \frac{\sigma_{xy} + c3}{\sigma_x\sigma_y + c3}, \tag{9.4}$$

where, μ_x and μ_y are the local sample mean, σ_x^2 and σ_y^2 are local variance, σ_{xy} is a local sample correlation coefficient (covariance) of x and y, and these local sample statistics are computed within overlapping windows. Moreover, $c1 = (k1L)^2$ and $c2 = (k2L)^2$ are two variables to stabilize the division with weak denominator (when denominator(s) becomes small), and L is the dynamic range of the pixel values ($2^{\#\ \text{number of bits per pixel}} - 1$), and k1=0.01, and k2=0.03. Moreover, the c3 is equal to c2/2.

Then, SSIM index is a weighted combination of those comparative measures $l(x, y)$, $c(x, y)$, and $s(x, y)$), and expressed as:

$$SSIM(x, y) = [l(x, y)^\alpha . c(x, y)^\beta . s(x, y)^\gamma], \tag{9.5}$$

where, weights α, β, and γ are used to indicate the relative importance of the three components. If we consider all the three components are equally important ($\alpha = \beta = \gamma = 1$), the Equation 9.5 becomes:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)}. \tag{9.6}$$

Results, presented in Figure 9.4, show that the distortion of LF image is clearly replicated in the estimated depth maps, and that the measure of distortion on depth map (computed as a SSIM) is related to perceptual quality of the depth maps.

(a) Reference   (b) GN (0.873)   (c) JP2 (0.805)   (d) JPEG (0.964)   (e) S&P (0.951)
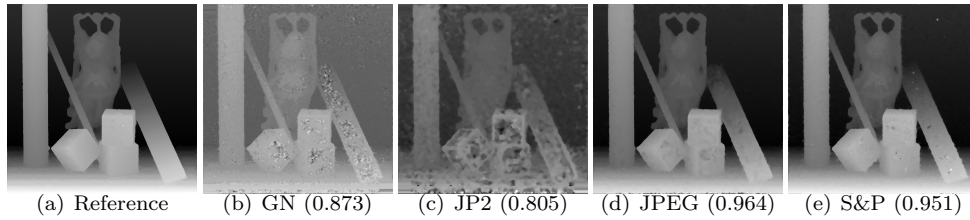
Figure 9.4: Depth map (estimated by MRDM) of the test light filed images and measure of distortion on depth map as SSIM. Test LF images have been created by considering JPEG and JPEG 2000 (JP2) compression and Gaussian and Salt & Pepper noise)

9.3.3   LF image quality estimation

In literature, many studies have been performed to understand the human perception and the stimulus. As mentioned before, the fundamental principle relating the human perception to the relative change of stimuli is well defined by Gustav Fechner, as Weber-Fechner Law [197]: "the differential perception is directly proportional to the relative change of the physical stimulus". Therefore, a mapping model is needed to estimate the perceptual quality score, Q', from a measure of distortion on depth map, Dis. In this study, as a mapping model, four parameter logistic function (Equation 9.7) has been used [228].

$$Q' = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{Dis - \beta_3}{|\beta_4|}\right)}}, \tag{9.7}$$

where, Q' is the estimated perceptual quality score from the measure of depth map distortion, Dis.

9.4   Subjective light field image quality assessment

In this section, the steps that have been followed for the subjective experiment are reported.

9.4.1   Test light field images

Source sequences (SRCs)

The SRCs have been taken from the dataset presented in [252]. A brief introduction of the SRCs is already presented in Section 8.3. The variety of scene content provided by the selected SRCs is confirmed by analysis results presented in Figure 10.3: SRCs are distributed over a wide range of spatial perceptual information and colorfulness plane.

Distorted LF images

In this work, encoding and noise artifacts are considered as the HRCs. To the best of our knowledge, there are no standard LF image compression method, and thus the standard image compression methods, JPEG and JPEG 2000, have been considered. Moreover, as a consequence of image acquisition and processing such as encoding and decoding, Gaussian and Salt & Pepper noise artifacts have been considered.

   To create a test image sequences, four to five levels of distortion are sufficient [30] [186]. In this work, the distortion strengths are selected manually. In particular, a large set of distorted LF images is generated, and a subset of test images spanning a desired image quality range, is chosen by the author. In other words, the distortion levels are selected based on the noticeable perceptual quality deference, JND.
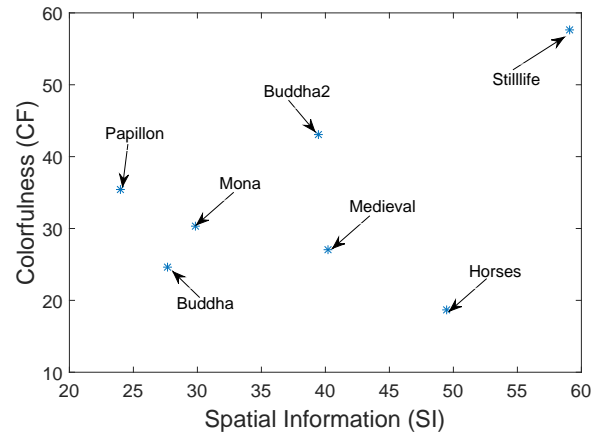
Figure 9.5: Contents variations in reference LF images measured in terms of spatial perceptual information (SI) and colorfulness (CF).

9.4.2 Experiment setup

Test methodology: According to ACR method [112], based on the perceived quality, the stimuli are rated from one to five (bad to excellent quality).

Display device: The LF images were displayed on a 2D DELL monitor (model no: DELL U2413f). *Stimuli arrangement:* To compensate the effect of a potential bias based on order or position of stimuli in the averaged results, the stimuli have been shown in pseudo-random order according to the distortion level. Moreover, test LF images from the same SRC are not displayed at least for next four conjugative test LF image pairs, in order to remove the memory and contextual effect in quality judgement. Four experimental sessions have been scheduled to minimize the effect of viewers fatigue on quality assessment. Moreover, at least 30-minute gap between the first two and the last two sessions was maintained to retain the attention of the subject [112] towards the stimuli.

Subjects: The reliability of achieving results also depends on the number of subjects that have been used for the subjective experiment. To this aim, we exploited 19 subjects, for quality evaluation as suggested in [121] [109]. The subjects were drawn from a pool of undergraduate and graduate students from Università degli Studi Roma TRE. The students were relatively naive concerning the impairments and associated terminology.

Training: Before starting the experiment, verbal instructions were given to the subjects. Then a training phase was scheduled. Each subject was shown the original LF images followed by examples of the images with the strongest artifacts found in the experiment. For this purpose, a subset of four LF images has been taken from the pool of the test images. The purpose of the training session was to get each subject familiar with the assessment procedure and to establish the annoyance value range according to the distortion levels.

9.4.3   Subjective experiment results analysis

For the collected opinion scores processing, the procedures presented in Section 4.4 is adopted. In brief, followed by the outlier detection procedure: detecting and removing scores given by the subjects whose score is very far from the mean behavior, the MOS and 95% CI are computed. In this experiment, no outliers were found, and thus, the scores given by all the subjects have been considered.

Perceived quality of test LF images

The quality scores distribution expressed in terms of MOS for the test LF images is shown in Figure 9.6. It can be noticed that the perceived quality of test images (MOS) is uniformly distributed over the range between 1 to 5 (bad to excellent quality). This result indicates that the selected levels of the distortions are sufficient to cover a wide range of quality levels.
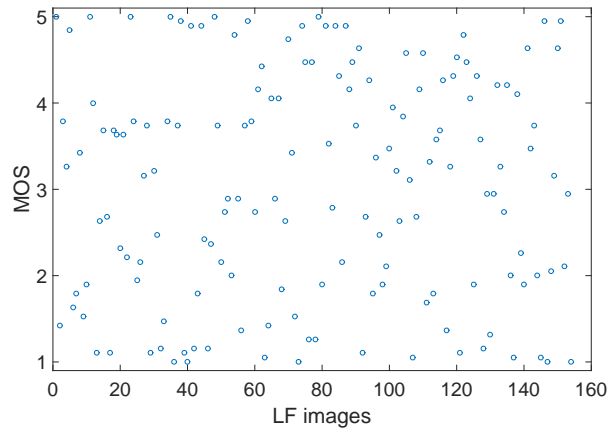
Figure 9.6: Scatter Plot of the MOS scores obtained for test LF images.

The average perceived quality of the distorted LF images is shown in Figure 9.7. We can notice that the perceived quality of the images for HRCs have a noticeable quality difference, JND. The result is confirmed by ANOVA, the result ($F(18, 114) = 62.91$ and $P_{value} \simeq 0$,) indicates the QoE for the HRCs is significantly different

9.5   Results and discussion

In this section, by exploiting the results of the subjective experiment, the performance of the proposed LF IQA framework is evaluated; the performance of 2D IQMs, when applied to LF image, is tested; and finally, perceptual quality of standard image compression methods has been evaluated for LF image.
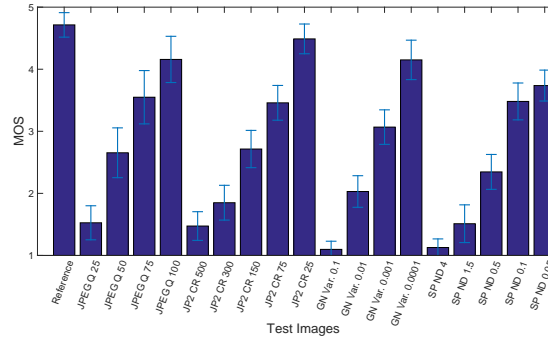
Figure 9.7: MOS scores and 95% CI for the test LF images. In the figure, Q = quality levels for JPEG compression, CR = compression ratio for JPEG 2000 (JP2) compression, Var.= variance for Gaussian Noise (GN), and ND = Noise Density for Salt & Pepper (SP) Noise.

9.5.1   Validation of the proposed Reduced Reference light field image quality assessment framework

At first, the relationship between a measure of distortion on the depth map (Dis) and MOS is studied. For the study, Dis is estimated for test LF images by using the depth map estimation methods, SLTDM, ADM, and MRDM, and the result is plotted vs MOS. Dis is the similarity measure between the distorted depth map and reference depth map estimated through SSIM. The plots, in Figures 9.8, 9.9, and 9.10 show that there is a strong relationship between Dis and MOS for all the considered depth map estimation methods.

To analyze the performance of the proposed RR LF IQM, the procedure described in the final report from the VQEG on the validation of objective models of VQA [228] has been followed. The estimated perceptual quality is computed by using a four parameter logistic function (Equation 9.7).

To set the optimal parameter of the logistic function, a nonlinear regression has been performed. Initial parameters for the logistic function have been selected as specified in [228]. The proposed approach needs a training process to find the parameters, thus the dataset has been divided in two randomly chosen non-overlapping parts. As general practice [163, 132, 159, 160], training set includes 80% of data and 20% has been used for testing purpose.

Finally, correlation analysis between estimated subjective score and ground truth subjective scores (MOS) have been performed by means of PLCC, SRCC, and KTCC. The result, in Table 9.1, shows that there is a very high correlation between estimated quality scores with MOS for the considered depth estimation methods. As mentioned before, in Figure 9.2, the quality of estimated depth maps (computed by using different depth estimation methods) is significantly different. Though, the result shown in Table 9.1, correlation between the estimated quality and MOS, is very high for all the methods. This result is due to the exploitation of hidden reference removal technique [112]:

Figure 9.8: Relationship between MOS and depth map distortion measure, Dis. The curves are result of second order polynomial: $MOS = p1 * Dis^2 + p2 * Dis + p3$ and the goodness of the fitting is expressed in terms of R-Square and Root Mean Square Error (RMSE). The depth maps have been extracted by using SLTDM and level of distortion in the depth map is estimated as SSIM.

considering the reference information (reference depth map), to compute distortion on depth map (Dis), and ultimately for estimating the perceived quality of LF image. We can also notice that the performance of metric is varied for depth estimation methods. Among considered depth estimation methods a recently proposed algorithm ADM has a high value of correlation coefficients.

To confirm these results, PCA has been performed. The result of PCA (bipolar plot), Figure 9.11, shows ADM has a small angle with respect to subjective score. Thus, we can conclude that among the considered depth map estimation algorithms, ADM has a better correlation compared to other considered algorithms. This could be due to the fact that ADM method is effective in terms of utilizing the sub-pixel shift in the frequency domain. The employed aggregation of the gradient costs and confident matching technique helps to enhance the depth map accuracy. Also, in the point of accurate depth estimation capability, the ADM outperforms other advanced methods [115].
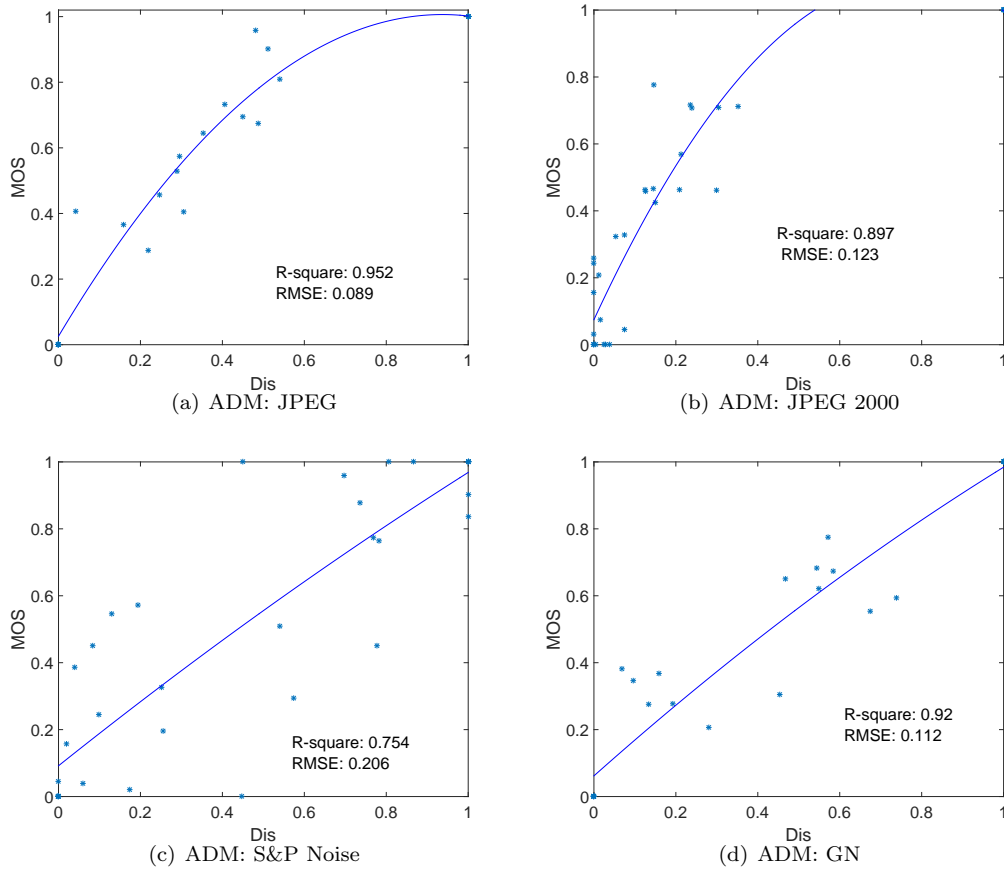
Figure 9.9: Relationship between MOS and depth map distortion measure, Dis. The curves are result of second order polynomial: $MOS = p1 * Dis^2 + p2 * Dis + p3$ and the goodness of the fitting is expressed in terms of R-Square and Root Mean Square Error (RMSE). The depth maps have been extracted by using ADM and level of distortion in the depth map is estimated as SSIM.

To compare the performance of the proposed LF IQM, the 2D quality metrics have been evaluated for LF image in the following subsection.

9.5.2 Performance analysis of 2D image quality metrics, when applied to light field image

To analyze performance of 2D IQMs, when applied to LF image, widely used NR and FR metrics were selected.

- *NR Metrics:* Blind Image Quality Index (BIQI) [162], Naturalness Image Quality Evaluator (NIQE) [160], and Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [159].

- *FR Metrics:* Mean of Squared Errors (MSE), PSNR, SSIM, Multi-scale Structural Similarity
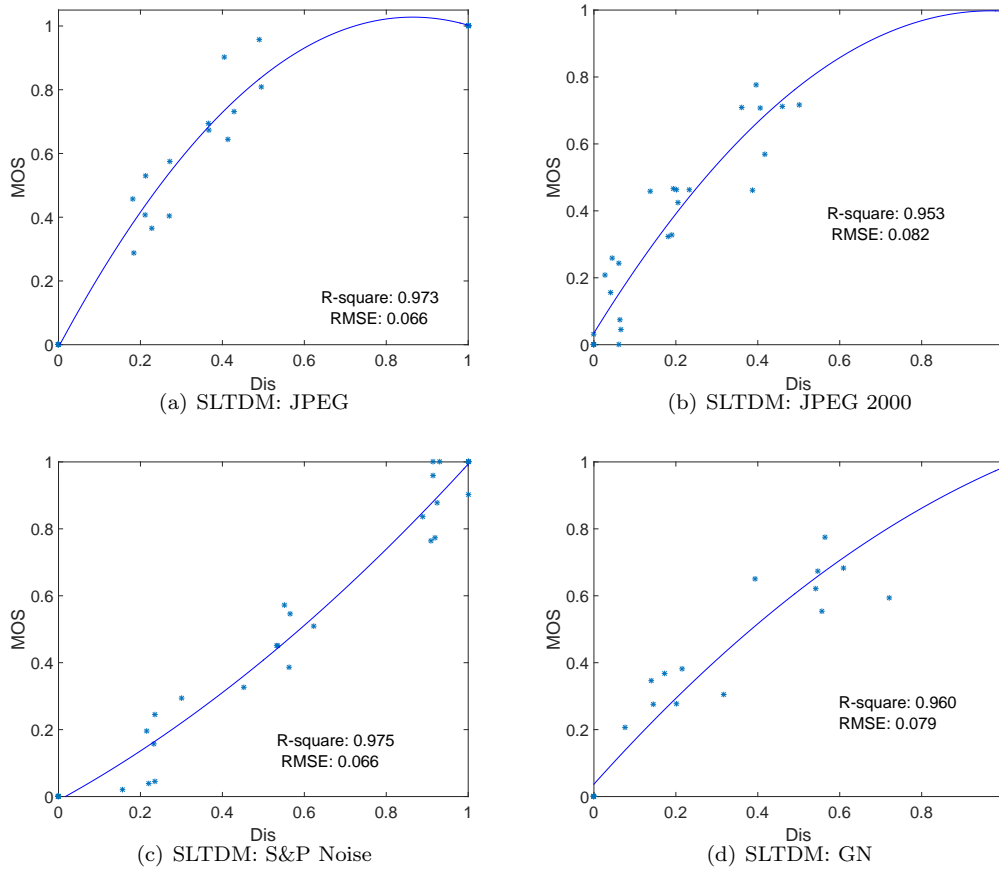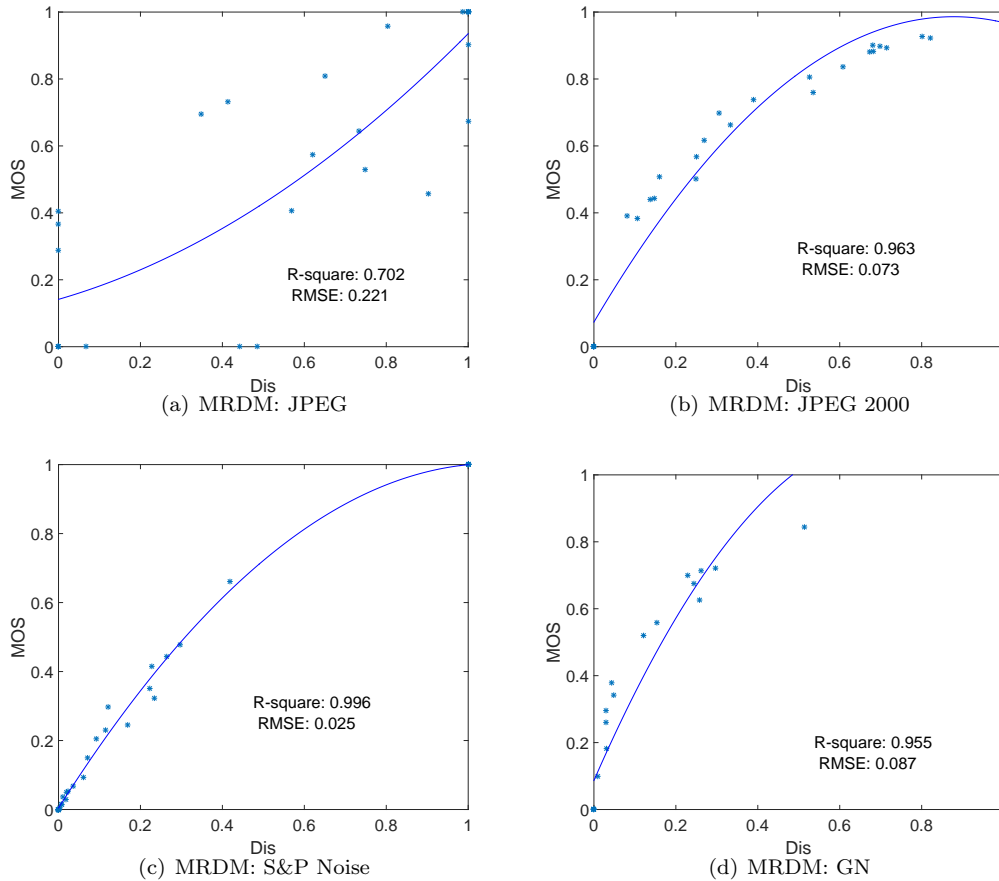
Figure 9.10: Relationship between MOS and depth map distortion measure, Dis. The curves are result of second order polynomial: $MOS = p1 * Dis^2 + p2 * Dis + p3$ and the goodness of the fitting is expressed in terms of R-Square and Root Mean Square Error (RMSE). The depth maps have been extracted by using MRDM and level of distortion in the depth map is estimated as SSIM.

Index (MSSIM) [251], Visual Signal-to-Noise Ratio (VSNR) [37], Image information and Visual quality (VIFP) [210], UQI [248], Signal-to-Noise Ratio (SNR), Weighted SNR (WSNR) [158], PSNR-HVS [57], and PSNR-HVS-M [187].

To evaluate the performance of the objective metrics, the procedure recommended in [228] (as discussed in Section 9.5.1)was followed. The achieved results from the correlation coefficient analysis are summarized in Table 9.2. The result shows that the metrics VIFP and UQI have higher values of PLCC, PSNR and MSSIM have higher values of the SRCC, and MSSIM has higher KTCC. These results indicate most of the considered metrics perform almost similarly. However, the correlation coefficients are not significantly high. To confirm the result, PCA has been performed. The result of PCA, Figure 9.12, shows that two principal components explain around 83% of the total perceptual

Table 9.1: Correlation between the estimated quality score and MOS. Performance of proposed quality assessment framework is evaluated for depth map estimation algorithms, MRDM, SLTDM, and ADM

	PLCC	SRCC	KTCC
MRDM	0.89	0.89	0.71
SLTDM	0.90	0.90	0.74
ADM	0.94	0.94	0.80

quality, and the metrics PSNR, VIFP, UQI, PSNR-HVS, PSNR-HVS-M, WSNR, SSIM, and MSSIM are concentrated in the same region of the plot, and the angle with MOS is also comparable.

Table 9.2: Performance analysis of the IQMs when applied for LF image. The correlation coefficient between subjective quality score (MOS) and estimated quality score by the IQM is computed.

IQMs	PLCC	SRCC	KTCC
BIQI	0.413	0.506	0.320
NIQE	0.843	0.837	0.646
BRISQUE	0.476	0.103	0.080
MSE	0.836	0.813	0.662
PSNR	0.818	**0.864**	0.700
SSIM	0.825	0.849	0.673
MSSIM	0.710	**0.876**	**0.720**
VSNR	0.816	0.865	0.680
VIFP	**0.847**	0.826	0.646
UQI	**0.847**	0.858	0.680
WSNR	0.808	0.856	0.700
PSNR-HVS	0.732	0.840	0.660
PSNR-HVS-M	0.727	0.849	0.680

The results, presented in Table 9.2 and Figure 9.12, indicate that the performance of most of the metrics is comparable. Moreover, there is no metric that has a significantly higher value of the correlation coefficient. This result strengthens the needs of a new IQM for LF image.

Moreover, to analyze the impact of the image scene on the estimated quality a two-way-ANOVA is performed. The result (P_{value} for interaction is equal to one indicates that there is no evidence of the interaction of the image scene on the estimated quality scores. In other words, the estimated quality scores are not influenced by the image content or SRCs.

Finally, the presented results in Table 9.1 (PLCC = 0.890, 0.899, and 0.940 for MRDM, SLTDM, and ADM respectively) show that the proposed RR light filed IQM have the higher values of correlation coefficients compared to well known 2D metrics (as presented in Table 9.2) for all the considered depth map estimation methods.
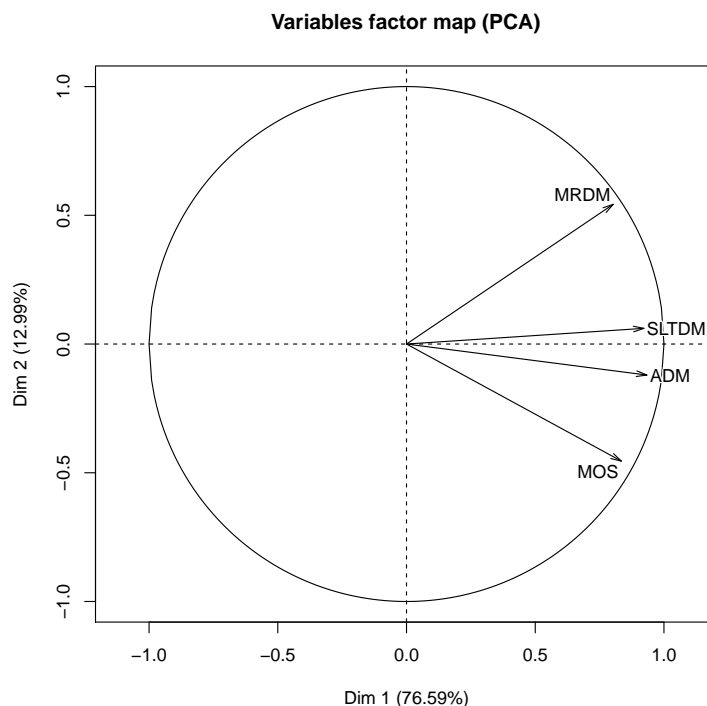
Figure 9.11: The performance analysis of the proposed RR LF IQA framework for depth map estimation algorithms MRDM, SLTDM, and ADM.

9.5.3 Perceptual quality analysis of encoding methods for light filed image

During the experiment design, two compression methods JPEG and JPEG2000 have been used to generate PISs. The results achieved from the subjective experiment allow to evaluate perceptual quality of compression methods for LF images. The result of the analysis is plotted in Figures 9.13 and 9.14. As for 2D images, JPEG2000 better performs than JPEG for LF images.

However, from Figures 9.13 and 9.14 we can also notice that for some images (in particular, Buddha, Buddha2, Horses, and Mona), at high level of compression, i.e. at low level of bpp, the JPEG compression results in a high MOS compared to JPEG2000. This is due to the fact that, at high compression rate JPEG2000 produces prominent blurring and ringing artifacts, having high impact on the light filed image QoE compared to the blocking artifact produced by JPEG. This result indicates that, at a high level of compression the JPEG2000 is not an appropriate choice for the images that include noisy textures, as in the images Horses and Buddha2 [114].
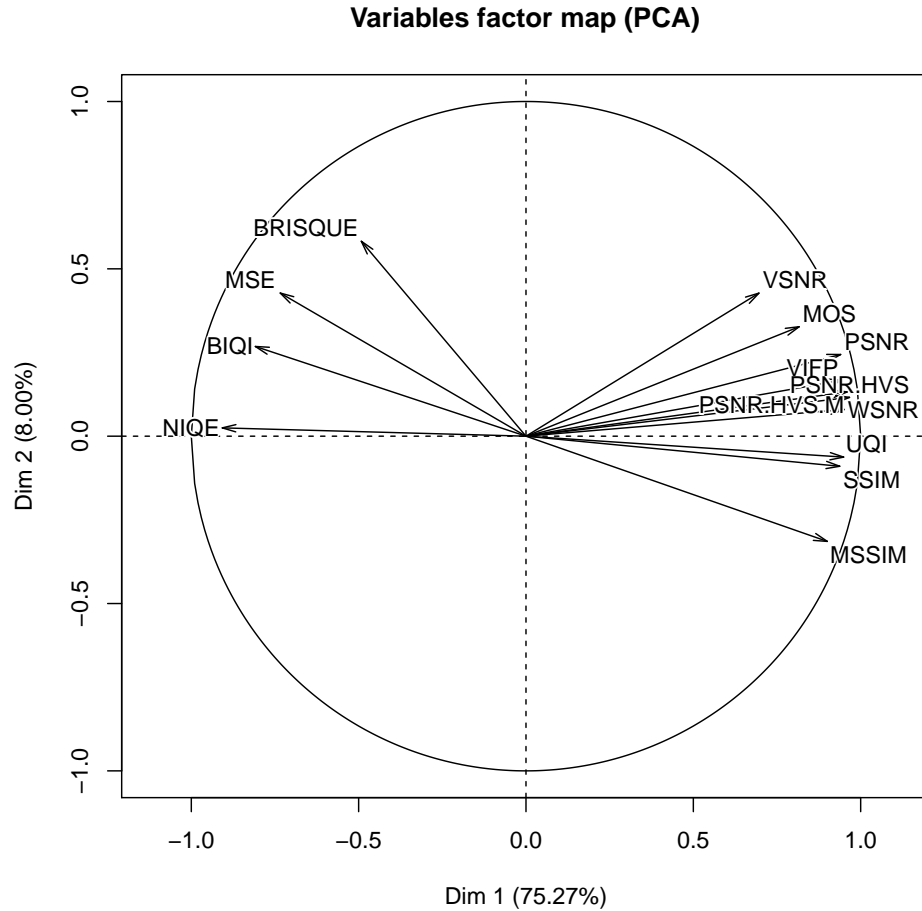
Figure 9.12: Performance analysis of 2D image/VQMs, when applied for LF images. The small angle between the variables (MOS and metrics) corresponds a high correlation.

9.6 Conclusion

In this contribution, a RR LF IQA framework is proposed. The measure of distortion in depth information is used to compute the estimated quality score. Achieved results show high correlation between estimated quality score and corresponding subjective quality ratings.

To collect the subjective quality ratings for test light filed images, a subjective experiment has been performed. Given the novelty of the imaging system, there is no validated subjective quality assessment procedure for the image, the steps that have been followed during the experiment design are tailored in such a way that it can be used as a reference material for the subjective study on this topic.
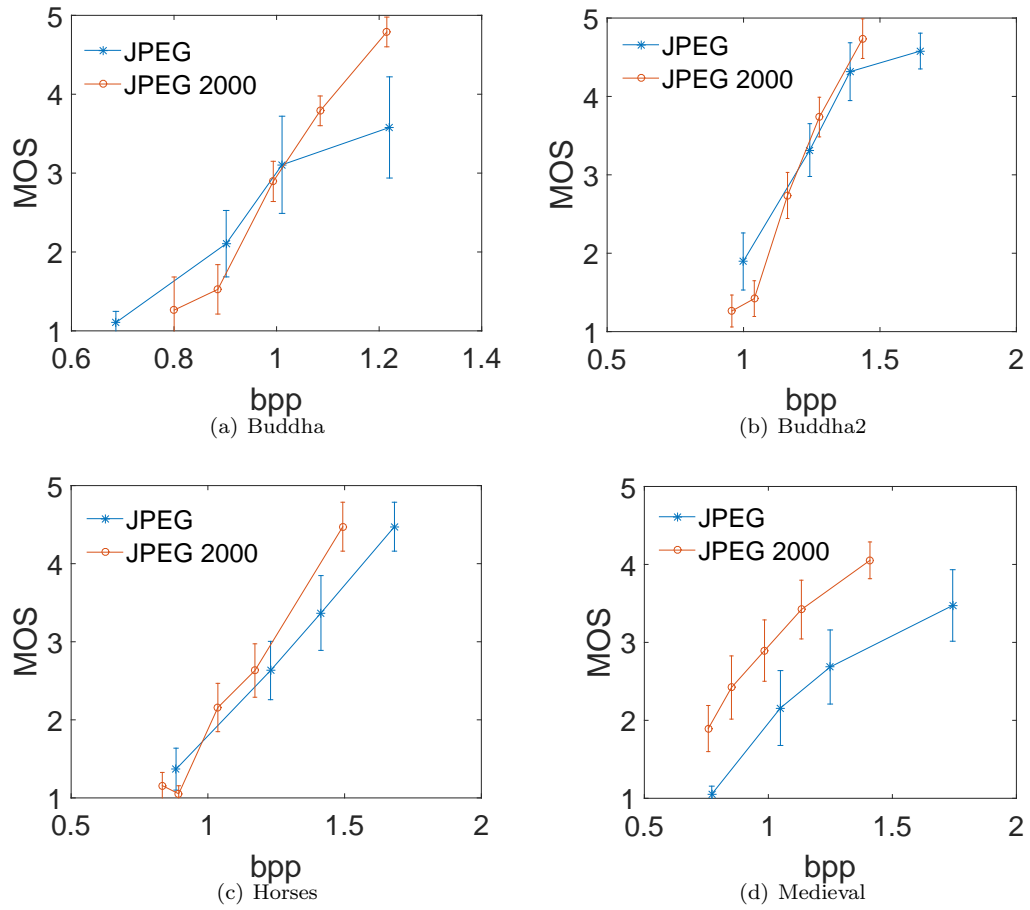
Figure 9.13: Quality of experience of JPEG and JPEG2000 compressed four LF images: Buddha, Buddha2, Horses, and Medieval.

The dataset is also used to evaluate the performance of state-of-the-art IQMs, when applied to LF images. The results of the analysis strengthen the need of a new IQM tuned to LF image.
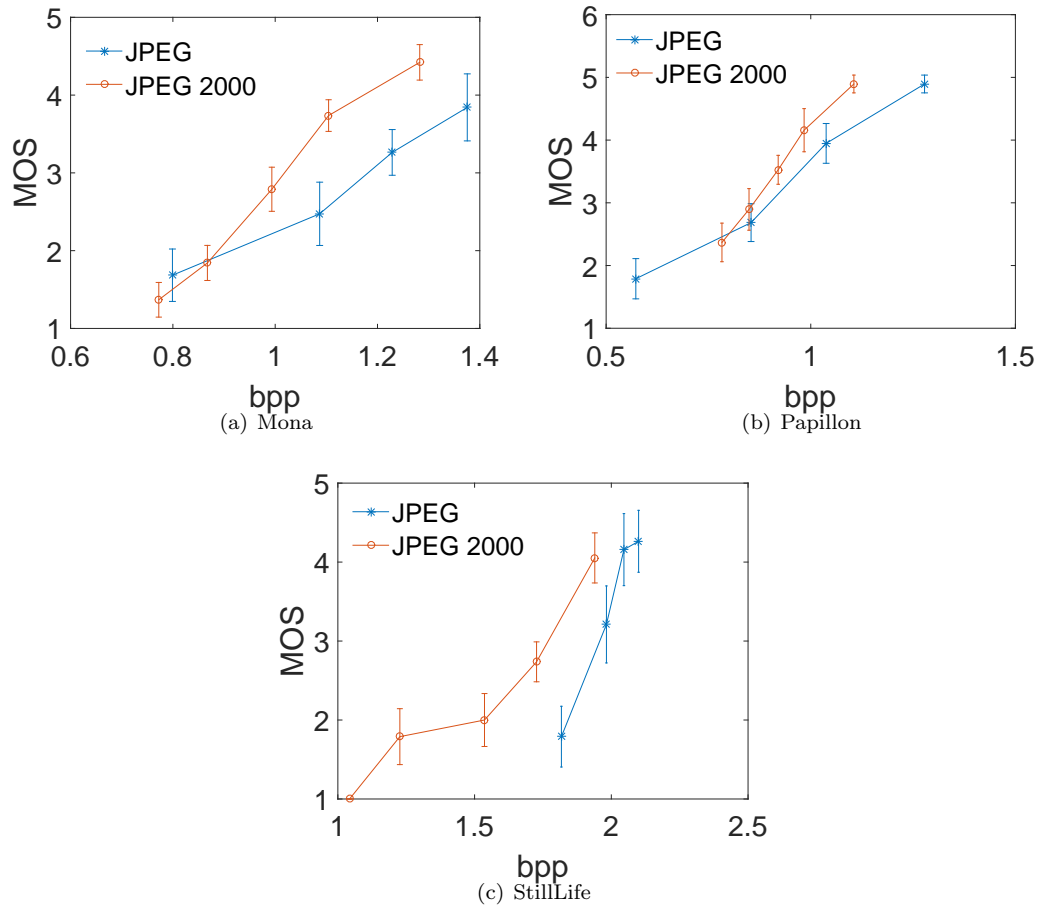
Figure 9.14: Quality of experience of JPEG and JPEG2000 compressed three LF images: Mona, Papillon, and StillLife.

Chapter 10

Subjective quality assessment of light field images

The major contribution of this chapter is threefold: a generic LF IQA framework is proposed. Since, there is no standard or guideline for LF image quality assessment, the adopted subjective experiment procedure has been tailored in such a way that it could be used as a reference material for developing the standard quality assessment protocol for LF image; by following the framework, a LF image quality dataset, called SMART LF image quality dataset, is created and making freely available to the research community; the dataset is extended by considering more SRCs and rendering methods; and finally a LF image processing experiment setup is proposed for subjective quality assessment.

10.1   Introduction

The wide range of possible applications and the rapidly developing LF technology pulled the attention of the consumer, industry, and academics. LF images are subject to a wide variety of distortions during acquisition, processing, compression, storage, transmission, and rendering; any of these steps may result in a visual quality degradation. The rapidly developing LF technology and consumer interest towards this technology is pushing the need for QoE evaluation of such contents.

As mentioned before, due to the novelty in technology, there is a lack of standard LF image quality methods (subjective as well as objective). In literature, only a few works are performed towards the quality assessment direction. However, the knowledge of degraded quality (quality level) is crucial to benchmark the developing LF image processing algorithms. In this scenario, in previous chapters, particularly in Chapters 7, 8, and 9 significant efforts are given to understand the quality issues in LF imaging. Now, it is the time to propose a subjective quality assessment framework for light filed image, and it is the scope of this chapter. This chapter presents a general subjective quality

assessment framework for LF image. The proposed framework is equally applicable for the LF images captured using different techniques: mircrolens array (Lytro Illum and Raytrix), synthetic, grid of a multi-camera system, etc.

10.1.1   Related works

In [174], also mentioned in Chapter 8, the perceptual quality of watermarked and distorted LF image is presented. Moreover, a LF image quality dataset and a Reduced Reference quality metric tuned to the LF image are presented in Chapter 9. In both of the works Granty type LF images are used. Since, hand held LF camera (Lytro Illum) is already available in the market for general photographers, and for research purposes. It would be nice to consider the Lytro Illum recorded LF image.

Moreover, a method for subjectively assessing the quality of plenoptic content is presented in [239] [238]. In the method after demosaicing and divignetting of raw sensor data, encoding is performed. Following the encoding process, color and gamma correction is applied to the encoded lenslet image. The proposed method could be a good use case for encoding of plenoptic content for reducing the size of camera recorded information. However, this quality evaluation framework has two major shortcomings: i) color and gamma correction is performed after the decoding process, thus the impact of encoding in color component will not be evaluated during the subjective/objective quality assessment; ii) after the decoding the camera calibration data are being used during the color and gamma correction: it limits the usability of the method in image/video communication environment. In particular, the need of (big size of) camera calibration data at the receiver side for color and gamma correction may not be practical. In this scenario, an availability of generic subjective quality assessment framework for LF image is crucial.

10.1.2   Light field image datasets

The list and basic features of state-of-the-art LF image quality datasets are reported in Table 10.1. In literature, no dataset is available with test LF images and annotated subjective quality ratings.

Advances in LF imaging technology and the availability of commercial LF cameras (Lytro Illum and Raytrix) allow the consumer to exploit such a technology. The above mentioned datasets are not sufficient to deal with new challenges (perceptual quality evaluation, performance testing for processing algorithms, etc.) arising with the advancement of the LF technology.

Among the datasets, the LF Image Dataset [200] includes the light field images captured by using Lytro Illum camera. However, as clarified in [176], the motivations behind the image content selection are not being reported and images present a huge amount of redundant information.

The novelty of the media has an impact on the subjective evaluation too. To collect the subjective opinion scores for the test LF images a subjective quality assessment experiment need to be designed. In literature, many standard guidelines [121] [112] [108], have been recommended to design the

Datasets	Acquisition Devices	Purpose	Features
SMART (2016) [176]	Lytro Illum	general	It includes well characterized 16 LF images, and useful for design, testing, and bench-marking LF image processing algorithms
LF Image Dataset (2016) [200]	Lytro Illum	general	It includes 118 LF images of the category Buildings, Grids, Mirrors and Transparency, Landscapes, Nature, ISO and Color charts, People, Studio, Urban, and Light)
Lytro Illum LF Dataset (2016) [7]	Lytro Illum	general use	It includes 43 LF images including indoor and outdoor, varying lightening conditions, translucent, texture, and calibration grid images
4D LF Dataset (2016) [246]	Lytro Illum	material recognition	It contains LF images of 12 material categories: fabric, foliage, fur, glass, leather, metal, plastic, paper, sky, stone, water, and wood, and each categories with 100 LF images
LCAV-31 (2013) [73]	Lytro	Object Recognition	It provides LF images of 31 object categories captured from ordinary household objects and designed for object recognition purpose.
Lytro dataset (2015) [165]	Lytro	LF Reconstruction	It provides 30 images, with indoor and outdoor, motion blur, long exposure time, and flat image.
LF Saliency Dataset (2014) [138]	Lytro	saliency map estimation	It provides 100 LF images with 60 indoor scenes and 40 outdoor scenes.
GUC LF Face and Iris Database (2016) [188]	Lytro and DSLR	face and iris Recognition	It provides two biometric image databases collected by using a Lytro camera on multiple faces and visible iris (2986 faces from 112 subjects and 55 pairs of eyes).
The (New) Stanford LF Archive (2008) [233]	Lego Gantry	General	It includes 13 LF images, resolution: 17x17 and various image resolutions
	LF microscope		3 LF images, low resolution
	LF Gantry		4 LF images, resolution 21x5x650x515 and 16x16x650x515
	Camera Array		2 LF image, 88 views of 640x480 and 45 views of 640x480
The (Old) Stanford LF Archive [1]	Synthetic	general	It provides low resolution LF images.
Synthetic LF Archive (2013) [3]	Synthetic	Compression	It provides 13 LF images (5x5 or 7x7 views) including images with transparencies, occlusions, and reflections.
Datasets and Benchmarks for Densely Sampled 4D LF (2013) [252]	Blender Software and Gantry device	Depth Map	It provides 7 Blender and 6 Gantry images; however, images do not cover the wide range of natural scenes.

Table 10.1: LF image datasets with corresponding basic features.

subjective experiment for images and videos. However, currently there is no guideline defined for LF images.

10.2 A generic subjective quality assessment framework

In the context of the lack of standard/validated subjective quality assessment procedure for LF image, before designing the subjective experiment to create a LF image quality dataset, a brief discussion on the generic LF IQA framework is presented.

10.2.1 Source sequence

As mentioned before, the scene selection is one of the most important steps in the experimental design. Some of the scene selection criteria are already defined in Section 3.2.1. Moreover, in literature, some efforts have been given to define the scene selection criteria for 2D and 3D visual contents. In brief, techniques for choosing video sequences for subjective experiment is presented in [180]. In this article, a semi-automatic scene selection technique is proposed by using the attributes: brightness, colorfulness, amount of background motion, amount of foreground motion, amount of scene cuts, amount of cartoon content, interesting content/easy to watch, particularity level, spatial detail complexity, amount of camera motion, and content type. Article [256] presents metrics: view mismatch, disparity range, divergence, and disparity change, to estimate parameters relevant for stereoscopic 3D video content in order to measure viewing comfort or 3D QoE. In [232], image condition indoor/outdoor, descriptions of semantic level features such as news, person, and playing, shooting conditions (distance and angle), source compression methods, spatial and temporal information about the scene and depth map, disparity histogram, disparity range, and coding parameters are used to describe the 3D video sequences.

In this dissertation, reference images are chosen in such a way that the selected SRCs should span a wide range of content features. For this purpose, together with basic visual signal features (SI, CF, contrast, correlation, homogeneity, brightness, hue, and saturation), LF camera specific capabilities are also considered.

General image features

The features need to be considered during the scene selection is already presented in Section 3.2.1. The inclusion of those features is also important in LF imaging.

LF image related features

In the context of LF imaging, inclusion of other key capabilities of the imaging system is also important, and some of them are:

- *transparency and reflections:* A LF camera provides information about depth dependence and Lambertian lighting. The depth dependence implies multiple depth of semitransparent objects and the Lambertian surface reflects light with equal intensity in all directions [28]. The depth dependence information can be exploited during coding, and the variation in depth of field information could give different compression levels at a same quality level. Reflections and transparency are prevalent in natural images, that is, reflected and transmitted lights are super-imposed on each other. The image can be modeled as a linear combination of transmitting layer, which contains the scene of interest, and a secondary layer, which contains the reflection or transparency [220] [243]. The decomposition of the images into two layers is an ill-posed problem in the absence of additional information about the scene [134]. The LF camera recorded information, particularly multiple views of a single scene, can be exploited to solve the problem. Therefore, in a test dataset images with transparency and reflections are needed.

- *Depth distribution:* Depth distribution is one of the most important information for many LF image applications such as refocused views and 3D views. Thus, inclusion of the scene with a wide depth of field variation is crucial.

- *Parallax:* The parallax is the one of the most exciting feature provided by the LF imaging. Therefore, the inclusion of LF image with rich parallax is advantageous.

- *Occlusion:* Handling occlusion related problems is always a most difficult and complex issue in the image processing field. Due to the advancement in imaging technology, in particular with LF imaging, this problem is becoming easier. The applicability of the dataset can be increased by including the scenes with occluded objects.

Together with the low level attributes, the high level features or content category, such as buildings, grids, mirrors and transparency, landscapes, nature, ISO and color charts, people, urban, and light, are considered during the content selection.

Generic procedure for source sequence selection

As already presented in Section 3.2.1, the steps included in scene selection procedure are:

- the image quality attributes are computed on LF image;

- images with high value of the attributes are selected: the result is a primary list of possible SRCs;

- from the primary list, SRCs are selected keeping in the mind that the final list of SRCs must cover all the considered content category or high level features. Moreover, the high preference is given to SRCs with the high occluded area and wide depth of field (appropriate for parallax and refocused views).

Followed by the procedures, the selected SRCs are verified subjectively based on the selected metrics and high level features.

10.2.2   Hypothetical reference circuit

To create Processed LF Image Sequences (PISs) from SRCs, HRCs are defined. For HRCs, it is worthwhile to recall the distortion model presented in Section 7.3.1; where a basic distortion model for LF image is presented (Equation 7.1).
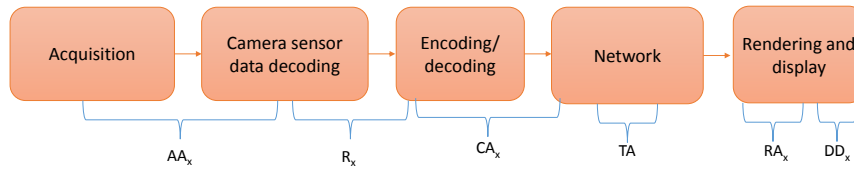

Figure 10.1: Generic HRC model in LF image communication.

Briefly, the distortion introduced by each step of the LF image communication chain is depending on the selected techniques or algorithms. An overview of the distortion stages is shown in Figure 10.1. By selecting different methods or algorithms all the stages can be varied for different HRCs. In Figure 10.1, AA_x is the acquisition artifacts. The R_x is the artifact introduced in representation phase. CA_x is used to represent the encoding and decoding artifacts. The artifacts introduced by the noisy and the bandwidth limited transmission network are represented by TA. Whereas, RA_x is the distortion produced by the rendering methods. The display devices also poses its own artifacts DD_x, to the based on its properties.

For traditional 2D display devices: normal monitor and normal TV, the LF image can visualize in different ways, such as refocused views, all-in-focused views or extended depth of field, and pseudo-video by exploiting viewing trajectory. Each of the displaying methods could have different user QoE even for the same content. For examples, in refocused view(s), out of focus area of the scene will be blurred, and level and quality blurriness is depending on the used refocusing algorithms. In all-in-focused view, only a few portions of the recorded LF information can be seen by the observer, as a result complete information of the light will be lost. The pseudo-video can be created by using sup-aperture views using the viewing trajectory, however, due to the lack of temporal information there could be the viewing trajectory distortion. Next, the many refocused views can also be used to create the pseudo-video, in that case handling inter-perspective aliasing ("jumps" between the views), as a result of large depth variation, is challenging together with the artifacts introduced by the selected refocusing algorithms.

10.2.3 Subjective quality evaluation protocol

In literature, many methods are recommended for multimedia services such as 2D/3D image and videos, and some of them are summarized in Section 3.2.3. In the context of LF imaging, during the selection of subjective quality assessment methods, the following issues need to be taken into account.

- Subjects are not familiar with this type of content (LF images) and the selected visualization techniques.

- LF camera recorded information decoding process results the artifacts (such as color distortions, blurring, Gaussian noise, and Salt and Pepper noise), and the quality of LF image is reduced, and ultimately, the low resolution of the image makes it worse.

Based on the issues, to reduce the impact of the pre-introduced artifacts, the double stimulus paradigm is an appropriate choice for the subjective quality assessment. Based on the small size of the pictures (views), experiment time, and poor quality of reference and test LF image, some of the possible candidate methods are: DSIS, PC, and DSCQS.

10.3 SMART light field image quality dataset

In this section, the steps followed to create the LF image quality dataset are detailed.

10.3.1 Source sequences

In this work, during the SRCs selection, together with the generic visual signal features, LF camera specific capabilities: transparency, reflections, and wide Depth of Field (DoF) are also considered. Next, as specified in ISO 20462 standard [121], to get relative quality values in JNDs, the selected attributes should appear in at least three images. A single image can cover more than one attribute; in this work two to three significant attributes per image have been considered. As a consequence, we have captured 16 images. Thumbnails of the selected SRCs are shown in Figure 10.2 and the corresponding key features are summarized in Table 10.2. As can be noticed, the considered images cover a large number of quality attributes and content variations. The analysis of the SRCs in Figure 10.3, shows that they cover a wide range of key quality attributes [176].

10.3.2 Hypothetical reference circuits

Figure 10.4 shows that before the test image showing to the subject, the LF image goes through encoding and rendering steps. As presented in Figure 10.1 there are many factors influencing the QoE of LF image. The test LF images can be created by varying the artifacts produced in any sections such as acquisition, raw sensor data decoding, encoding, network, and rendering. Among

(a) Book   (b) Building   (c) Car   (d) Chair

(e) Flower   (f) Grid   (g) Ladder   (h) Person

(i) Pillars   (j) River   (k) Sky   (l) Stone

(m) Table   (n) Tile   (o) Whitesky   (p) Window
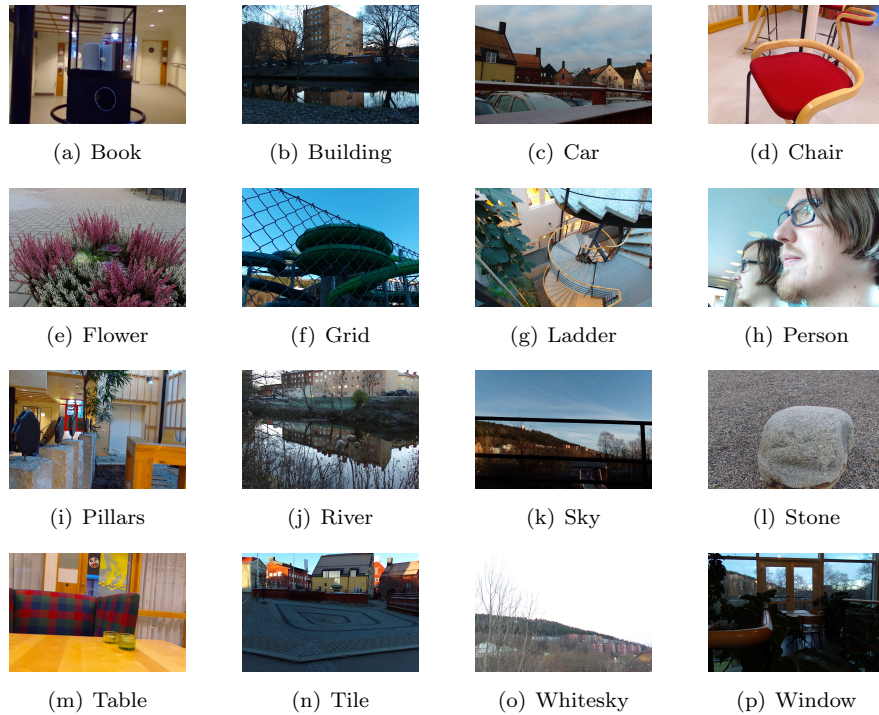
Figure 10.2: Thumbnail of source sequences in SMART dataset.

them, artifacts produced by the encoding methods is the scope of this section, and thus, the rest of the parameters including rendering method are keeping fixed, as shown in Figure 10.5. To the best of our knowledge, there is no standard encoding and rendering method for LF image. As presented in Section 7.2.3, many ongoing efforts have been given to devise the LF image compression techniques. However, in this work, the encoding techniques: JPEG, JPEG2000, HEVC intra [30] [5] and an ad hoc designed plenoptic image compression system (Sparse Set and Disparity Coding-SSDC) [139] are considered.

As mentioned before, four to five levels of distortion are sufficient to create the PISs. The distortion strengths have been selected after initial tests: a large set of test images has been generated and a subset of these images that spanned a wide range of visual quality scores (minimum to maximum quality) have been selected for the dataset, as detailed in [205]. The Quantization Parameters (QPs) used in literature for comparing performance of H.265/HEVC, VP9, and H. 264/AVC encoders are 22, 27, 32, and 37 [76]. From a preliminary test, the difference in perceptual quality of rendered images with original image was very small for HEVC intra and SSDC for QPs 22 and 27. Therefore, higher values of QPs (42 and 47) have been used. The selected compression levels for each compression method are shown in Table 10.3.

Index	Name	Description	Key Features	Remarks
(a)	Book	Book inside a transparent box	Homogeneity, Transparency	Indoor
(b)	Building	Building and its reflection on the river	SI, contrast, reflection	Outdoor
(c)	Car	Car roof and building with sky	Homogeneity, DoF	Outdoor
(d)	Chair	Chair on the floor	Colorfulness, DoF	Indoor
(e)	Flower	Flower with tile on the floor	SI, hue	Outdoor
(f)	Grid	Grid with natural scenes	DoF, hue	Outdoor
(g)	Ladder	Ladder top view	DoF, SI	Outdoor
(h)	Person	Close-up picture of a person with reflection	Reflection, contrast	Indoor
(i)	Pillars	Pillars	Colorfulness, DoF	Outdoor
(j)	River	Flower and river with reflection of the building	Contrast, DoF	Outdoor
(k)	Sky	Sky with natural scenes	Homogeneous, correlation	Outdoor
(l)	Stone	Stone on the concrete ground	SI, contrast	Outdoor
(m)	Table	Table with sofa	Colorfulness, Correlation	Indoor
(n)	Tile	Tile with background building	brightness , hue	Outdoor
(o)	Whitesky	Natural scene with white sky	brightness, correlation	Outdoor
(p)	Window	Natural outdoor scene with indoor objects	Transparency, DoF	Outdoor/Indoor

Table 10.2: The brief description of SRCs. The SRCs cover a wide range of key image quality attributes and image content variations.

JPEG (Quality Level)	30	50	70	90
JPEG2000 (Compression Ratio)	25	50	100	200
HEVC intra (QPs)	32	37	42	47
SSDC (QPs)	32	37	42	47

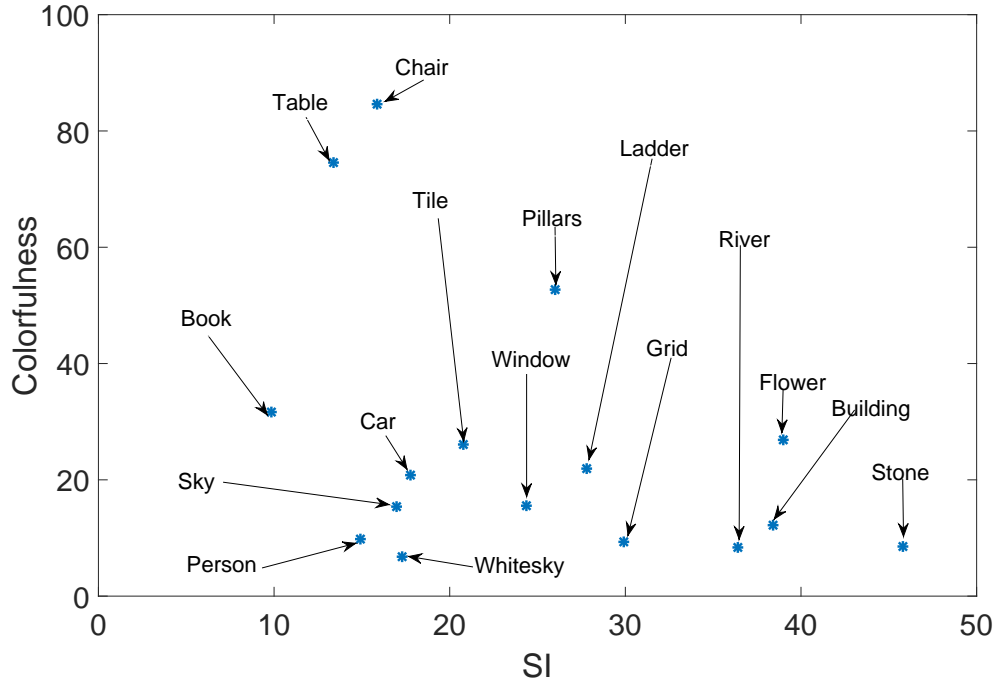Table 10.3: Compression methods and corresponding compression levels.

Figure 10.3: Spatial and colorfulness information distribution of SRCs.

10.3.3   Experimental setup for light field image processing

To create the test LF image sequences, the steps presented in Figure 10.6 were followed. In brief, LF Toolbox v0.4 [54] [50] has been used to convert the recorded raw sensor data to a LF data structure. After the decoding, demosaicing, devineting, and color and gamma correction are applied to the lenslet image. The color corrected lenslet image is changed to a 4D LF data structure. Then, the 4D LF image is arranged as a 2D image, by attaching the microlens image (like a lenslet image). Then the image is clipped to 8 bits by discarding the least significant bits, and converted into YCbCr 4:2:0 format for encoding. To create JPEG and JPEG2000 compressed images *MATLAB imwrite* function and for HEVC compression *HM software* (Encoder Version [11.0][Windows][VS 1700][64 bit]) [5] are used. For HEVC compression test parameter configurations have been selected as detailed in [30] and [66].

Among many possibilities (as mentioned in Section 7.3.2), in this work, the central view all-in-focused image was evaluated by the subjects. It was also possible to take other views all-in-focused, since, by just taking the central view, a lot of data may not be assessed. However, in principle, this means that the coding method may be good for just this view, but not for others. On the other hand, since the encoding is not aware of which view is being assessed, the selection of assessment is independent of the compression even if is does not give a complete assessment of the LF data. To
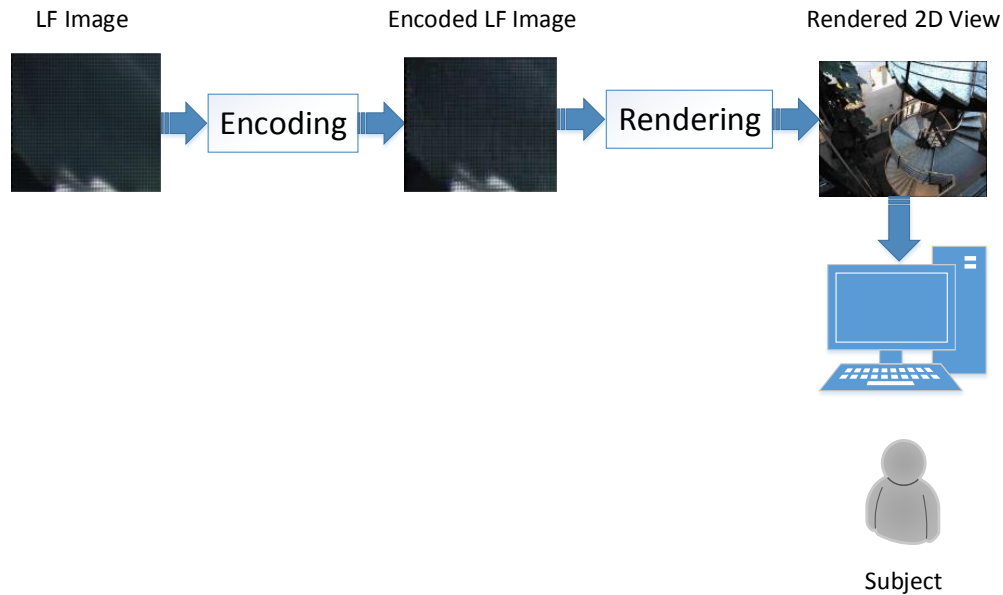
Figure 10.4: Signal processing steps and experimental design.

create the view, a basic or full resolution rendering method has been exploited, since considered LF data structure was the same as used for focused plenoptic camera in [71].

10.3.4 Subjective experiment design

In the following, the protocol adopted during the subjective experiment is described.

Subjective experiment methodology

To collect the opinion scores for the rendered 2D views, the PC, is selected. The main advantage of the PC is its high discriminatory power, which is of particular value when several test items are nearly equal in quality [112]. Briefly, a pair of images is presented to the observer who selects the one that has better image quality. The Graphical User Interface (GUI) used in this experiment is shown in Figure 10.7. At the end of each paired presentation, the subject expressed his/her preference by ticking the boxes. The number of pairs is $n \times (n - 1)$ for a n number of images. In our case, there

Figure 10.5: SMART LF image quality dataset: HRC model.

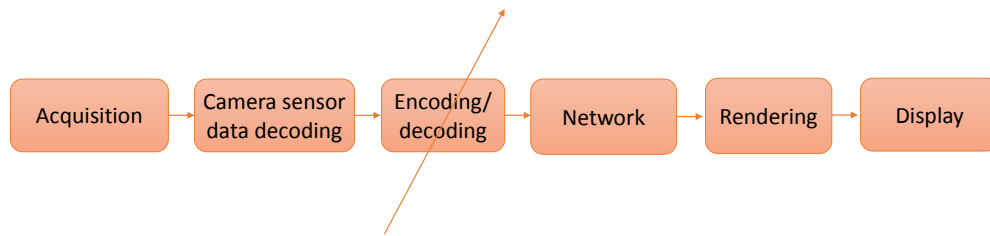were 17 test images (272 pairs) for each SRC, and we had 16 SRCs.

Subjects

To collect reliable results, 19 subjects have been selected [121] [109]. The subjects were drawn from a pool of undergraduate to post-doctorate students from Università degli Studi Roma TRE, Rome, Italy. The subjects were naive concerning the image impairments and the associated terminology. They were asked to wear any vision correcting devices (glasses or contact lens) that they normally wear.

Experiment length

To minimize the effect of viewers' fatigue on quality assessment, four experimental sessions have been scheduled. The experiment length is maintained shorter than 30 minutes by dividing each session into two sub-sessions to retain the attention of subjects [109]. Each sub-sessions lasted 12 to 15 minutes including evaluation and training time of 2 minutes, and at least 5 minutes gap between each sub-sessions.

Stimuli arrangement

In the experiment, PISs from 2 SRCs have been evaluated in each sub-session. The PISs from the same SRC are displayed at a time. To compensate the effect of a potential bias based on order or position of stimuli in the averaged results [126], stimuli are shown in random order for each subject according to their compression and intensity.

Training

Before the experiment, the subject is verbally given the instructions, followed by written instruction [112]. In the training stage, the subject is shown pairs of images having different levels of impairment, from the lowest to the highest found in the experiment. In this phase, each subject
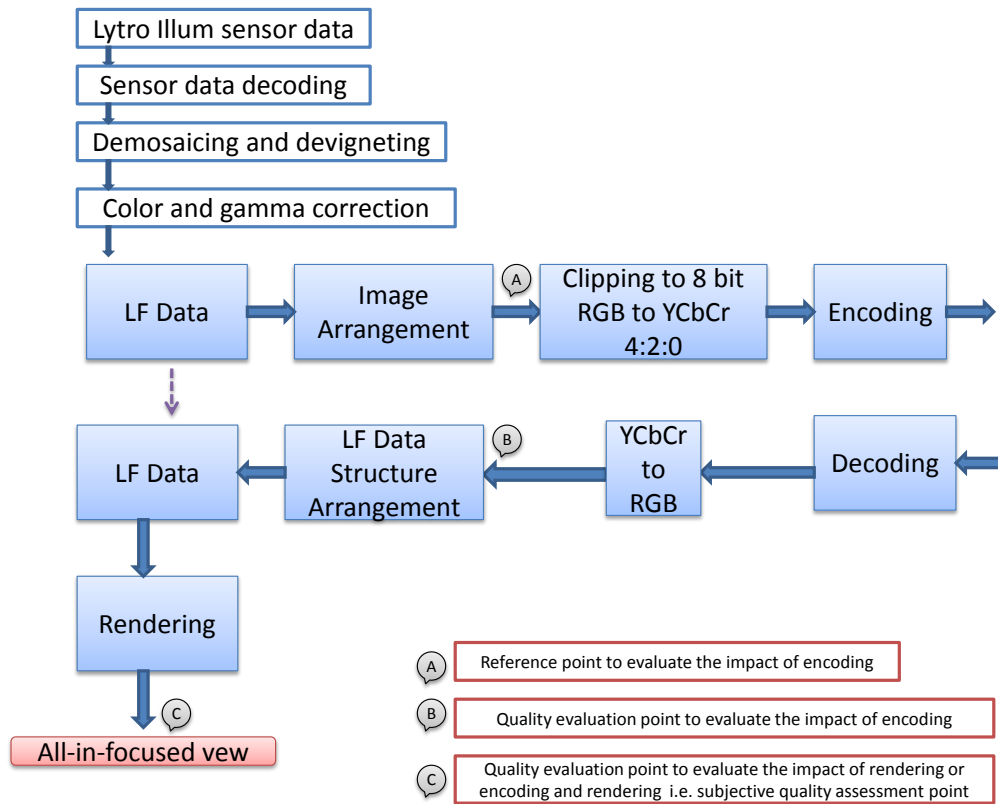
Figure 10.6: Experimental setup for LF image processing.

gets familiar with the assessment procedure and establish the annoyance values range. The images used in the training session are different from the test images.

Apparatus and environment

The experiment is conducted in a controlled environment in order to produce reliable and reproducible results by avoiding involuntary influence of external factors [109]. The characteristics of the display device and system are used in the experiment are briefly described in Table 10.4.

10.3.5   Collected quality scores for test light field images

As mentioned before, the result of the subjective experiment is the PCM matrix. To analyze the perceived quality, it is important to convert the results in a continuous rating scale [129]. To this aim, a Bradley-Terry (BT) [31] model has been considered [83]. The result of the PC experiment is expressed in terms of BT score, as a MOS.

Figure 10.7: Subjective experiment setup.

Table 10.4: System and display parameters.

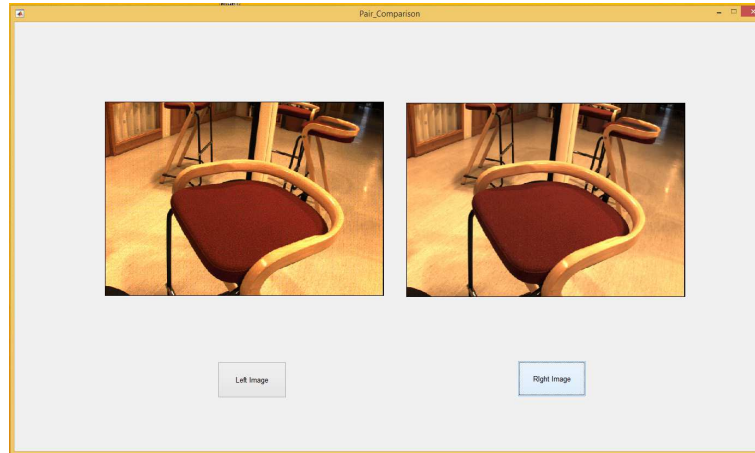| (a) System parameters.		(b) Display parameters.	
Parameters	Values	Parameters	Values
Processor	Intel(R)Core(TM)i7-4770	Display Device	DELL U2413f
Processor Speed	@3.40GHZ	Screen Refresh Rate	60Hz
RAM	8GB	Screen Resolution	1920x1200 pixels
System type	64-bit OS	Brightness and Contrast	50
Operating System	Windows 8.1	Sharpness	50
GUI	MATLAB R2015a	Aspect Ratio	Wide 16:10

The distribution of the BT scores for the test images is shown in the histogram in Figure 10.8. The histogram shows that the HRCs are well selected to create a set of data to cover a wide range of perceptual quality. The detailed analysis of the achieved results is presented in the next chapter.

10.4 IRCCyN/IVC&RM3/COMLAB light field image quality dataset

In SMART dataset, an all-in-focused view is used to evaluate the QoE of distorted LF image: an all-in-focused view is a very small portion of LF data, therefore the distortion created in complete LF data may not be evaluated by this visualization technique. Thus, this section is an attempt to evaluate the LF image QoE by considering more information during the subjective evaluation process. In particular, the different rendering methods (pseudo-video of the refocused views and sub-aperture images) are used to display the distorted LF image. Meanwhile, sources sequences are selected by considering the generic visual features and LF imaging capabilities, focusing to scenes with high parallax, occlusion, and wide depth of field. The adopted experimental setup is presented
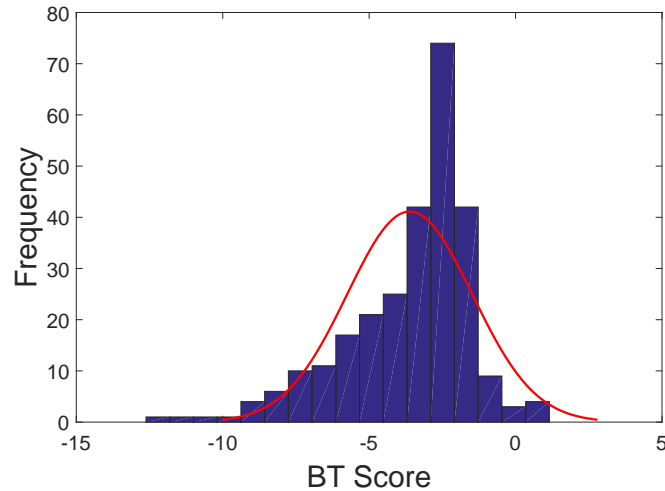
Figure 10.8: Distribution of BT scores. The test LF images have a wide range of perceived quality scores.

in the following.

10.4.1 Source Sequence

As mentioned before, SRCs are selected based on the key image quality attributes and LF camera specific capabilities. In brief,

- generic scene features: basic features such as CF, SI, and contrast are computed from thumbnail image of LF data; assuming most of the content or scene related information is available from thumbnail image. The thumbnails are extracted by using Lytro Desktop software;

- depth distribution: as a measure of depth distribution, 95% confidence interval of pixel values is computed from the depth map histogram. For depth map analysis, depth maps are estimated by using the Lytro Desktop software. In literature, many depth map estimation methods are available. However, due to the lack of the standard depth map estimation method, for a safe practice, depth map estimated by Lytro Desktop is used;

- parallax: to measure the parallax, the average absolute difference between (diagonal) extreme views (95% confidence interval in pixel values) is computed. Where, extreme views (view (4, 4), view (13, 13), view (13, 4), and view (4, 13)) is extracted by using the MATLAB LF Toolbox.

- occlusion: as a measure of occlusion in the scene, numbers of pixels on the occluded area boundary is counted. The adopted occluded area boundary estimation method is presented in

article [245]. In particular, to compute the pixels in occluded area, the .LFR file is decoded using the MATLAB LF Toolbox, and decoded 5D LF image is given as an input to the algorithm. As a result of the algorithm, the boundary of occluded area is computed, and pixels in the occluded area boundary are counted.

Together with the low level attributes, the high level features or content categories have been considered during the content selection. Followed by the procedures presented in Section 10.2.1, SRCs are selected; and verified subjectively based on the metrics and high level features; where higher preference is given to the images with higher occluded area, depth distribution, and parallax.

Thumbnail of selected SRCs is shown in Figure 10.12. As shown in Figure 10.9 selected SRCs covers a wide range of content features. Each SRC covers more than a feature, however the principal feature for each SRCs is indicated in Table 10.5. As well known, the selection of SRCs is also depends on the purpose of the study. In our study refocused and parallax views will be shown to the subjects, thus suitability of the selected SRCs for different applications (parallax and refocusing) is also indicated in Table 10.5.

10.4.2 LF image distortion model

As mentioned before in Figure 10.1, many possibilities have been available to create the test or distorted light filed images for subjective experiment. In fact, the artifacts (types and strengths) can be introduced at every step of the image communication chain. However, in this work, only two stages encoding and rendering is varied and other steps are kept constant, shown in Figure 10.10.

For encoding artifacts, the four image based encoding techniques are selected. As a baseline, standard image encoding technique JPEG is considered. The rest of three encoding methods are the HEVC based encoding, and these are: HEVC intra, Sparse SSDC, and bi-predicted self-similarity compensation (BPSS) [48]. Four levels of distortion are selected for each encoding method. The distortion strengths have been selected after the initial subjective experiment: a large set of test images is generated and a subset of these images that spanned a wide range of visual quality scores (bad to excellent quality) are selected for the dataset.

Then, the distorted LF images are rendered for the visualization. The selection of rendering methods is depending on the target displaying technique and device. In this experiment, two visualization techniques are selected.

- Refocused views: As a basic refocusing technique, a shift-sum approach is used to create the refocused views. The MATLAB toolbox is used to perform the refocusing by using *LFFilt-ShiftSum* function. This function shifts all the sub-aperture images according to a slope, which determines the focal plane. For Lytro Illum camera recorded content, the slope of -2 to 2 is selected. The resulted refocused (at the slope of -2 to 2) views are used as frames to create the pseudo video. The refocused views (slope -2 to 2 to -2 to 0) are displayed as a pseudo video.

SRCs	Name	Key features	Category	Applications
SRC1	Black_fence	Contrast	Building	Parallax
SRC2	Perforated_Metal_2	DD	Grid	Parallax
SRC3	Spear_Fence_1	DD	Grid	Parallax
SRC4	Friends_1	DD	People	Refocusing
SRC5	Zwahlen_&_Mayr	CF	Urban	Refocusing
SRC6	Parc_du_Luxembourg	CF	Nature	Parallax and Refocusing
SRC7	Palais_du_Luxembourg	Occlusion	Building	Parallax and Refocusing
SRC8	Fountain_&_Bench	DD	Mirrors and Transparency	Parallax and Refocusing
SRC9	Mirabelle_Prune_Tree	Parallax	Nature	Parallax
SRC10	Bikes	Parallax	Urban	Parallax and Refocusing
SRC11	Danger_de_Mort	DD	Grid	Parallax and Refocusing
SRC12	Bush	SI	Nature	Refocusing
SRC13	Slab_&_Lake	DD	Sky	Refocusing
SRC14	Sophie_&_Vincent_on_a_Bench	SI	People	Refocusing
SRC15	Game_Board	DD	Light	Parallax and Refocusing
SRC16	Sphynx	DD	Nature	Parallax and Refocusing
SRC17	Car_dashboard	DD	Reflection and Transparency	Parallax and Refocusing
SRC18	Rose	CF	Nature	Refocusing
SRC19	Duck	SI	Nature	Refocusing
SRC20	Pillers	CF	Pillars	Parallax and Refocusing
SRC21	ISO_Chart_22	SI	ISO Chart	
SRC22	ColorChart_2	CF	Color Chart	

Table 10.5: Selected source sequence description with covered key features. SRC18 and SRC19 are taken from Lytro Illum LF Dataset, SRC20 is taken from SMART dataset, and rest of SRCs are taken from LF Image Dataset.

(a) SI, CF, and Depth distribution (DD).

(b) SI, CF, and Contrast.

(c) SI, CF, and Parallax.

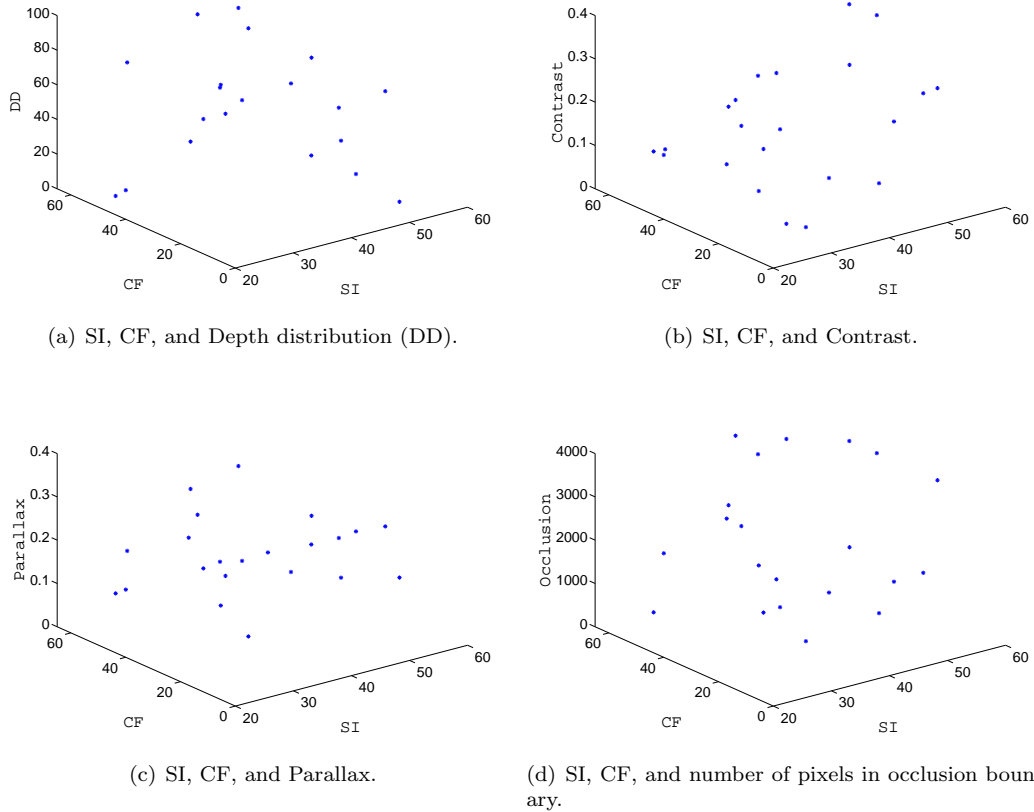(d) SI, CF, and number of pixels in occlusion boundary.

Figure 10.9: Basic features of selected source sequences. The selected SRCs are well distributed over a range for all the considered content characterization metrics.

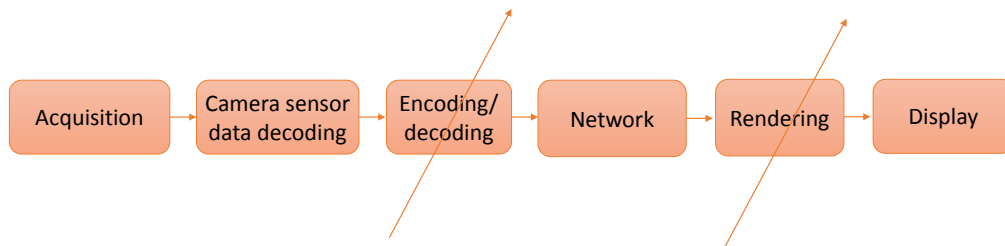Figure 10.10: Selected distortion model. To create the HRCs, the encoding and rendering steps are varied, and rest of the stages are kept fixed.

- Parallax: To parallax display, trajectory of view swap is exploited. In particular, 2D slices of the LF are displayed with a rudimentary parallax effect; the display motion is preset in a circular path of radius 60 degrees. For pseudo video, the views are selected as specified in

MATLAB toolbox function *LFDispVidCirc*.

10.4.3 Selected subjective quality assessment method

By considering the facts, presented in Section 10.2.3, DSIS method is selected as a subjective experiment protocol. In particular, the selection was based on the small size of the (view) pictures, experiment time, and poor quality of reference and test LF image.

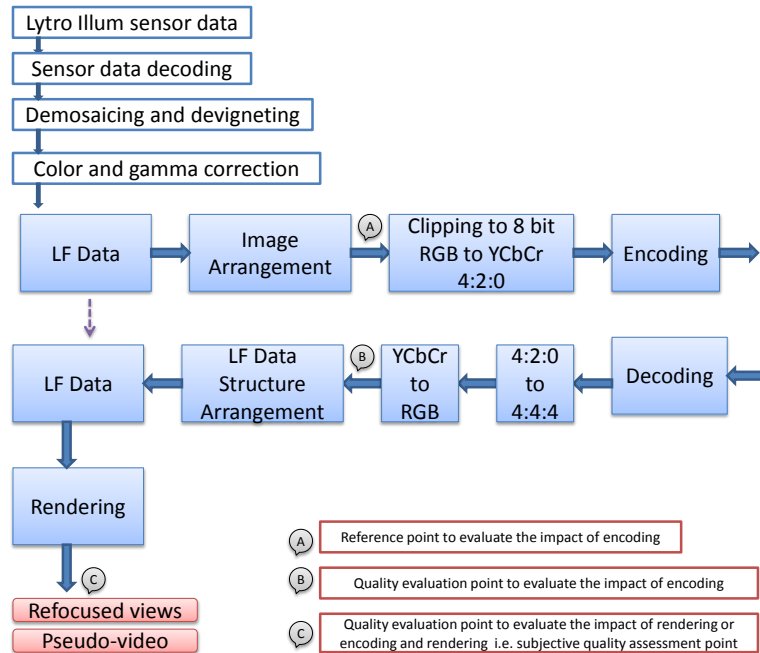### 10.4.4 Proposed LF image processing experimental setup

Figure 10.11: LF image processing experimental setup. (proposed subjective quality assessment framework for LF image)

The adopted LF image processing setup is shown in Figure 10.11. This is the proposed LF image processing experiment setup for subjective quality assessment of LF image. Therefore, every step of the framework is briefly explained below.

- *LF data:* The input to the quality assessment framework is 4D LF data. For maintaining the generality of LF, the LF data are expected to be in two plane representation $(L(u, v, s, t))$. If the data are taken from a multi camera array, some operation such as rectification and synchronization may be needed. On the other hand, for microlens array cameras such as Lytro Illum and Raytrix recorded data, some preprocessing steps may be needed. For an example,

for Lytro Illum camera captured LF, we may have to pass through the following steps: raw sensor data decoding, demosaicing, devigneting, color and gamma correction [49, 50].

- *Image arrangement:* The arrangement is depending on the selection of the encoder, particularly, the data input format of the encoder. For examples: i) for image based encoders: 4D LF data may be converted in to a 2D image structure. Because, most of the standard and recently proposed image encoding techniques (JPEG, JPEG, HEVC intra related methods) are designed for 2D image (structure). This step is more important for multi-camera captured LF data. Whereas, for microlens array (Lytro Illum) recorded data the lenslet image also can be given directly as an input for encoding; and ii) for video based encoders: the aperture views of the LF image can consider as frames, and pseudo video can be created by considering the relative position of the sub-aperture images. The final image, what we got in this stage, can be considered as a reference point to evaluate the impact of encoding (in Figure 10.11 it is indicated as point A). At the receiver side, 4D LF image may be need from the encoded LF data for the rendering. However, it depends on the selection of rendering methods.

- *Image format conversion:* This step is also based on the selected encoding method. For an example, in this chapter HEVC intra related encoders are used for encoding, and thus, 16 bit precision data is clipped to 8 bits by dropping the least significant bits, and RGB 8 bit uncompressed image is converted to YCbCr4:2:0 color space. The resulted LF image is encoded and decoded. After decoding, the decoded image is converted to YCbCr4:4:4 and back to RGB.

- *Encoding:* The above mentioned steps: image arrangement and image format conversion are particularly for encoding purposes, and these are based on the selected encoding technique. From another viewpoint, if our focus is not for encoding artifacts, these steps (image arrangement, image format conversion, and encoding) are not needed. From the reference LF data, we can directly go to the rendering step.

- *Rendering:* The test LF image needs to be converted into a suitable format for the visualization by using the appropriate rendering method. In particular, different visualization techniques such as refocused views, parallax, and 3D image demands different rendering methods.

10.4.5 Ongoing works

Due to the time taken by selecting, encoding methods to encode the LF image, the HRCs creation process is running. Once, HRCs is created the subjective experiments will be scheduled. As per our plan, after the subjective experiment, achieved initial results will be analyzed, and the collected opinion scores and distorted LF images will be used to benchmark the objective IQMs.

10.5 Conclusion

In this chapter, first a generic subjective quality assessment framework is discussed with highlighting source sequence selection methods, distortion model for creating test sequences, and subjective quality evaluation protocol.

Following this framework, a new LF image quality dataset has been created. For dataset population, based on the key image quality attributes and LF camera specific capabilities the SRCs are selected and captured by Lytro Illum Camera. The captured SRCs are processed and compressed. An experiment was scheduled to collect the subjective quality rating for the processed image sequences. Source sequences, test sequences, subjective quality scores, and adopted experimental setup procedures are made freely available for the research community.

Further studies are performed in the framework, as a result, a LF image processing experiment setup is proposed. The proposed setup is used to design the subjective experiment for LF images. By using the proposed setup, a new LF image quality dataset is going to be created by considering a wide range of source sequences and rendering methods.

(a) Black_fence

(b) Perforated_Metal_2

(c) Spear_Fence1

(d) Friends_1

(e) Zwahlen_&_Mayr

(f) Parc_du_Luxembourg

(g) Palais_du_Luxembourg

(h) Fountain_&_Bench

(i) Mirabelle_Prune_Tree

(j) Bikes

(k) Danger_de_Mort

(l) Bush

(m) Slab_&_Lake

(n) Sophie_&_Vincent_on_a_Bench

(o) Game_Board

(p) Sphynx

(q) Car_dashboard

(r) Rose

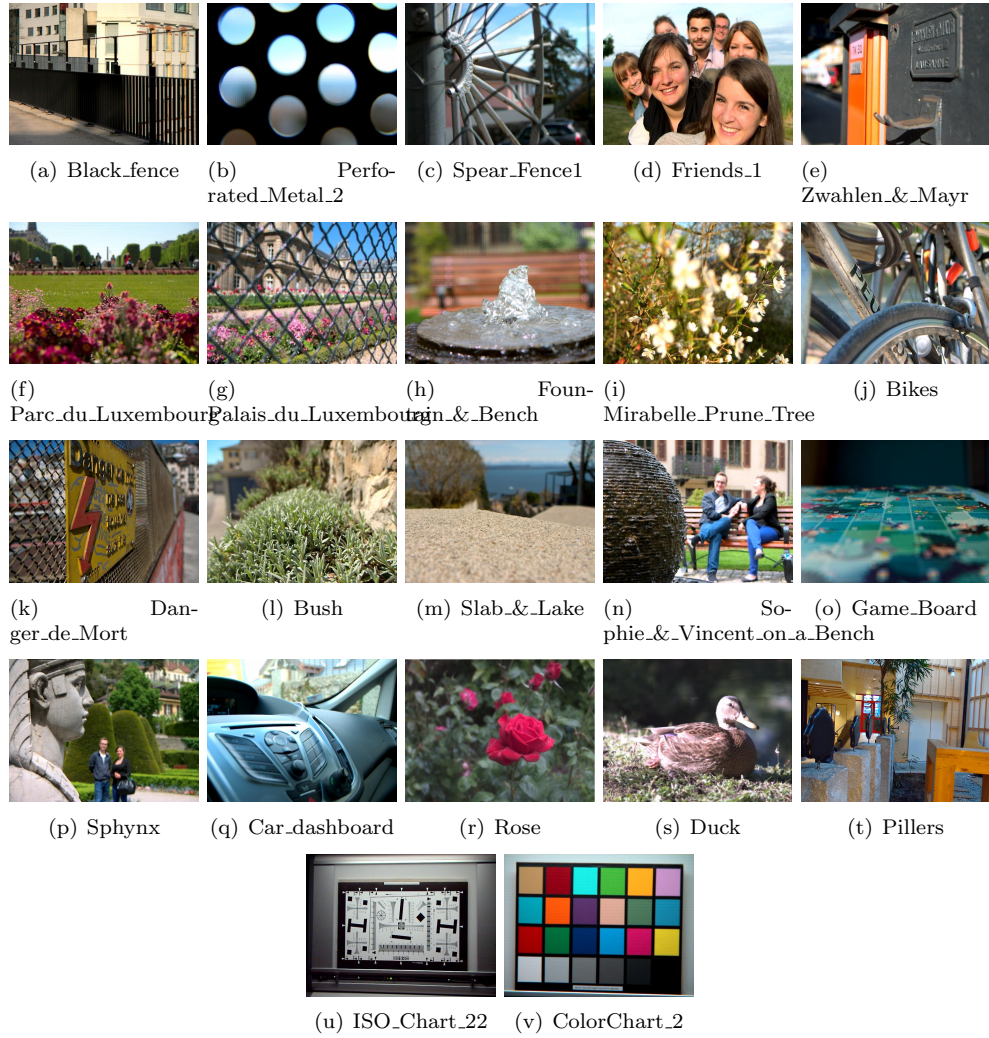(s) Duck

(t) Pillers

(u) ISO_Chart_22

(v) ColorChart_2

Figure 10.12: Thumbnail of the selected source sequences..

Chapter 11

Perceptual quality of compressed light field images

This chapter presents the study of the perceptual quality of the encoded LF image by using the dataset presented in Section 10.3. In particular, the major contributions are:

- study on the impact of encoding artifacts on LF image QoE;

- performance evaluation of the encoding methods based on the human perceptual quality of compressed LF image;

- study on the influence of scene or image content on visual QoE;

- benchmarking of the 2D IQMs, when applied to LF image.

11.1 Introduction

Evaluation of LF image QoE, as well as testing of new processing tools, or even assessing the effectiveness of objective quality metrics, relies on the availability of test dataset and corresponding quality ratings. For this purpose, the SMART LF image quality dataset presented in Section 10.3 is used. The dataset provides test LF images and corresponding subjective opinion scores.

As mentioned in Section 7.2.3, many ongoing efforts have been given to develop LF image compression methods. The performance analysis of these methods is important. To this aim, a study on the impact of compression artifacts on perceived quality is necessary. Moreover, performance evaluation of the encoding methods based on the QoE of encoded image is of crucial importance.

As well known, QoE of images and the performance of compression methods is significantly influenced by the image content characteristics and HVS [175]. Thus, the study of the influence of the image content with respect to HVS on the perceived quality of LF image is important.

Index	Artifacts	Index	Artifacts
1	Reference		
2	JPEG (q=30)	10	HEVC intra (QP=32)
3	JPEG (q=50)	11	HEVC intra(QP=37)
4	JPEG (q=70)	12	HEVC intra (QP=42)
5	JPEG (q=90)	13	HEVC intra (QP=47)
6	JPEG2000 (CR=25)	14	SSDC (QP=32)
7	JPEG2000 (CR=50)	15	SSDC (QP=37)
8	JPEG2000 (CR=100)	16	SSDC (QP=42)
9	JPEG2000 (CR=200)	17	SSDC (QP=47)

Table 11.1: Indexes and there corresponding artifacts (q=quality level, CR=compression ratio, QP=quantization parameter level).

As for conventional image and video processing, the selection of an objective quality metric is necessary to benchmark the processing algorithms such as encoding methods. In literature, to benchmark LF image encoding methods a well known full reference metric, PSNR, is used. Because, there are no quality metrics specifically designed for LF image. In literature, many IQMs have been proposed, and some of them are very common in the image processing field. In this situation, performance evaluation of existing 2D IQMs, when applied to LF image is necessary.

11.2 Compressed light field image quality of experience

To study the impact of compression methods on LF image, the analysis of the perceived quality of the compressed LF image is presented. For the analysis, the numbers 1 to 17 have been used as indexes to represent 17 HRCs. The index numbers and corresponding HRC description are shown in Table 11.1. The impact of compression on the perceived quality is analyzed in two steps. First, the overall impact of the HRCs on perceived quality is presented and similar analysis is performed for the individual SRCs.

11.2.1 Overall impact of encoding

To analyze the perceived quality of LF images of different HRCs, the opinion scores given by all the subjects and for all the SRCs are used. The results reported in Figure 11.1, show that the perceived quality of processed LF images varies with the selected HRCs. This result is further confirmed by NAOVA: the result $F(16, 255) = 18.1$ and p-value $\simeq 0 < 0.05$, indicates that the quality scores for the HRCs are significantly different.

11.2.2 Impact of encoding on individual source sequences

To study the influence of HRCs for individual SRCs, the result is further analyzed with the help of box plot. From Figure 11.2 it can be noticed that some of the points are outliers and that the range
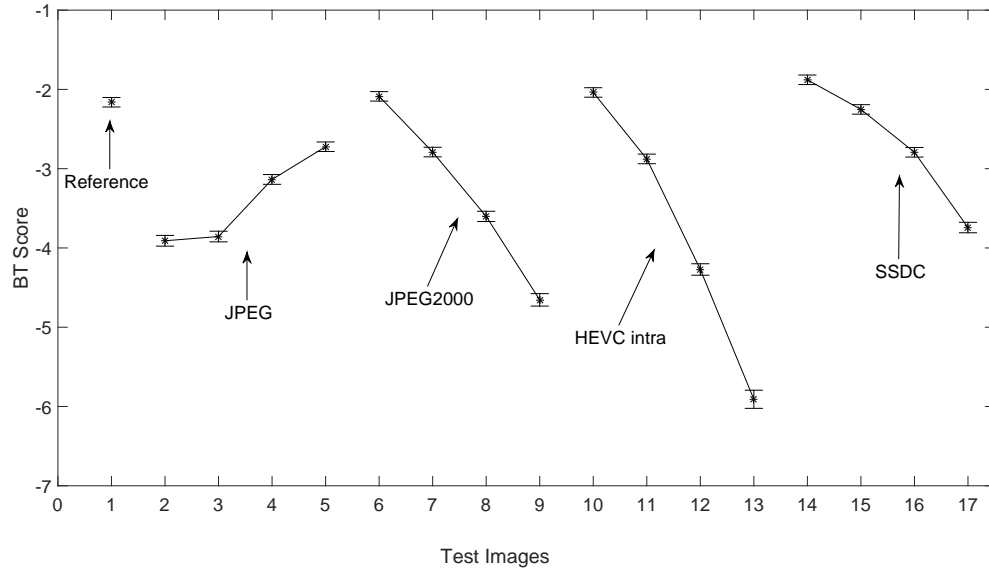
Figure 11.1: Overall impact of encoding on LF QoE: BT score with 95% of Confidence Interval (CI).

of the first and third quartile is noticeable. This result indicates that the BT scores are different for the SRCs even for the same level of compression. Therefore, it is worthwhile to analyze the perceptual quality of individual scene at different compression levels.

The perceived quality of the SRCs for different HRCs is shown in Figures 11.10, 11.11, 11.12, and 11.13. As can be noticed, the patterns of quality scores are noticeably different for the SRCs even at the same level of compression. As an example, variation in the BT scores is smaller for *Building* compared to *Car* for the same HRCs or at a same level of compression.

To confirm the results, Barnards test [22] has been considered and the test is used to check whether the probability, P_{ij} (scores given for a test sequence) is significantly different from a probability of 0.5 (i.e., whether the observers are undecided) or not [137]. In plots for exact test, Figures 11.10, 11.11, 11.12, and 11.13, the difference of the perceived quality of test images is expressed for each possible pair. It is assumed that each pair of images follows the commutative law. Therefore, our region of interest is the left side of the matrix plot. In the plot, if the opinion scores for the pairs of images are not significantly different, the corresponding square is filled with white box otherwise it is black, whereas gray box is used to express the test is not necessary. The Figures 11.10, 11.11, 11.12, and 11.13 show that a different set of SRCs has different numbers of white squares. This result indicates the perceptual quality of LF images is also influenced by the SRC or image content.

Moreover, a one-way-ANOVA has been performed for a group of SRCs at different HRCs to confirm the result. The result, $F(15, 256) = 3.71$ and p-value $\simeq 0 < 0.05$, indicates that the
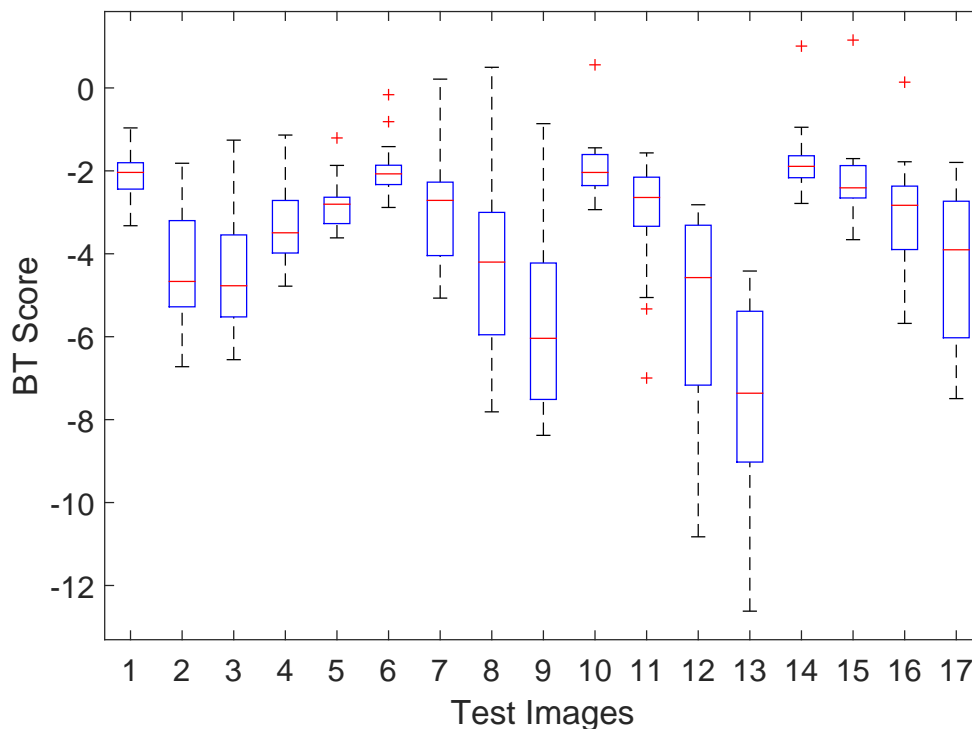
Figure 11.2: Box plot of the BT scores at different level of HRCs for 16 SRCs. On each box, the central mark is the median score, the edges of the box are the 25th and 75th percentiles, and the outliers are plotted individually.

perceived quality of the test images are significantly different for the SRCs. This is because the image content has a significant impact on the QoE [175].

11.3 Effect of image content on perceived light field image quality

To study the effect of scene content on QoE of LF image, the performed analysis results are reported in Figure 11.3. It shows that the perceptual quality is different for SRCs. This result strengthens the need to investigate the impact of image content on the perceived quality.

At the moment, no standard content descriptors have been defined. In this work, key image quality attributes defined in Section 3.2.1, have been used. Based on the descriptors scores and corresponding image perceived quality scores, the correlation analysis is performed.

For the analysis, PLCC and SRCC have been used. The correlation between the content descriptors scores and corresponding perceived quality scores is shown in Table 11.2, and it shows that
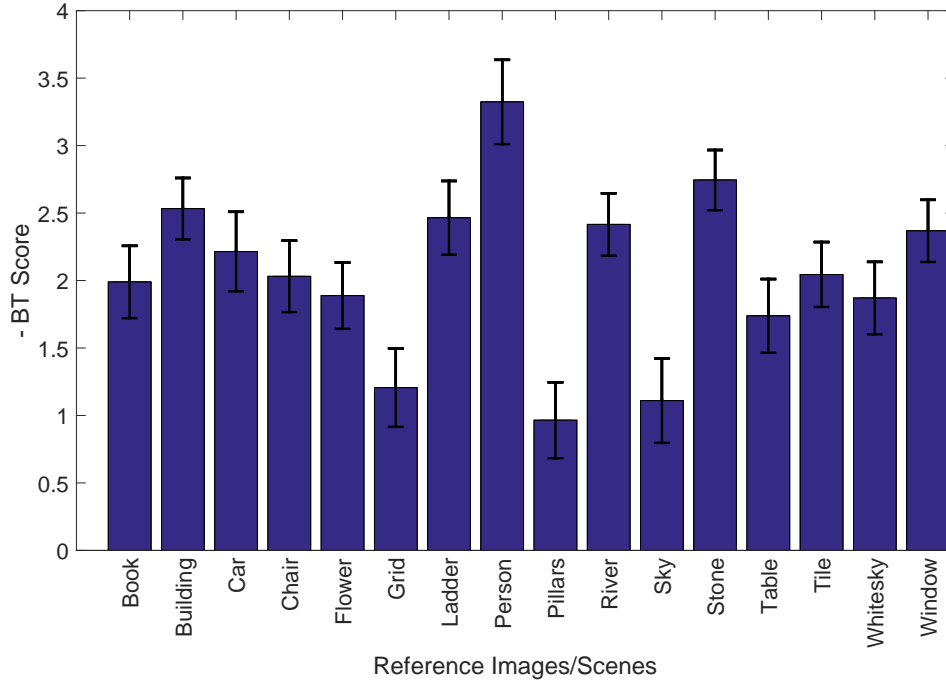
Figure 11.3: BT score with 95% CI for reference LF images. The perceived quality is different for reference images.

there is no significant correlation between the considered descriptor with the perceived quality of the LF image.

The result is further confirmed with the help of PCA and the result, Figure 11.4, shows that the descriptor saturation has a small angle with BT score compared to other descriptors. However, the angle between the BT score and saturation is noticeable.

The results show that there is no strong correlation between the SRC descriptors and perceived quality. Though, the perceived quality of the SRCs is significantly different. This result could be due to the fact that together with system factors (such as transmission impairments and compression

Descriptors	SI	CF	Contrast	Homogeneity
PLCC	-0.184	0.317	-0.344	0.347
SRCC	-0.267	0.352	-0.423	0.452
Descriptors	Hue	Saturation	ColorValue	Brightness
PLCC	-0.047	**0.520**	0.018	-0.150
SRCC	-0.229	0.394	0.191	-0.050

Table 11.2: Correlation coefficient: correlation between scene/content descriptors and LF image QoE.
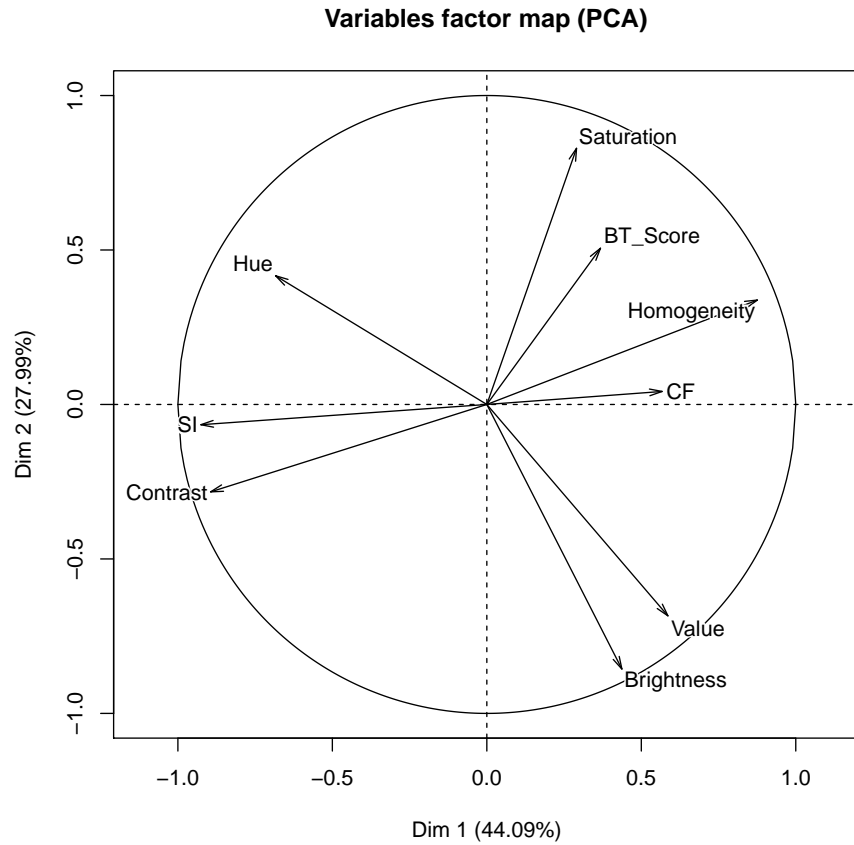
Figure 11.4: Result of PCA. The high angles between the content descriptors and BT score indicate there is no strong correlation between the perceived quality and the attributes.

artifacts) the perceived quality is also influenced by human and context factors [131]. However, the considered descriptors mostly explain the system factors and the perceived quality is more about the subjects, and hence physiological, psychological, social, and role-related aspect of the subject also influence the quality score [234].
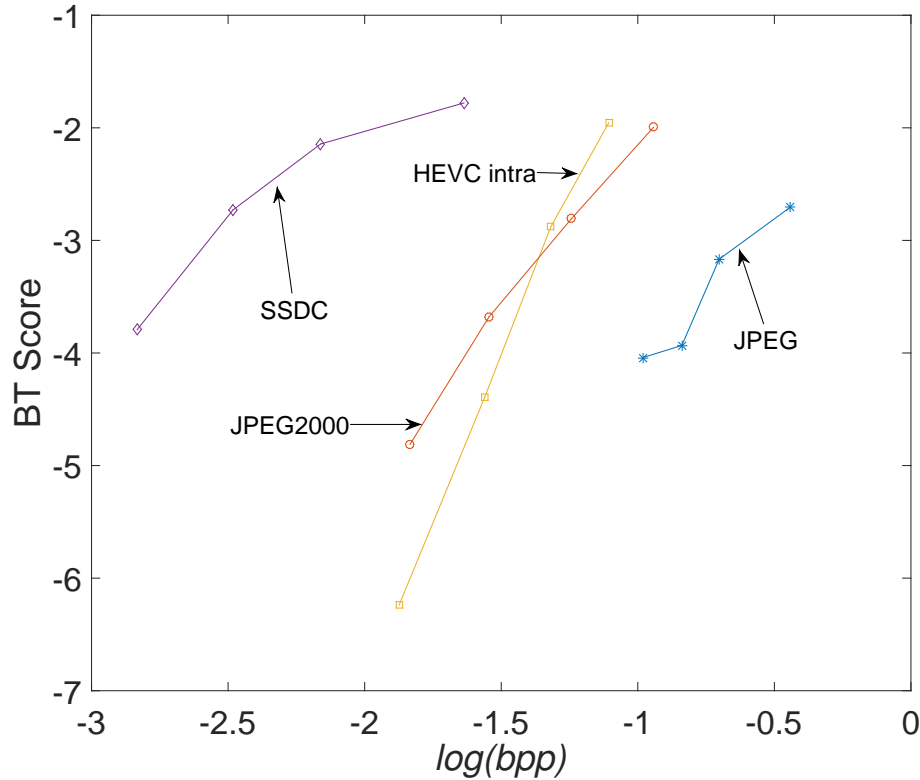
Figure 11.5: Performance of the considered compression methods: perceived quality of LF images at different level of compression by the different compression methods. The recently proposed LF image compression method, SSDC, has a significantly high level of perceived quality compared to other methods.

11.4 Perceived quality analysis of the compression methods for light field image

From Figure 11.1, it can be noticed that the perceptual quality of SSDC compressed LF image is high compared to other compression methods, and the same result is replicated in Figures 11.10, 11.11, 11.12, and 11.13 for different SRCs. In those plots the test points for the compression methods are different: quality level for JPEG, compression ratio for JPEG2000, and QPs for HEVC intra and SSDC. To compare the performance the bit per pixel (bpp) for each test image is computed and compared with the corresponding subjective quality scores.

From Figure 11.5, we can observe that the perceptual quality of JPEG2000 is significantly higher compared to JPEG at the same level of bpp and a similar trend is shown for SSDC and HEVC intra. These results indicate that the SSDC outperforms the other methods for the LF images. A detailed

(a) JPEG2000 (-0.941)   (b) HEVC intra (-1.052)   (c) JPEG2000 (-1.834)
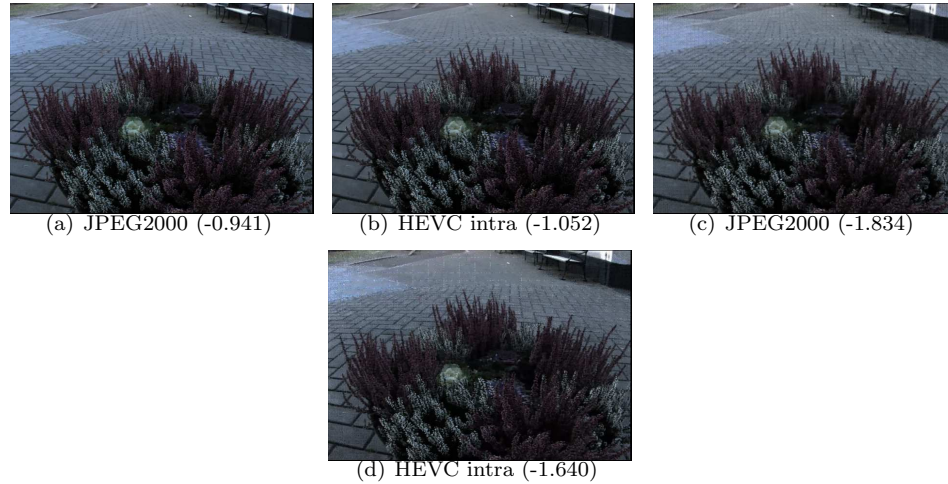
(d) HEVC intra (-1.640)

Figure 11.6: Perceptual quality of LF images at low and high level of compression for JPEG2000 and HEVC intra, and level of compression is expressed in terms of *log(bpp)*.

description of SSDC is available in [139], which also reports on the objective performance of this compression method on LF images from focused plenoptic cameras such as Raytrix. The results of our work are different in two ways: 1) comparison of the compression methods is done on LF images from a LF camera, Lytro Illum, and 2) the perceptual quality of the compressed image has been evaluated by using a subjective assessment, and results show that the SSDC performs better than other considered methods.

Moreover, it is noticed that there is a crossover between the curves for JPEG2000 and HEVC intra. This crossover indicates that at a high level of compression, JPEG2000 produced blurring and ringing artifacts, that have a small impact on the perceptual LF image quality compared to the block artifact produced by the HEVC compression [171]. This result is further confirmed by Figure 11.6. Figure 11.6 (a) and (b) show that at low level of compression the HEVC intra compressed LF image has a high perceptual quality even the level of compression is higher than for JPEG2000 but at a high level of compression (Figure 11.6 (c) and (d)) the JPEG2000 compressed image has a high perceptual quality. The most important point that we need to notice that, as mentioned before, the employed values of QP (42 and 47) are significantly higher than the commonly used QPs (22, 27, 32, and 37) [76].

Influence of image content: The result shown in Figure 11.7 indicates that the even at a same level of QPs, the SRCs have a different level of compression. Moreover, the influence of the image content in the perceptual quality of LF image is analyzed with the help of content descriptors for top performing LF image compression methods, HEVC intra and SSDC. The result, Table 11.3, shows that the level of compression significantly depend on the image color value and the brightness of the LF image. Therefore, inclusion of a wide range of content variations is important in the dataset.
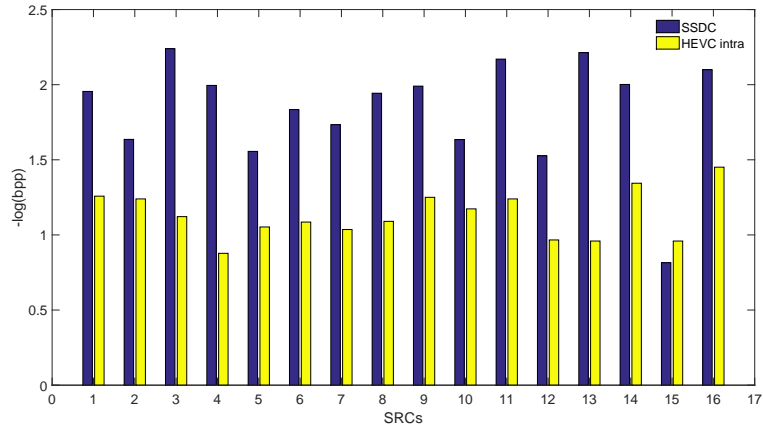
Figure 11.7: The compression level of SRCs at QP = 32 for top performing compression methods, SSDC and HEVC intra.

Compressions	SI	CF	Contrast	Homogeneity
SSDC	0.013	-0.302	0.036	-0.046
HEVC intra	-0.008	0.399	0.089	-0.153
	Hue	Saturation	Value	Brightness
SSDC	-0.404	-0.536	0.451	0.617
HEVC intra	-0.290	-0.243	0.844	0.780

Table 11.3: PLCC: correlation coefficient between image attributes and image compression levels measured in terms of bpp (computed at QP = 32).

11.5   Performance analysis of image quality metrics

In this section, the performance of the 2D IQMs, when applied to LF image is analyzed. The considered 2D visual quality metrics are:

- Mean of Squared Errors (MSE) [69]

- Peak Signal to- Noise Ratio (PSNR) [69]

- Signal-to-Noise Ratio (SNR) [69]

- Weighted SNR (WSNR) [158]

- PSNR based on Human Visual System (PSNR-HVS) [57]

- PSNR-HVS-M [187]

- Visual Signal-to-Noise Ratio (VSNR) [37]
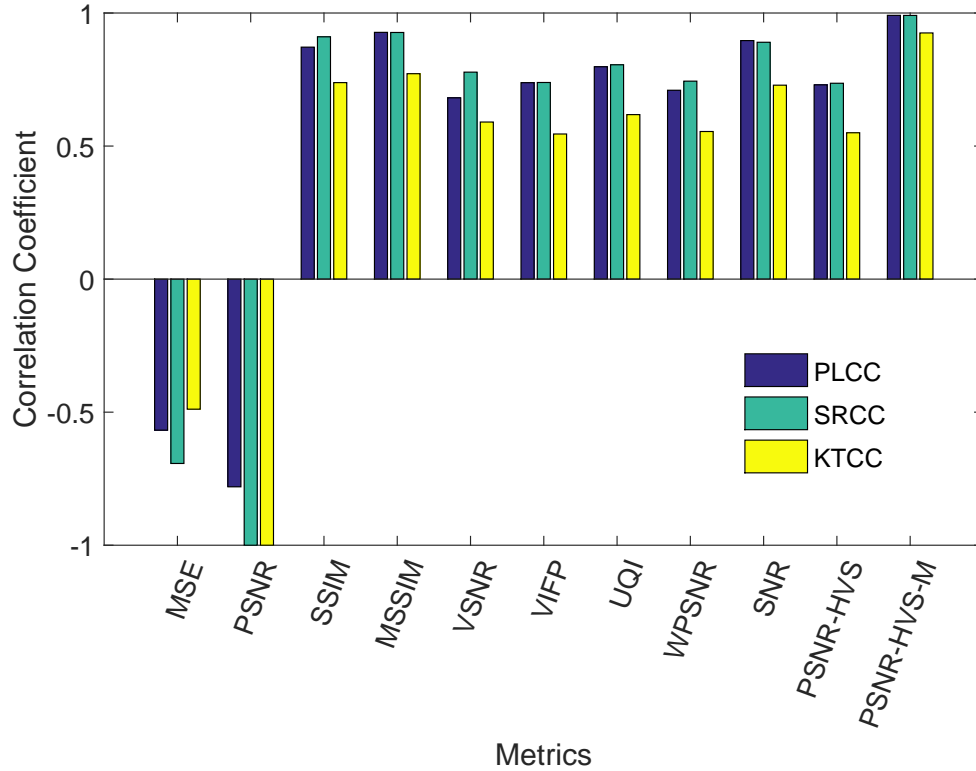
- Structural Similarity Index (SSIM) [250]

Figure 11.8: Performance of 2D IQM, when applied for Litro Illum LF image: correlation coefficient between the estimated quality scores and subjective opinion scores.

- Multi-scale Structural Similarity Index (MSSIM) [251]

- Image information and Visual quality (VIFP) [210]

- Universal image Quality Index (UQI) [248]

Correlation between the subjective opinion scores and objective quality scores has been computed by using PLCC, SRCC, and KTCC. Figure 11.8 illustrates that the objective IQMs perform differently, and the PSNR-HVS-M has higher values of PLCC and SRCC compared to other metrics. This result could be due to the fact that the PSNR-HVS-M takes into account the HVS and the contrast sensitivity of the image [187]. A similar result is yielded from PCA, as shown in Figure 11.9, the PSNR, PSNR-HVS, and PSNR-HVS-M have a small angle with perceptual quality score (BTScore) with respect to other metrics.
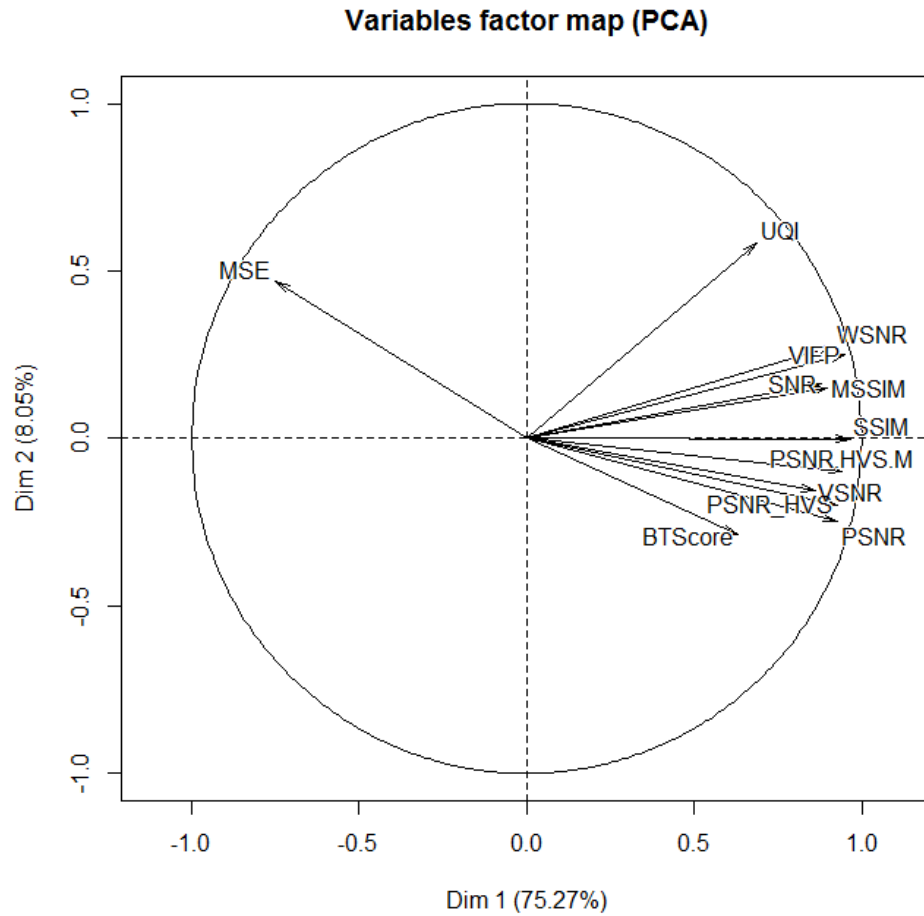
Figure 11.9: PCA results: performance analysis of 2D IQM, when applied for Litro Illum LF image.

11.6   Conclusion

In this chapter, a study on the perceived quality of the compressed LF image is presented. Major contributions of the article are described in the following.

- The impact of the compression artifacts on the perceived LF image quality is studied. The results show that the compression methods significantly degrade the perceived quality of LF image, and the level of the quality degradation varies for the images with different content.

- The impact of the image content on the perceived quality is studied with the help of key image quality attributes and subjective quality scores of the LF images. The results show that there was no strong correlation of the descriptors with the corresponding perceptual quality. However, the results presented in Section 11.4 show that the level of compression of LF images

varies for different content even at a same level of QPs. Therefore, the inclusion of SRCs with a wide range of content variation is important.

- The performances of compression methods are evaluated for LF images. The results show that the recently proposed plenoptic image compression method, SSDC [139], has a better perceived quality at a same level of compression compared to other considered compression methods. The results demonstrate the importance of specific compression algorithms for LF data in order to reach a best possible quality.

- The performances of widely used 2D IQMs are evaluated for the LF images. The results show that, among the considered metrics, the quality scores predicted by PSNR-HVS-M has a better correlation with corresponding subjective quality scores.
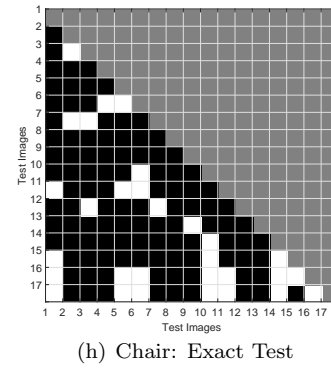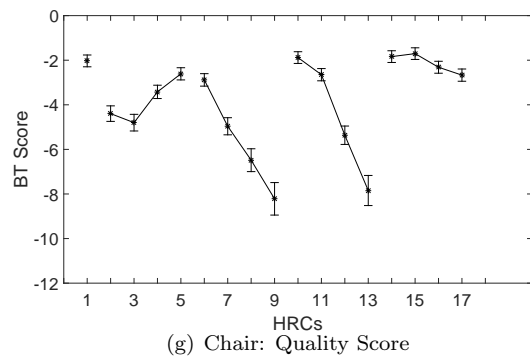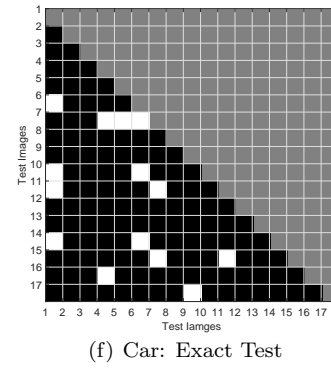
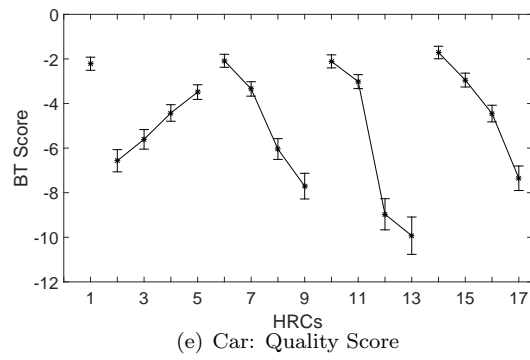(a) Book: Quality Score

(b) Book: Exact Test

(c) Building: Quality Score

(d) Building: Exact Test

(e) Car: Quality Score

(f) Car: Exact Test

(g) Chair: Quality Score

(h) Chair: Exact Test

Figure 11.10: BT Score with 95% confidence interval and result of exact test for four images: Book, Building, Car, and Chair.

(a) Flower: Quality Score

(b) Flower: Exact Test

(c) Grid: Quality Score

(d) Grid: Exact Test

(e) Ladder: Quality Score

(f) Ladder: Exact Test

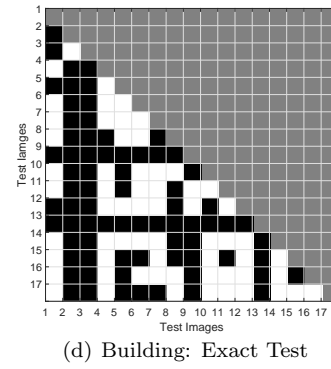(g) Person: Quality Score
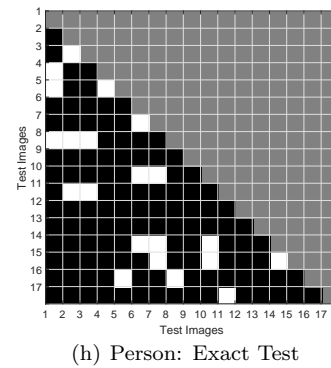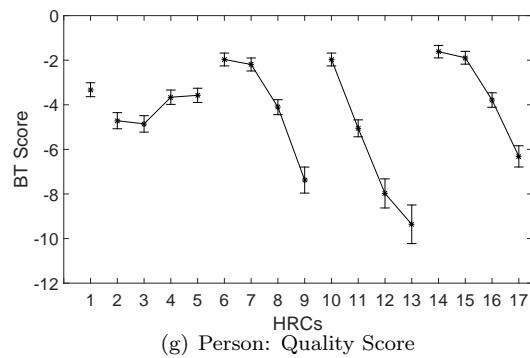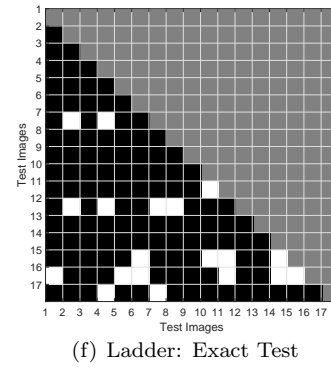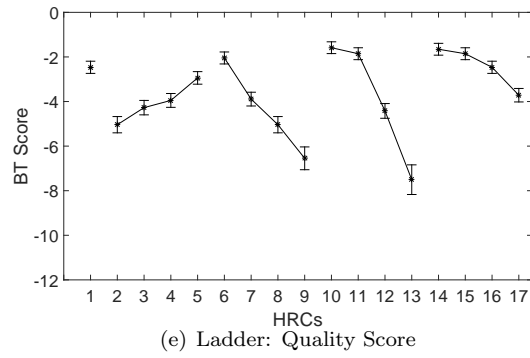
(h) Person: Exact Test

Figure 11.11: BT Score with 95% confidence interval and result of exact test for four images: Flower, Grid, Ladder, and Person.

(a) Pillars: Quality Score

(b) Pillars: Exact Test

(c) River: Quality Score

(d) River: Exact Test

(e) Sky: Quality Score
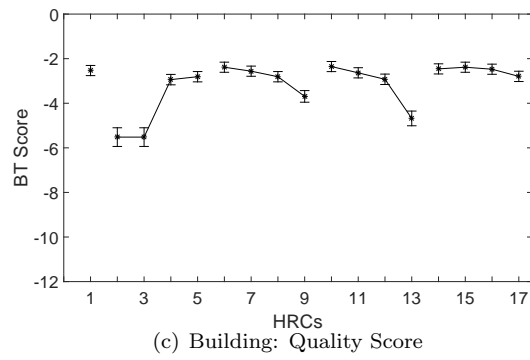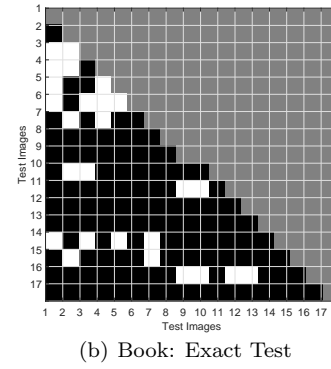

(f) Sky: Exact Test

(g) Stone: Quality Score
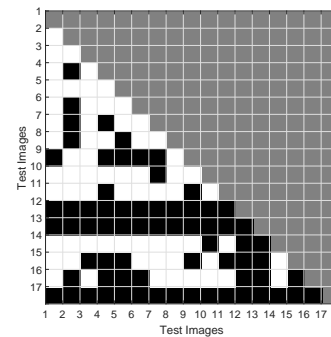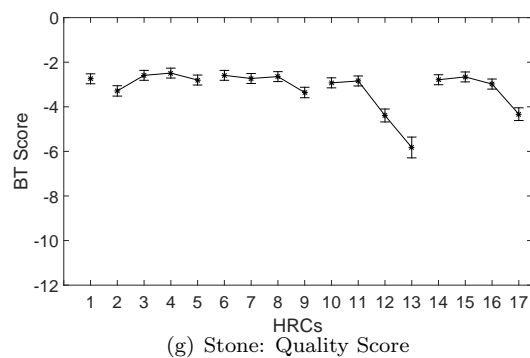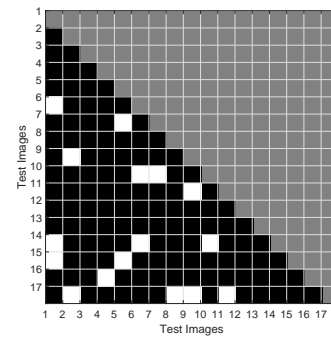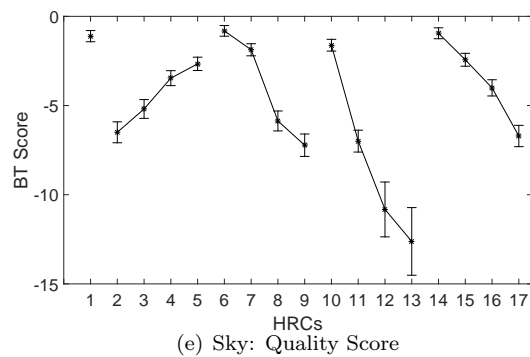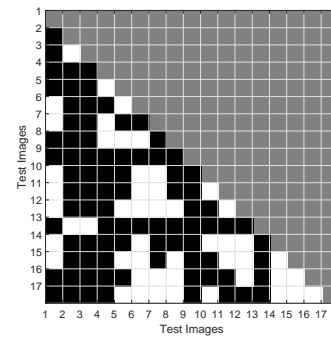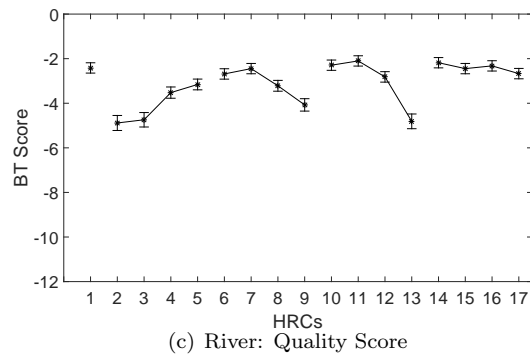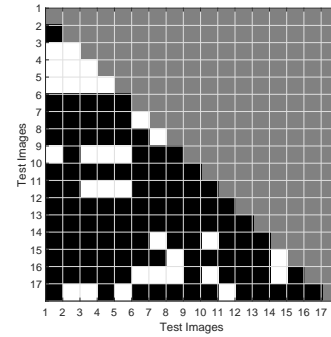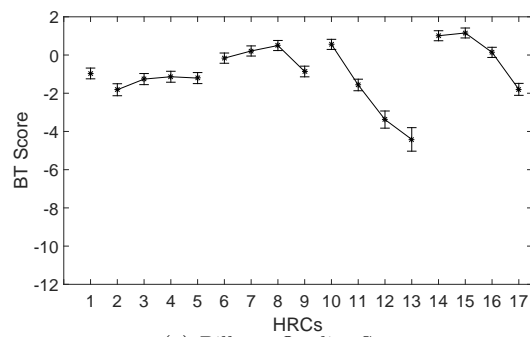
(h) Stone: Exact Test

Figure 11.12: BT Score with 95% confidence interval and result of exact test for four images: Pillars, River, Sky, and Stone.
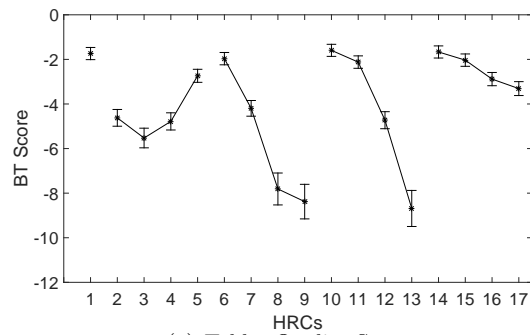
(a) Table: Quality Score
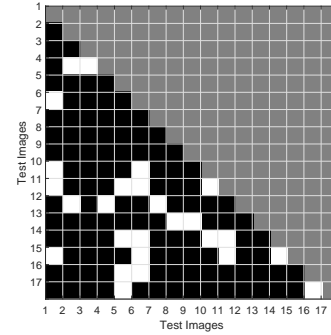
(b) Table: Exact Test
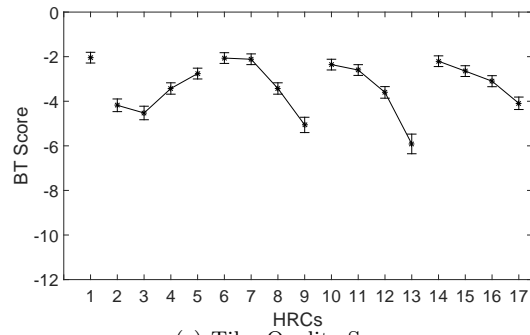
(c) Tile: Quality Score
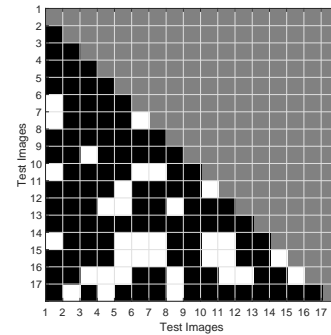
(d) Tile: Exact Test
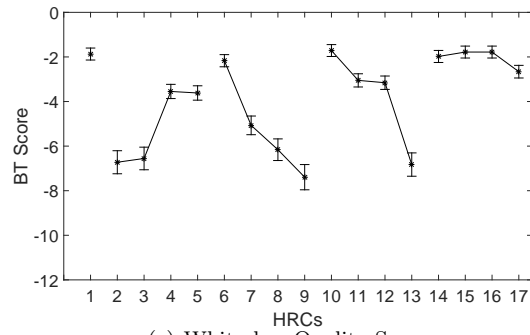
(e) Whitesky: Quality Score
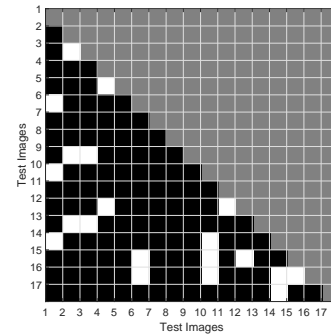
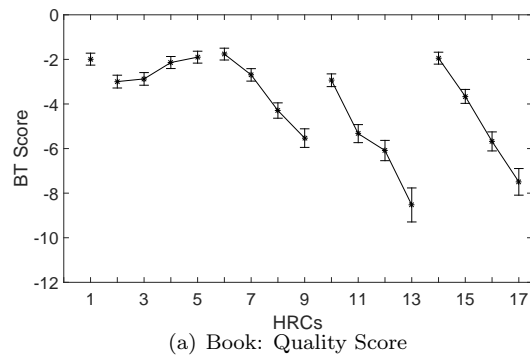(f) Whitesky: Exact Test
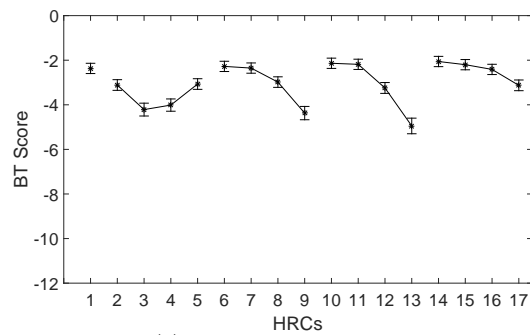
(g) Window: Quality Score
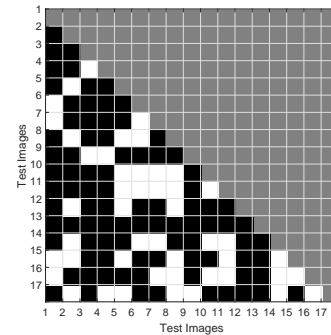
(h) Window: Exact Test

Figure 11.13: BT Score with 95% confidence interval and result of exact test for four images: Table, Tile, Whitesky, and Window.

Chapter 12

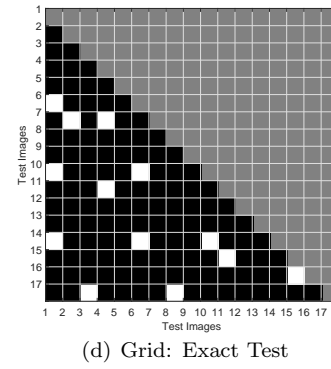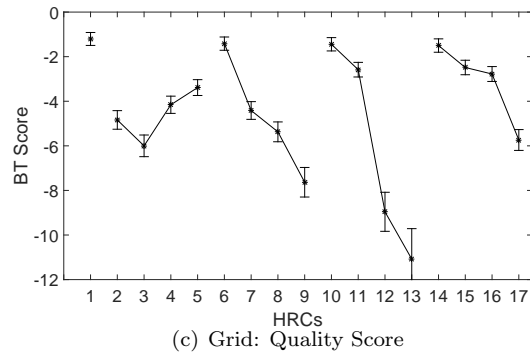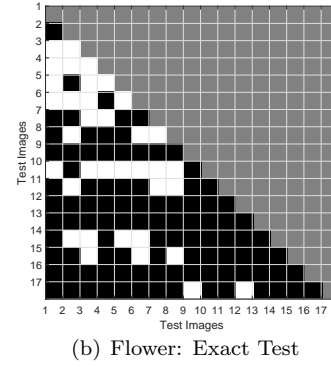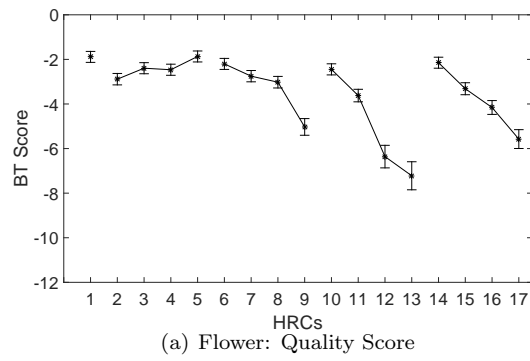# Conclusions

The aim of this dissertation is to explore the interesting but complex multidimensional problem, QoE in the multimedia communication services. The broad scope of the topic is making focused by considering two types of multimedia services: video communication and LF image communication.

The first two chapters (Chapiters 2 and 3) of this dissertation are focused to develop the theoretical and conceptual framework for assessing the QoE. In particular, the background study of multimedia communication, factors influencing multimedia quality, and QoE influencing factors are presented in Chapter 2. Followed by the background study, a QoE model and QoE centric multimedia management and optimization framework are presented. The model provides a holistic approach for devising an effective and efficient QoE assessment method, and the framework highlights the necessity for the quality assessment. In the first part of the Chapter 3, subjective experiment procedures followed by the data processing tools and result analysis techniques are presented. Whereas, the approaches used to design the objective quality assessment methods are presented in the second part of the Chapter 3.

To study the impact of transmission impairments on Video QoE, a video quality dataset: ReTRiEVED Video Quality Database, is presented in Chapter 4. The dataset was designed by considering QoS parameters: delay, jitter, packet loss, and bandwidth, and a real world streaming environment. The collected quality scores and corresponding test videos are already made available for the research community. The dataset is used to study the impact of transmission impairments on video QoE. The achieved results indicate that the adopted initial delay values do not have a significant impact on the perceived video quality, while jitter, PLR, and bandwidth do. In the performed experiments, the values 2 Mbps of bandwidth, 2 ms of jitter, and 3% of PLR can be considered as JND limits, because above and below these thresholds the perceived video quality changes significantly.

It is also noticed that if the communication channel is influenced significantly by impairments (high values of PLR and jitter, and low values of available bandwidth) the perceived quality is

mainly dependent on channel conditions. Therefore, the perceived quality is more dependent on the impairments rather than the video content and human factors. In another viewpoint, if the channel is less influenced by impairments, the perceived video quality is also depending on video content and human factors. Moreover, the influence of video content, and content related attributes on video QoE is studied by exploiting the ReTRiEVED dataset. To study the influence of video content related attributes such as spatial-temporal perceptual information, colorfulness, contrast, brightness, and HVS characteristics on QoE, another experiment was scheduled by considering the large number of SRCs with a wide rage of content variations. The achieved analysis results indicate that the except video frame size, the rest of the considered content related attributes do not have a significant relationship with the video QoE. The dataset and analysis of the results are presented in Chapter 5.

Followed by the above mentioned studies, the dataset is also used to benchmark the available VQMs. The achieved result strengthens the need of a new VQM. Therefore, a QoE metric suitable for communication environment is presented in Chapter 6. The proposed metric is evaluated and compared with other metrics. The achieved results indicate the supremacy of the proposed metric.

The next part of the dissertation is focused in LF imaging. An introduction on LF imaging, steps involving in the LF image communication system, generic LF image quality distortion model, and quality evaluation issues and their implications on LF IQA process are briefly presented in Chapter 7.

A watermarking technique for protecting the LF images from manipulations and unauthorized reproduction or diffusion is presented in Chapter 8. The impact of watermark on the quality of LF image is studied by using the objective and subjective measure. The achieved results indicate that the watermark technique does not significantly influence the quality of LF image. In particular, the subjective results define a visibility threshold ($\alpha = 0.3$) for the embedded data, while the objective test demonstrates that the reconstructed depth map is almost not affected by the watermarking system thus allowing a perfect rendering of the scene. Moreover, robustness of the proposed watermarking method are evaluated on the focused LF images for different compression techniques. The achieved results show that the watermarking scheme is robust only for very low level of compression for JPEG, JPEG2000, HEVC intra, and HEVC SS.

Due to the novelties on technology, there is a lack of subjective an objective quality assessment framework for LF image. Therefore, the last parts of this dissertation are focused in developing the quality assessment frameworks. First, an objective LF IQA technique is presented in Chapter 9. The depth map information about LF image is used as a reduced information. The measure of difference between the reference depth map and distorted depth map is used to estimate the QoE of the LF image. To train, test, and benchmark the objective metric a LF image quality dataset is designed. First, the dataset is used to evaluate the performance of the existing quality metrics, when applied to LF image. The achieved results strengthen the need of a new quality metric tuned to LF image.

The proposed quality assessment method is also evaluated on the dataset; the achieved results prove the importance of the proposed metric to the LF image. The dataset and the achieved results are also included in Chapter 9.

A generic subjective quality assessment framework comprises of source sequence selection, distortion model, and selection of the subjective quality assessment protocol to LF image is resented in Chapter 10. It is worthwhile to recall that the LF camera records, multiple views of a scene in a single shoot, and thus LF image composes lots of redundant information. Therefore, an aggressive compression of LF image is necessary for communication because of the redundancy and big size. The compression method reduces the size in the cost of artifacts; the produce artifacts ultimately degrades the QoE level. Therefore, a LF image quality dataset, referred as SMART LF image quality dataset, is created by considering Lytro Illum images and compression methods by following the framework. In this dataset, only an all-in-focused view (small part of the LF image) of the LF image is evaluated by the subjects. Therefore, by considering the limitation of the dataset, an another experiment is scheduled with a wide range of SRCs and rendering methods. Meanwhile, a LF image processing experimental setup is proposed for the subjective experiment of LF image. The setup presents all the steps from scene selection to the display. By exploiting the setup a massive LF image quality data set is going to be created aiming to benchmark the existing metrics and to develop the metric tuned to LF image.

Finally, by using the proposed dataset, SMART LF image quality dataset, perceptual quality of encoded image is studied. The achieved results show that the compression methods significantly degrade the perceived quality of LF image, and the level of the quality degradation varies for the images with different content. The performances of compression methods are also evaluated for LF images. The results show that the recently proposed plenoptic image compression method, SSDC [139], has a better perceived quality at a same level of compression compared to other considered compression methods: JPEG, JPEG2000, and HEVC intra. The results demonstrate the importance of specific compression algorithms for LF data in order to reach a best possible quality. Next, the dataset is also used to benchmark the widely used 2D IQMs, when applied to LF images. The results show that, among the considered metrics, the quality scores predicted by PSNR-HVS-M has a better correlation with corresponding subjective quality scores. However, there is still enough space to improve the performance of the metrics.

...

Bibliography

[1] The (old) stanford light fields archive. [online]`https://graphics.stanford.edu/software/lightpack/lifs.html`.

[2] Raytrix light field technology. [online]`https://www.raytrix.de/`.

[3] Synthetic light field archive. [online]`http://web.media.mit.edu/~gordonw/SyntheticLightFields/`.

[4] The consumer digital video library, 2016. [online] `http://www.cdvl.org/`.

[5] High Efficiency Video Coding (HEVC): HM Version 16.8, 2016. [online] `https://hevc.hhi.fraunhofer.de//`.

[6] JVET JEM software, 2016. [online] `https://jvet.hhi.fraunhofer.de/`.

[7] Lytro Illum light field dataset, 2016. [online] `http://www.irisa.fr/temics/demos/IllumDatasetLF/index.html`.

[8] Video quality assessment databases, 2016. [online] `http://ivc.univ-nantes.fr/en/pages/view/24/`.

[9] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *International Symposium on Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.

[10] John L Ackrill et al. *Categories and De interpretatione.* Clarendon Press, 1975.

[11] Edward H Adelson and James R Bergen. *The plenoptic function and the elements of early vision.* Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[12] Amar Aggoun, Emmanuel Tsekleves, Mohammad Rafiq Swash, Dimitrios Zarpalas, Anastasios Dimou, Petros Daras, Paulo Nunes, and Luí Ducla Soares. Immersive 3D holoscopic video system. *IEEE MultiMedia*, 20(1):28–37, 2013.

[13] Fritz Albregtsen et al. Statistical texture measures computed from gray level coocurrence matrices. *Image processing laboratory, department of informatics, university of oslo*, 5, 2008.

[14] Mohammed Alreshoodi and John Woods. Survey on QoE\ QoS correlation models for multimedia services. *arXiv preprint arXiv:1306.0221*, 2013.

[15] Louis Anegekuh, Lingfen Sun, Emmanuel Jammeh, Is-Haka Mkwawa, and Emmanuel Ifeachor. Content-based video quality prediction for HEVC encoded videos streamed over packet networks. *IEEE Transactions on Multimedia*, 17(8):1323–1334, 2015.

[16] Evlambios E Apostolidis and Georgios A Triantafyllidis. Watermark selection for light field rendering in FTV. In *2008 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 385–388. IEEE, 2008.

[17] John G Apostolopoulos and Amy R Reibman. The challenge of estimating video quality in video communication applications [in the spotlight]. *IEEE Signal Processing Magazine*, 29(2):160–158, 2012.

[18] Luigi Atzori, Chang Wen Chen, Tasos Dagiuklas, and Hong Ren Wu. QoE management in emerging multimedia services. *IEEE Communications Magazine*, 50(4):18–19, 2012.

[19] Sung-Ho Bae and Munchurl Kim. A novel image quality assessment with globally and locally consilient visual quality perception. *IEEE Transactions on Image Processing*, 25(5):2392–2406, 2016.

[20] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang. A quest for an internet video quality-of-experience metric. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pages 97–102. ACM, 2012.

[21] Amin Banitalebi-Dehkordi, Mahsa T Pourazad, and Panos Nasiopoulos. A study on the relationship between depth map quality and the overall 3D video quality of experience. In *3DTV-Conference: The True Vision-Capture, Transmission and Dispaly of 3D Video (3DTV-CON), 2013*, pages 1–4. IEEE, 2013.

[22] GA Barnard. A new test for 2× 2 tables. *Nature*, 156:177, 1945.

[23] Federica Battisti, Marco Carli, and Alessandro Neri. No reference quality assessment for mpeg video delivery over IP. *EURASIP Journal on Image and Video Processing*, 2014(1):1–19, 2014.

[24] Federica Battisti, Marco Carli, and Pradip Paudyal. QoS to QoE mapping model for wired/wireless video communication. In *Euro Med Telco Conference (EMTC), 2014*, pages 1–6. IEEE, 2014.

[25] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau. Quality assessment of stereoscopic images. *EURASIP journal on image and video processing*, 2008(1):1–13, 2009.

[26] Regina Bernhaupt, Marianna Obrist, and Manfred Tscheligi. Usability and usage of iTV services: lessons learned in an austrian field trial. *Computers in Entertainment (CIE)*, 5(2):6, 2007.

[27] Nigel Bevan, James Carter, and Susan Harker. ISO 9241-11 revised: What have we learnt about usability since 1998? In *International Conference on Human-Computer Interaction*, pages 143–151. Springer, 2015.

[28] Tom E Bishop, Sara Zanetti, and Paolo Favaro. Light field superresolution. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–9. IEEE, 2009.

[29] Henry David Block, Jacob Marschak, et al. Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, 2:97–132, 1960.

[30] Frank Bossen et al. Common test conditions and software reference configurations. *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-F900*, 2011.

[31] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[32] Tomás Brandão and Maria Paula Queluz. No-reference quality assessment of H.264/AVC encoded video. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11):1437–1447, 2010.

[33] Rec. ITU-R BT.1438. Subjective assessment of stereoscopic television pictures. 2000.

[34] Francisco C Calderon, Carlos A Parra, and Cesar L Niño. Depth map estimation in light fields using an stereo-like taxonomy. In *2014 XIX Symposium on Image, Signal Processing and Artificial Vision*, pages 1–5. IEEE, 2014.

[35] Patrizio Campisi, Marco Carli, Gaetano Giunta, and Alessandro Neri. Tracing watermarking for multimedia communication quality assessment. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 2, pages 1154–1158. IEEE, 2002.

[36] Damon M Chandler. Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Processing*, 2013, 2013.

[37] Damon M Chandler and Sheila S Hemami. VSNR: a wavelet-based visual signal-to-noise ratio for natural images. *IEEE transactions on image processing*, 16(9):2284–2298, 2007.

[38] K-T Chen, C-C Tu, and W-C Xiao. OneClick: a framework for measuring network quality of experience. In *INFOCOM 2009, IEEE*, pages 702–710. IEEE, 2009.

[39] Ni Chen, Xiuhua Jiang, and Caihong Wang. Impact of packet loss distribution on the perceived IPTV video quality. In *Image and Signal Processing (CISP), 2012 5th International Congress on*, pages 38–42. IEEE, 2012.

[40] Peter Y Chen and Paula M Popovich. *Correlation: Parametric and nonparametric measures*. Number 137-139. Sage, 2002.

[41] Wen-Hsiung Chen, CH Smith, and SC Fralick. A fast computational algorithm for the discrete cosine transform. *IEEE Transactions on communications*, 25(9):1004–1009, 1977.

[42] Yanjiao Chen, Kaishun Wu, and Qian Zhang. From QoS to QoE: a tutorial on video quality assessment. *IEEE Communications Surveys & Tutorials*, 17(2):1126–1165, 2015.

[43] Timmerer Christian, Pereira Fernando, and Ebrahimi Touradj. Quality of experience in multimedia systems and services: A journey towards the quality of life. http://www.slideshare.net/christian.timmerer/quality-of-experience-in-multimedia-systems-and-services-a-journey-towards-the-quality-of-life, 2016.

[44] CISCO. Cisco visual networking index: Forecast and methodology, $2015 - 2020$. *White Paper*, 2016.

[45] Mark Claypool and Jonathan Tanner. The effects of jitter on the peceptual quality of video. In *Proceedings of the seventh ACM international conference on Multimedia (Part 2)*, pages 115–118. ACM, 1999.

[46] Caroline Conti, Paulo Nunes, et al. Inter-layer prediction scheme for scalable 3-D holoscopic video coding. *IEEE Signal Processing Letters*, 20(8):819–822, 2013.

[47] Caroline Conti, Paulo Nunes, and Luis Ducla Soares. New HEVC prediction modes for 3D holoscopic video coding. In *2012 19th IEEE International Conference on Image Processing*, pages 1325–1328. IEEE, 2012.

[48] Caroline Conti, Paulo Nunes, and Luís Ducla Soares. HEVC-based light field image coding with bi-predicted self-similarity compensation. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*, pages 1–4. IEEE, 2016.

[49] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1027–1034, 2013.

[50] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams. Linear volumetric focus for light field cameras. *ACM Transactions on Graphics (TOG)*, 34(2):15, 2015.

[51] Francesca De Simone, Marco Tagliasacchi, Matteo Naccari, Stefano Tubaro, and Touradj Ebrahimi. A H.264/AVC video database for the evaluation of quality metrics. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2430–2433. IEEE, 2010.

[52] Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, Ibrahim Sezan, Thomas Huang, and Avideh Zakhor. Applications of video-content analysis and retrieval. *IEEE multimedia*, 9(3):42–55, 2002.

[53] Minh N Do, Davy Marchand-Maillet, and Martin Vetterli. On the bandwidth of the plenoptic function. *IEEE Transactions on Image Processing*, 21(2):708–717, 2012.

[54] G. Dansereau Donald. Light Field Toolbox for Matlab, 2015. [online]`https://it.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4`.

[55] Touradj Ebrahimi. Quality of multimedia experience: past, present and future. In *MM" 09: Proceedings of the seventeen ACM international conference on Multimedia*, number MMSPL-CONF-2010-003, pages 3–4, 2009.

[56] Touradj Ebrahimi. JPEG PLENO abstract and executive summary. 2015. [online]`https://jpeg.org/items/20150320_pleno_summary.html`.

[57] Karen Egiazarian, Jaakko Astola, Nikolay Ponomarenko, Vladimir Lukin, Federica Battisti, and Marco Carli. New full-reference quality metrics based on HVS. In *CD-ROM proceedings of the second international workshop on video processing and quality metrics, Scottsdale, USA*, volume 4, 2006.

[58] Muslim Elkotob, Daniel Grandlund, Karl Andersson, and Christer Ahlund. Multimedia QoE optimized management using prediction and statistical learning. In *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, pages 324–327. IEEE, 2010.

[59] ETSI . Speech and multimedia Transmission Quality (STQ); end-to-end jitter transmission planning requirements for real time services in an ngn context. *Technical Specification ETSI TS 103 210 V1.2.1 (2014-05)*, 2014.

[60] ETSI TS 103 294 V1.1.1. Speech and multimedia transmission quality (STQ); quality of experience; a monitoring architecture, 2014.

[61] Christos Faloutsos, Ron Barber, Myron Flickner, Jim Hafner, Wayne Niblack, Dragutin Petkovic, and William Equitz. Efficient and effective querying by image content. *Journal of intelligent information systems*, 3(3-4):231–262, 1994.

[62] Elena A Fedorovskaya, Huib de Ridder, and Frans JJ Blommaert. Chroma variations and perceived quality of color images of natural scenes. *Color research & application*, 22(2):96–110, 1997.

[63] Markus Fiedler, Tobias Hossfeld, and Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *IEEE Network*, 24(2):36–41, 2010.

[64] Markus Fiedler, Sebastian Möller, and Peter Reichl. Quality of experience: From user perception to instrumental metrics (dagstuhl seminar 12181). *Dagstuhl Reports*, 2(5), 2012.

[65] Karel Fliegel. QUALINET multimedia databases v5. 5. 2014.

[66] D Flynn and C Rosewarne. Common test conditions and software reference configurations for HEVC range extensions. In *Proceedings of the 14th Meeting of Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, 2013.

[67] International Organization for Standardization. *ISO 8402: Quality Management and Quality Assurance-Vocabulary*. International Organization for Standardization, 1994.

[68] Adrian Ford and Alan Roberts. Colour space conversions. *Westminster University, London*, 1998:1–31, 1998.

[69] Matthew Gaubatz and SS Hemami. MeTriX MuX visual quality assessment package. *foulard. ece. cornell. edu/gaubatz/metrix_mux*, 2011.

[70] Todor Georgiev and Andrew Lumsdaine. Depth of field in plenoptic cameras. In *Proc. Eurographics*, volume 2009, 2009.

[71] Todor Georgiev and Andrew Lumsdaine. Focused plenoptic camera and rendering. *Journal of Electronic Imaging*, 19(2):021106–021106, 2010.

[72] Deepti Ghadiyaram, Alan C Bovik, Hojatollah Yeganeh, Roman Kordasiewicz, and Michael Gallant. Study of the effects of stalling events on the quality of experience of mobile streaming videos. In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pages 989–993. IEEE, 2014.

[73] Alireza Ghasemi, Nelly Afonso, and Martin Vetterli. LCAV-31: a dataset for light field object recognition. In *IS&T/SPIE Electronic Imaging*, pages 902014–902014, 2013.

[74] Gheorghita Ghinea, Christian Timmerer, Weisi Lin, and Stephen R Gulliver. Mulsemedia: State of the art, perspectives, and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(1s):17, 2014.

[75] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen. The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 43–54. ACM, 1996.

[76] Dan Grois, Detlev Marpe, Amit Mulayoff, Benaya Itzhaky, and Ofer Hadar. Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders. In *Picture Coding Symposium (PCS), 2013*, pages 394–397. IEEE, 2013.

[77] Video Quality Experts Group et al. Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii (fr_tv2). *ftp://ftp. its. bldr-doc. gov/dist/ituvidq/Boulder_VQEG_jan_04/VQEG_PhaseII_FRTV_Final_Report_SG9060E. doc, 2003*, 2003.

[78] Video Quality Experts Group et al. Validation of reduced-reference and noreference objective models for standard definition television, phase i, 2009.

[79] Video Quality Experts Group et al. Report on the validation of video quality models for high definition video content. *VQEG, Geneva, Switzerland, Tech. Rep.[Online]. Available: http://www. its. bldrdoc. gov/vqeg/projects/hdtv/hdtv. aspx*, 2010.

[80] Video Quality Experts Group et al. Test plan for evaluation of video quality models for use with high definition tv content. *http://www.its.bldrdoc.gov/vqeg/projects/hdtv/hdtv.aspx*, 2010.

[81] Video Quality Experts Group et al. Video Quality Experts Group (VQEG) database, 2010. [online]http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx.

[82] Kwangsung Ha and Munchurl Kim. A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video. In *2011 18th IEEE International Conference on Image Processing*, pages 2525–2528. IEEE, 2011.

[83] John C Handley. Comparative analysis of bradley-terry and thurstone-mosteller paired comparison models for image quality assessment. In *PICS*, volume 1, pages 108–112, 2001.

[84] David Hands, Orion Vincent Barriac, and France Telecom. Standardization activities in the ITU for a QoE assessment of IPTV. *IEEE Communications Magazine*, 46(2):78–84, 2008.

[85] Philippe Hanhart, Emilie Bosc, Patrick Le Callet, and Touradj Ebrahimi. Free-viewpoint video sequences: A new challenge for objective quality metrics. In *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*, pages 1–6. IEEE, 2014.

[86] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.

[87] David Hasler and Sabine E Suesstrunk. Measuring colorfulness in natural images. In *Electronic Imaging 2003*, pages 87–95. International Society for Optics and Photonics, 2003.

[88] Takanori Hayashi. QoE-centric operation for optimizing user quality of experience. *Feature Articles: Network Science*, 13(9):1–5, 2015.

[89] Stephen Hemminger. Network emulation with NetEm, 2005. [online] `https://wiki.linuxfoundation.org/networking/netem/`.

[90] Chaminda TER Hewage, Stewart T Worrall, Safak Dogan, Stephane Villette, and Ahmet M Kondoz. Quality evaluation of color plus depth map-based stereoscopic video. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):304–318, 2009.

[91] Holografika. Full-angle 3D light field displaying, 2016. [online] `http://www.holografika.com/Holovizio80WLT/`.

[92] Tobias Hoßfeld, David Hock, Phuoc Tran-Gia, Kurt Tutschku, and Markus Fiedler. Testing the IQX hypothesis for exponential interdependency between QoS and QoE of voice codecs iLBC and G. 711. In *18th ITC Specialist Seminar Quality of Experience, Karlskrona, Sweden*, pages 105–114, 2008.

[93] David Hoyle. ISO 9000: quality systems handbook. 2001.

[94] Wu-Hsiao Hsu and Chi-Hsiang Lo. QoS/QoE mapping and adjustment model in the cloud-based multimedia infrastructure. *IEEE Systems Journal*, 8(1):247–255, 2014.

[95] Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Transactions on Graphics (TOG)*, 34(4):60, 2015.

[96] Syed Hussain, Richard Harris, Amal Punchihewa, and Zafar Iqbal. Dominant factors in the network domain that influence the QoE of an IPTV service. In *Modelling Symposium (EMS), 2013 European*, pages 629–634. IEEE, 2013.

[97] Institute for Telecommunication Sciences. Spatial Information (SI) Filter. 2016. [online]`http://www.its.bldrdoc.gov/resources/video-quality-research/guides-and-tutorials/spatial-information-si-filter.aspx`.

[98] Omneya Issa, Filippo Speranza, Tiago H Falk, et al. Quality-of-experience perception for video streaming services: Preliminary subjective and objective results. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–9. IEEE, 2012.

[99] Omneya Issa, Filippo Speranza, Wei Li, and Hong Liu. Estimation of time varying QoE for high definition IPTV distribution. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, pages 326–330. IEEE, 2012.

[100] ITS. General Model (aka VQM) and full reference calibration standards. *http://www.its.bldrdoc.gov/resources/video-quality-research/standards/hidden-general-model.aspx*, 2003.

[101] ITU. Test pictures and sequences for subjective assessments of digital codecs conveying signals produced according to recommendation itu-r bt.601. *Rec. BT.802-1*, 1994.

[102] ITUTG ITU. Opinion model for video-telephony applications.

[103] ITUTJ ITU. Draft: "hybrid perceptual/bitstream models for objective video quality measurements".

[104] ITUTJ ITU. Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.

[105] ITUTJ ITU. Requirements for an objective perceptual multimedia quality model.

[106] ITUTP ITU. Parametric non-intrusive assessment of audiovisual media streaming quality.

[107] ITUTP ITU. Parametric non-intrusive bitstream assessment of video media streaming quality.

[108] RBT ITU. Methodology for the subjective assessment of video quality in multimedia applications. *Rapport technique, International Telecommunication Union*, 2007.

[109] BT.500 ITU-R RECOMMENDATION. Methodology for the subjective assessment of the quality of television pictures. 2012.

[110] ITU-T E.860. Framework of a service level agreement.

[111] ITU-T Rec. E.802 . Rec. ITU-T E.802 : Framework and methodologies for the determination and application of qos parameters. *International Telecommunication Union Recommendation*, 2007.

[112] P ITU-T RECOMMENDATION. Subjective video quality assessment methods for multimedia applications. 1999.

[113] ITU-T Recommendation F.700. Framework recommendation for multimedia services, 2000.

[114] Aleks Jakulin. Baseline JPEG and JPEG2000 artifacts illustrated. *unpublished. Available: http://zeus. fri. unilj. si/~ aleks/jpeg/artifacts. htm*, 2004.

[115] Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon. Accurate depth map estimation from a lenslet light field camera. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1547–1555. IEEE, 2015.

[116] Young Ju Jeong, Jin-Ho Lee, Yang Ho Cho, Dongkyung Nam, Du-Sik Park, and C-C Jay Kuo. Efficient light-field rendering using depth maps for 100-Mpixel multi-projection 3D display. *Journal of Display Technology*, 11(10):792–799, 2015.

[117] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[118] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. Rendering for an interactive 360 light field display. In *ACM Transactions on Graphics (TOG)*, volume 26, page 40. ACM, 2007.

[119] Taichi Kawano, Kazuhisa Yamagishi, and Takanori Hayashi. Performance comparison of subjective assessment methods for 3D video quality. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 218–223. IEEE, 2012.

[120] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE, 2006.

[121] Brian W Keelan and Hitoshi Urabe. ISO 20462: a psychophysical image quality measurement standard. In *Electronic Imaging 2004*, pages 181–189. International Society for Optics and Photonics, 2003.

[122] Christian Keimel, Julian Habigt, Tim Habigt, Martin Rothbucher, and Klaus Diepold. Visual quality of current coding technologies at high definition IPTV bitrates. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 390–393. IEEE, 2010.

[123] Christian Keimel, Arne Redl, and Klaus Diepold. The TUM high definition video datasets. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 97–102. IEEE, 2012.

[124] Hyun Jong Kim, Dong Geun Yun, Hwa-Suk Kim, Kee Seong Cho, and Seong Gon Choi. QoE assessment model for video streaming service using QoS parameters in wired-wireless network. In *Advanced Communication Technology (ICACT), 2012 14th International Conference on*, pages 459–464. IEEE, 2012.

[125] Woo-Shik Kim, Antonio Ortega, PoLin Lai, Dong Tian, and Cristina Gomila. Depth map distortion analysis for view rendering and depth coding. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 721–724. IEEE, 2009.

[126] Jari Korhonen, Nino Burini, Junyong You, and Ehsan Nadernejad. How to evaluate objective video quality metrics reliably. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 57–62. IEEE, 2012.

[127] Alper Koz, Cevahir Cigla, and A Aydin Alatan. Free-view watermarking for free-view television. In *2006 International Conference on Image Processing*, pages 1405–1408. IEEE, 2006.

[128] Alper Koz, Cevahir Cigla, and A Aydın Alatan. Watermarking of free-view video. *IEEE Transactions on Image Processing*, 19(7):1785–1797, 2010.

[129] Lukáš Krasula, Karel Fliegel, Patrick Le Callet, and Miloš Klíma. Using full-reference image quality metrics for automatic image sharpening. In *SPIE Photonics Europe*, pages 913807–913807. International Society for Optics and Photonics, 2014.

[130] Sébastien Lê, Julie Josse, François Husson, et al. FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, 25(1):1–18, 2008.

[131] Patrick Le Callet, Sebastian Möller, Andrew Perkis, et al. Qualinet white paper on definitions of quality of experience. *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, 2012.

[132] Dohyoung Lee and Konstantinos N Plataniotis. Toward a no-reference image quality assessment using statistics of perceptual color descriptors. *IEEE Transactions on Image Processing*, 25(8):3875–3889, 2016.

[133] Jong-Seok Lee, Francesca De Simone, and Touradj Ebrahimi. Subjective quality evaluation via paired comparison: application to scalable video coding. *IEEE Transactions on Multimedia*, 13(5):882–893, 2011.

[134] Anat Levin and Yair Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1647, 2007.

[135] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.

[136] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. Light field microscopy. *ACM Transactions on Graphics (TOG)*, 25(3):924–934, 2006.

[137] Jing Li, Marcus Barkowsky, and Patrick Le Callet. Subjective assessment methodology for preference of experience in 3DTV. In *IVMSP Workshop, 2013 IEEE 11th*, pages 1–4. IEEE, 2013.

[138] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu. Saliency detection on light field. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2806–2813, 2014. [online`https://www.eecis.udel.edu/~nianyi/LFSD.htm`].

[139] Yun Li, Mårten Sjöström, Roger Olsson, and Ulf Jennehag. Scalable coding of plenoptic images by using a sparse set and disparities. *IEEE Transactions on Image Processing*, 25(1):80–91, 2016.

[140] Joe Yuchieh Lin, Rui Song, Chi-Hao Wu, TsungJung Liu, Haiqiang Wang, and C-C Jay Kuo. MCL-V: A streaming video quality assessment database. *Journal of Visual Communication and Image Representation*, 30:1–9, 2015.

[141] Weisi Lin. Perceptual quality evaluation for image and video: from modules to systems. *APSIPA Annual Summit and Conference (ASC): Tutorial*, 2013.

[142] Weisi Lin and C-C Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011.

[143] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, and Wenjun Zeng. Pseudo-sequence-based light field image compression. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*, pages 1–4. IEEE, 2016.

[144] Jichun Liu, Yang Geng, Deyuan Wang, Wenjing Li, and Xuesong Qiu. An objective multi-factor QoE evaluation based on content classification for H.264/AVC encoded video. In *2013 IEEE Symposium on Computers and Communications (ISCC)*, pages 000137–000142. IEEE, 2013.

[145] Tao Liu, Yao Wang, Jill M Boyce, Hua Yang, and Zhenyu Wu. A novel video quality metric for low bit-rate video considering both coding and packet-loss artifacts. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):280–293, 2009.

[146] Xingang Liu, Chao Sun, Kai Kang, and Lan Zhang. Joint 3-D image quality assessment metric by using image view and depth information over the networking in IoT. *IEEE Systems Journal*, 10(3):1203–1213, 2016.

[147] R Duncan Luce. Individual choice behavior: A theoretical analysis, 1959.

[148] R Duncan Luce. Thurstone and sensory scaling: Then and now. 1994.

[149] Robert Duncan Luce and Patrick Suppes. *Preference, utility, and subjective probability*. Wiley, 1965.

[150] Andrew Lumsdaine and Todor Georgiev. The focused plenoptic camera. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–8. IEEE, 2009.

[151] Aldo Maalouf and Mohamed-Chaker Larabi. CYCLOP: a stereo color image quality assessment metric. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1161–1164. IEEE, 2011.

[152] Rafał K Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. In *Computer Graphics Forum*, volume 31, pages 2478–2491. Wiley Online Library, 2012.

[153] Robert McGill, John W Tukey, and Wayne A Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.

[154] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 39–46. ACM, 1995.

[155] Vlado Menkovski, Adetola Oredope, Antonio Liotta, and Antonio Cuadra Sánchez. Predicting quality of experience in multimedia streaming. In *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia*, pages 52–59. ACM, 2009.

[156] Tahir Nawaz Minhas, Oziel Gonzalez Lagunas, Patrik Arlos, and Markus Fiedler. Mobile video sensitivity to packet loss and packet delay variation in terms of QoE. In *2012 19th International Packet Video Workshop (PV)*, pages 83–88. IEEE, 2012.

[157] Karan Mitra, Arkady Zaslavsky, and Christer Åhlund. Context-aware QoE modelling, measurement, and prediction in mobile computing systems. *IEEE Transactions on Mobile Computing*, 14(5):920–936, 2015.

[158] Theophano Mitsa and Krishna Lata Varkur. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 5, pages 301–304. IEEE, 1993.

[159] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[160] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013.

[161] Ricardo Monteiro, Luís Lucas, Caroline Conti, Paulo Nunes, Nuno Rodrigues, Sérgio Faria, Carla Pagliari, Eduardo da Silva, and Luís Soares. Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*, pages 1–4. IEEE, 2016.

[162] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, 2010.

[163] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, 2011.

[164] Anush Krishna Moorthy, Lark Kwon Choi, Alan Conrad Bovik, and Gustavo De Veciana. Video quality assessment on mobile devices: Subjective, behavioral and objective studies. *IEEE Journal of Selected Topics in Signal Processing*, 6(6):652–671, 2012.

[165] Antoine Mousnier, Elif Vural, and Christine Guillemot. Partial light field tomographic reconstruction from a fixed-camera focal stack. *arXiv preprint arXiv:1503.01903*, 2015. [online]http://www.irisa.fr/temics/demos/lightField/index.html.

[166] Takeshi Naemura, Junji Tago, and Hiroshi Harashima. Real-time video-based modeling and rendering of 3D scenes. *IEEE Computer Graphics and Applications*, 22(2):66–73, 2002.

[167] Alessandro Neri, Marco Carli, and Federica Battisti. A multi-resolution approach to depth field estimation in dense image arrays. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 3358–3362. IEEE, 2015.

[168] Ren Ng. *Digital light field photography*. PhD thesis, stanford university, 2006.

[169] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005.

[170] James Nightingale, Qi Wang, Christos Grecos, and Sergio Goma. The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming. *IEEE Transactions on Consumer Electronics*, 60(2):242–250, 2014.

[171] Andrey Norkin, Gisle Bjontegaard, Arild Fuldseth, Matthias Narroschke, Masaru Ikeda, Kenneth Andersson, Minhua Zhou, and Geert Van der Auwera. HEVC deblocking filter. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1746–1754, 2012.

[172] Mikko Nuutinen, Toni Virtanen, Mikko Vaahteranoksa, Tero Vuori, Pirkko Oittinen, and Jukka Häkkinen. CVD2014A database for evaluating no-reference video quality assessment algorithms. *IEEE Transactions on Image Processing*, 25(7):3073–3086, 2016.

[173] Tobias Oelbaum, Christian Keimel, and Klaus Diepold. Rule-based no-reference video quality evaluation using additionally coded videos. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):294–303, 2009.

[174] P Paudyal, F Battisti, A Neri, and M Carli. A study of the impact of light fields watermarking on the perceived quality of the refocused data. In *2015 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2015.

[175] Pradip Paudyal, Federica Battisti, and Marco Carli. Impact of video content and transmission impairments on quality of experience. *Multimedia Tools and Applications*, pages 1–25, 2016.

[176] Pradip Paudyal, Roger Olsson, Mårten Sjöström, Federica Battisti, and Marco Carli. SMART: a light field image quality dataset. In *Proceedings of the 7th International Conference on Multimedia Systems*, page 49, 2016. [online] `http://www.comlab.uniroma3.it/SMART.html`].

[177] Maria Paula, Toms Brando, Lus Roque, Mrcio Gonalves, and Rui Almeida. Video dataset, 2008. [online]`http://amalia.img.lx.it.pt/~tgsb/H264_test/`.

[178] Marius Pedersen, Nicolas Bonnier, Jon Y Hardeberg, and Fritz Albregtsen. Attributes of a new image quality model for color prints. In *Color and Imaging Conference*, volume 2009, pages 204–209. Society for Imaging Science and Technology, 2009.

[179] Andrew Perkis. Quality of experience (QoE) in multimedia applications. *SPIE Newsroom, DOI: 10.1117/2.1201302*, 4591, 2013.

[180] Margaret H Pinson, Marcus Barkowsky, and Patrick Le Callet. Selecting scenes for 2D and 3D subjective video quality tests. *EURASIP Journal on Image and Video Processing*, 2013(1):1, 2013.

[181] Margaret H Pinson and Stephen Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on broadcasting*, 50(3):312–322, 2004.

[182] Yohann Pitrey, Marcus Barkowsky, Romuald Pépion, Patrick Le Callet, and Helmut Hlavacs. Influence of the source content and encoding configuration on the perceived quality for scalable video coding. In *IS&T/SPIE Electronic Imaging*, pages 82911K–82911K. International Society for Optics and Photonics, 2012.

[183] Yohann Pitrey, Ulrich Engelke, Marcus Barkowsky, Romuald Pépion, and Patrick Le Callet. Subjective quality of svc-coded videos with different error-patterns concealed using spatial scalability. In *Visual Information Processing (EUVIP), 2011 3rd European Workshop on*, pages 180–185. IEEE, 2011.

[184] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, and Marco Carli. Modified image visual quality metrics for contrast change and mean shift accounting. *Proceedings of CADSM*, pages 305–311, 2011.

[185] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database tid2013: Peculiarities and preliminary results. In *Visual Information Processing (EUVIP), 2013 4th European Workshop on*, pages 106–111. IEEE, 2013.

[186] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.

[187] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, and Vladimir Lukin. On between-coefficient contrast masking of DCT basis functions. In *Proceedings of the third international workshop on video processing and quality metrics*, volume 4, 2007.

[188] R Raghavendra, Kiran Bylappa Raja, and Christoph Busch. Exploring the usefulness of light field cameras for biometrics: An empirical study on face and iris recognition. *IEEE Transactions on Information Forensics and Security*, 11(5):922–936, 2016. [online]`http://www.nislab.no/biometrics_lab/liffid`.

[189] Pedro de la Cruz Ramos, Joaqun Navarro Salmern, Raquel Prez Leal, and Francisco Gonzlez Vidal. Estimating perceived video quality from objective parameters in video over IP services. In *ICDT: The Seventh International Conference on Digital Telecommunications*, 2012.

[190] ITU-R Rec. *Future spectrum requirements estimate for terrestrial IMT*. 2013.

[191] ITUT Rec. *Rec. ITU-T E.800: Definitions of terms related to quality of service*. 1994.

[192] ITUT Rec. P. 800: Methods for subjective determination of transmission quality. *International Telecommunication Union, Geneva*, 1996.

[193] ITUT Rec. *Rec. ITU-T J.143: User requirements for objective perceptual video quality measurements in digital cable television*. 2000.

[194] ITUT Rec. *Rec. ITU-T G.1000: Communications Quality of Service: A framework and definitions*. 2001.

[195] ITUT Rec. Rec. ITU-T P. 10/G. 100 amendment 1, new appendix i–definition of quality of experience (qoe). *International Telecommunication Union*, 2007.

[196] Recommendation ITU-R BT.709-6. *Parameter values for the HDTV standards for production and international programme exchange*. 2015.

[197] Peter Reichl, Sebastian Egger, Raimund Schatz, and Alessandro D'Alconzo. The logarithmic nature of QoE and the role of the weber-fechner law in QoE assessment. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010.

[198] Ng Ren. Light fields. 2016. [online]`https://inst.eecs.berkeley.edu/~cs194-26/fa16/Lectures/light_fields.pdf`.

[199] Martin Řeřábek and Touradj Ebrahimi. Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments. In *SPIE Optical Engineering+ Applications*, pages 92170U–92170U. International Society for Optics and Photonics, 2014.

[200] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. In *8th International Conference on Quality of Multimedia Experience (QoMEX)*, number EPFL-CONF-218363, 2016.

[201] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Transactions on Image Processing*, 21(8):3339–3352, 2012.

[202] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind prediction of natural video quality. *IEEE Transactions on Image Processing*, 23(3):1352–1365, 2014.

[203] Telecommunication Standardization Sector. Subjective audiovisual quality assessment methods for multimedia applications. 1998.

[204] Telecommunication Standardization Sector. Opinion model for video-telephony applications, 2007.

[205] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack. Study of subjective and objective quality assessment of video. *IEEE transactions on image processing*, 19(6):1427–1441, 2010.

[206] Lukas Sevcik, Miroslav Voznak, and Jaroslav Frnda. Qoe prediction model for multimedia services in ip network applying queuing policy. In *Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2014), International Symposium on*, pages 593–598. IEEE, 2014.

[207] Mirghiasaldin Seyedebrahimi, Colin Bailey, and Xiao-Hong Peng. Model and performance of a no-reference quality assessment metric for video streaming. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(12):2034–2043, 2013.

[208] Muhammad Shahid, Jacob Søgaard, Jeevan Pokhrel, Kjell Brunnström, Kun Wang, Samira Tavakoli, and Narciso Gracia. Crowdsourcing based subjective quality assessment of adaptive

video streaming. In *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*, pages 53–54. IEEE, 2014.

[209] Feng Shao, Weijun Tian, Weisi Lin, Gangyi Jiang, and Qionghai Dai. Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions. *IEEE Transactions on Image Processing*, 25(5):2059–2074, 2016.

[210] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[211] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.

[212] R Shmueli, R Huber, and O Hadar. Effects of frame rate, frame size and mpeg2 compression on the perceived compressed video quality transmitted over lossy ip networks. In *Information Technology: Research and Education, 2004. ITRE 2004. 2nd International Conference on*, pages 49–54. IEEE, 2004.

[213] Ron Shmueli, Ofer Hadar, Revital Huber, Masha Maltz, and Merav Huber. Effects of an encoding scheme on perceived video quality transmitted over lossy internet protocol networks. *IEEE Transactions on Broadcasting*, 54(3):628–640, 2008.

[214] Lea Skorin-Kapov and Martín Varela. A multi-dimensional view of QoE: the ARCU model. In *MIPRO, 2012 Proceedings of the 35th International Convention*, pages 662–666. IEEE, 2012.

[215] Alvy Ray Smith. Color gamut transform pairs. *ACM Siggraph Computer Graphics*, 12(3):12–19, 1978.

[216] Wei Song and Dian W Tjondronegoro. Acceptability-based QoE models for mobile video. *IEEE Transactions on Multimedia*, 16(3):738–750, 2014.

[217] Murray R Spiegel. *Theory and Problems*. Schaum Publishing, 1961.

[218] Nicolas Staelens, Glenn Van Wallendael, Rik Van de Walle, Filip De Turck, and Piet Demeester. High definition H.264/AVC subjective video database for evaluating the influence of slice losses on quality perception. In *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pages 130–135. IEEE, 2013.

[219] Dongeun Suh, Insun Jang, and Sangheon Pack. Qoe-enhanced adaptation algorithm over dash for multimedia streaming. In *The International Conference on Information Networking 2014 (ICOIN2014)*, pages 497–501. IEEE, 2014.

[220] Richard Szeliski, Shai Avidan, and P Anandan. Layer extraction from multiple images containing reflections and transparency. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 246–253. IEEE, 2000.

[221] Yuichi Taguchi, Amit Agrawal, Srikumar Ramalingam, and Ashok Veeraraghavan. Axial light field for curved mirrors: Reflect your perspective, widen your view. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 499–506. IEEE, 2010.

[222] Mehrdad Panahpour Tehrani, Shinya Shimizu, Gauthier Lafruit, Takanori Senoh, Toshiaki Fujii, Anthony Vetro, and Masayuki Tanimoto. Use cases and requirements on freeviewpoint television (FTV). *ISO/IEC JTC1/SC29/WG11 MPEG N*, 14104, 2013.

[223] Louis L Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.

[224] Huawei Tian, Zheng Wang, Yao Zhao, Rongrong Ni, and Lunming Qin. Spread spectrum-based multi-bit watermarking for free-view video. In *International Workshop on Digital Watermarking*, pages 156–166. Springer, 2011.

[225] Christian Timmerer, Touradj Ebrahimi, and Fernando Pereira. Toward a new assessment of quality. *Networks*, 7:8, 2015.

[226] Toshiko Tominaga, Takanori Hayashi, Jun Okamoto, and Akira Takahashi. Performance comparisons of subjective quality assessment methods for mobile video. In *Quality of multimedia experience (QoMEX), 2010 second international workshop on*, pages 82–87. IEEE, 2010.

[227] Kristi Tsukida and Maya R Gupta. How to analyze paired comparison data. Technical report, DTIC Document, 2011.

[228] ITUT Tutorial. Objective perceptual assessment of video quality: Full reference television. *ITU-T Telecommunication Standardization Bureau*, 2004.

[229] IT Union. Subjective assessment of standard definition digital television (SDTV/HDTV) systems. *Rec. ITU-R BT.1129/BT.710*, 1998.

[230] IT Union. Test materials to be used in assessment of picture quality. *Recommendation ITU-R BT.1210-4*, 1998.

[231] IT Union. Subjective methods for the assessment of stereoscopic 3DTV systems. *Recommendation ITU-R BT*, 2015.

[232] Matthieu Urvoy, Marcus Barkowsky, Romain Cousseau, Yao Koudota, Vincent Ricorde, Patrick Le Callet, Jesús Gutiérrez, and Narciso Garcia. NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high

quality 3d stereoscopic sequences. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 109–114. IEEE, 2012.

[233] Vaibhav Vaish and Andrew Adams. The (New) Stanford Light Field Archive, 2008. [online]`http:/lightfield.stanford.edu`.

[234] Martın Varela, Lea Skorin-Kapov, Katrien De Moor, Peter Reichl, Chang Wen Chen, Periklis Chatzimisios, Tasos Dagiuklas, and Luigi Atzori. QoE- defining a user-centric concept for service quality. *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, page 5, 2015.

[235] Tay Vaughan. *Multimedia: Making it work*. Tata McGraw-Hill Education, 2006.

[236] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69, 2007.

[237] Alexandre Vieira, Helder Duarte, Cristian Perra, Luis Tavora, and Pedro Assuncao. Data formats for high efficiency coding of lytro-illum light fields. In *Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on*, pages 494–497. IEEE, 2015.

[238] Irene Viola, Martin Rerabek, Tim Bruylants, Peter Schelkens, Fernando Pereira, and Touradj Ebrahimi. Objective and subjective evaluation of light field image compression algorithms. In *32nd Picture Coding Symposium*, number EPFL-CONF-221601, 2016.

[239] Irene Viola, Martin Řeřábek, and Touradj Ebrahimi. A new approach to subjectively assess quality of plenoptic content. In *SPIE Optical Engineering+ Applications*, pages 99710X–99710X. International Society for Optics and Photonics, 2016.

[240] Mario Vranješ, Snježana Rimac-Drlje, and Krešimir Grgić. Review of objective video quality metrics and performance comparison using different databases. *Signal Processing: Image Communication*, 28(1):1–19, 2013.

[241] Phong V Vu and Damon M Chandler. ViS3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices. *Journal of Electronic Imaging*, 23(1):013016–013016, 2014.

[242] Caihong Wang and Xiuhua Jiang. Video quality assessment for IPTV services: A survey. In *Computing and Convergence Technology (ICCCT), 2012 7th International Conference on*, pages 182–186. IEEE, 2012.

[243] Qiaosong Wang, Haiting Lin, Yi Ma, Sing Bing Kang, and Jingyi Yu. Automatic layer separation using light field imaging. *arXiv preprint arXiv:1506.04721*, 2015.

[244] Tianyi Wang, Anjum Pervez, and Hua Zou. VQM-based QoS/QoE mapping for streaming video. In *Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on*, pages 807–812. IEEE, 2010.

[245] Ting-Chun Wang, Alexei Alyosha Efros, and Ravi Ramamoorthi. Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions On Pattern Analysis And Machine Intelligence*.

[246] Ting-Chun Wang, Jun-Yan Zhu, Ebi Hiroaki, Manmohan Chandraker, Alexei A Efros, and Ravi Ramamoorthi. A 4D light-field dataset and CNN architectures for material recognition. In *European Conference on Computer Vision*, pages 121–138. Springer, 2016. [online]`http://cseweb.ucsd.edu/~viscomp/projects/LF/`.

[247] Zhou Wang. Objective image quality assessment: Facing the real-world challenges. *Electronic Imaging*, 2016(13):1–6, 2016.

[248] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002.

[249] Zhou Wang and Alan C Bovik. Mean squared error: love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[250] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[251] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. Ieee, 2003.

[252] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4D light fields. In *VMV*, pages 225–226, 2013. [online`https://hci.iwr.uni-heidelberg.de/node/5052`].

[253] Florian Wickelmaier and Christian Schmid. A Matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods, Instruments, & Computers*, 36(1):29–40, 2004.

[254] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. 24(3):765–776, 2005.

[255] Stefan Winkler. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 6(6):616–625, 2012.

[256] Stefan Winkler. Efficient measurement of stereoscopic 3D video content issues. In *IS&T/SPIE Electronic Imaging*, pages 90160Q–90160Q. International Society for Optics and Photonics, 2014.

[257] Lars C Wolf, Carsten Griwodz, and Ralf Steinmetz. Multimedia communication. *Proceedings of the IEEE*, 85(12):1915–1933, 1997.

[258] Yuanyi Xue, Beril Erkin, and Yao Wang. A novel no-reference video quality metric for evaluating temporal jerkiness due to frame freezing. *IEEE Transactions on Multimedia*, 17(1):134–139, 2015.

[259] Jason C Yang, Matthew Everett, Chris Buehler, and Leonard McMillan. A real-time distributed light field camera. *Rendering Techniques*, 2002:77–86, 2002.

[260] Junyong You, Liyuan Xing, Andrew Perkis, and Xu Wang. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In *Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA*, 2010.

[261] Cha Zhang and Tsuhan Chen. Spectral analysis for sampling image-based rendering data. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(11):1038–1050, 2003.

[262] Cha Zhang and Tsuhan Chen. Active rearranged capturing of image-based rendering scenes-theory and practice. *IEEE transactions on Multimedia*, 9(3):520–531, 2007.

[263] F Zhang, S Li, L Ma, YC Wong, and KN Ngan. IVP subjective quality video database, 2011.

[264] Yi Zhang and Damon M Chandler. 3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception. *IEEE Transactions on Image Processing*, 24(11):3810–3825, 2015.