



IEEE Standard for the Perceptual Quality Assessment of Three Dimensional (3D), Ultra-High Definition (UHD) and High Dynamic Range (HDR) Contents

IEEE Computer Society

Standards Activities Board Standards Committee

IEEE Std 3333.1.2™-2024
(Revision of IEEE Std 3333.1.2-2017)



STANDARDS

IEEE Standard for the Perceptual Quality Assessment of Three Dimensional (3D), Ultra-High Definition (UHD) and High Dynamic Range (HDR) Contents

Developed by the

Standards Activities Board Standards Committee
of the
IEEE Computer Society

Approved 15 February 2024

IEEE SA Standards Board

This is a Redline document produced by Accuris.

Redlines are designed to save time and improve productivity by using the latest software technology to find and highlight document changes from one edition to the next. It includes a full-text version of the new document, plus an indication of changes from the previous version.

Because it may not be technically possible to capture all changes accurately, it is recommended that users consult previous editions as appropriate. In all cases, only the current base version of this publication is to be considered the official document. This Redline is produced solely by Accuris. The publisher has not prepared this Redline and is not responsible for its accuracy.

Contact:

<https://store accuristech.com/pages/contact.html>

Redline Processing Notes:

1. ~~Red Text~~ - Red strikethrough text denotes deletions.
2. Blue Text - Blue underlined text denotes modifications and additions.

Abstract: The world is witnessing a rapid advance in stereoscopic 3D (S3D), and ultra-high definition (UHD) technology. As a result, there is a need for accurate, quality, and visual comfort assessment techniques to foster the display device industry, as well as signal processing areas. Thorough assessments with respect to the human visual system (HVS) for S3D and UHD contents are presented in this standard. Moreover, several image and video databases are also publicly provided for any research purpose.

Keywords: accommodation and vergence conflict, foveation, human visual system, HVS, IEEE 3333.1.2™, quality assessment, quality of experience, QoE, saliency detection, stereoscopic, stereoscopic display, subjective assessment, UHD, UHD display, ultra-high definition, visual contents analysis, visual (dis)comfort

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2024 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published 2 August 2024. Printed in the United States of America.

IEEE is a registered trademark in the U.S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

PDF: ISBN 979-8-8557-0932-2 STD27088
Print: ISBN 979-8-8557-0933-9 STDPD27088

IEEE prohibits discrimination, harassment, and bullying.

For more information, visit <https://www.ieee.org/about/corporate/governance/p9-26.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE Standards documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page (<https://standards.ieee.org/ipr/disclaimers.html>), appear in all IEEE standards and may be found under the heading “Important Notices and Disclaimers Concerning IEEE Standards Documents.”

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents are developed within IEEE Societies and subcommittees of IEEE Standards Association (IEEE SA) Board of Governors. IEEE develops its standards through an accredited consensus development process, which brings together volunteers representing varied viewpoints and interests to achieve the final product. IEEE standards are documents developed by volunteers with scientific, academic, and industry-based expertise in technical working groups. Volunteers involved in technical working groups are not necessarily members of IEEE or IEEE SA and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE makes no warranties or representations concerning its standards, and expressly disclaims all warranties, express or implied, concerning all standards, including but not limited to the warranties of merchantability, fitness for a particular purpose and non-infringement. IEEE Standards documents do not guarantee safety, security, health, or environmental protection, or compliance with law, or guarantee against interference with or from other devices or networks. In addition, IEEE does not warrant or represent that the use of the material contained in its standards is free from patent infringement. IEEE Standards documents are supplied “AS IS” and “WITH ALL FAULTS.”

Use of an IEEE standard is wholly voluntary. The existence of an IEEE standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity, nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document should rely upon their own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus balloting process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English language version published by IEEE is the approved IEEE standard.

Use by artificial intelligence systems

In no event shall material in any IEEE Standards documents be used for the purpose of creating, training, enhancing, developing, maintaining, or contributing to any artificial intelligence systems without the express, written consent of IEEE SA in advance. “Artificial intelligence” refers to any software, application, or other system that uses artificial intelligence, machine learning, or similar technologies, to analyze, train, process, or generate content. Requests for consent can be submitted using the Contact Us form.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE SA Standards Board Operations Manual is not, and shall not be considered or inferred to be, the official position of IEEE or any of its committees and shall not be considered to be, or be relied upon as, a formal position of IEEE or IEEE SA. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that the presenter’s views should be considered the personal views of that individual rather than the formal position of IEEE, IEEE SA, the Standards Committee, or the Working Group. Statements made by volunteers may not represent the formal position of their employer(s) or affiliation(s). News releases about IEEE standards issued by entities other than IEEE SA should be considered the view of the entity issuing the release rather than the formal position of IEEE or IEEE SA.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE or IEEE SA. However, **IEEE does not provide interpretations, consulting information, or advice pertaining to IEEE Standards documents.**

Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its Societies and subcommittees of the IEEE SA Board of Governors are not able to provide an instant response to comments or questions, except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in evaluating comments or revisions to an IEEE standard is welcome to join the relevant IEEE SA working group. You can indicate interest in a working group using the Interests tab in the Manage Profile & Interests area of the [IEEE SA myProject system](#).¹ An IEEE Account is needed to access the application.

Comments on standards should be submitted using the [Contact Us](#) form.²

¹ Available at: <https://development.standards.ieee.org/myproject-web/public/view.html#landing>.

² Available at: <https://standards.ieee.org/about/contact/>.

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not constitute compliance to any applicable regulatory requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

Data privacy

Users of IEEE Standards documents should evaluate the standards for considerations of data privacy and data ownership in the context of assessing and using the standards in compliance with applicable laws and regulations.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under U.S. and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, neither IEEE nor its licensors waive any rights in copyright to the documents.

Photocopies

Subject to payment of the appropriate licensing fees, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400; <https://www.copyright.com/>. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every 10 years. When a document is more than 10 years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit [IEEE Xplore](#) or [contact IEEE](#).³ For more information about the IEEE SA or IEEE's standards development process, visit the IEEE SA Website.

Errata

Errata, if any, for all IEEE standards can be accessed on the [IEEE SA Website](#).⁴ Search for standard number and year of approval to access the web page of the published standard. Errata links are located under the Additional Resources Details section. Errata are also available in [IEEE Xplore](#). Users are encouraged to periodically check for errata.

Patents

IEEE standards are developed in compliance with the [IEEE SA Patent Policy](#).⁵

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE SA Website at <https://standards.ieee.org/about/sasb/patcom/patents.html>. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

IMPORTANT NOTICE

Technologies, application of technologies, and recommended procedures in various industries evolve over time. The IEEE standards development process allows participants to review developments in industries, technologies, and practices, and to determine what, if any, updates should be made to the IEEE standard. During this evolution, the technologies and recommendations in IEEE standards may be implemented in ways not foreseen during the standard's development. IEEE standards development activities consider research and information presented to the standards development group in developing any safety recommendations. Other information about safety practices, changes in technology or technology implementation, or impact by peripheral systems also may be pertinent to safety considerations during implementation of the standard. Implementers and users of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, data privacy, and interference protection practices and all applicable laws and regulations.

³ Available at: <https://ieeexplore.ieee.org/browse/standards/collection/ieee>.

⁴ Available at: <https://standards.ieee.org/standard/index.html>.

⁵ Available at: <https://standards.ieee.org/about/sasb/patcom/materials.html>.

Participants

At the time this draft standard was completed, the Human Factors for Visual Experiences Working Group had the following membership:

Sanghoon Lee, *Chair*
Patrick LeCallet, *Vice Chair*

Hezerul Abdul Karim
Sewoong Ahn
Federica Battisti
Marco Carli
Sungho Choi
Seonghwa Choi
Ricky Christanto
Sangkwon Jeong

Xin Jin
Haksub Kim
Jaekyung Kim
Jinwoo Kim
Jongyoo Kim
Seongyong Kim
Woojae Kim
Patrick Le Callet

Maria G. Martini
Anh-Duc Nguyen
Heeseok Oh
An Ping
Liquan Shen
Chao Yang
Min Zhang
Yu Zhang

The following members of the individual Standards Association balloting group voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

Robert Aiello
Kwok Shum Au
William Byrd
Juan Carreon
Pin Chang

Ritwik Chowdhury
Werner Hoelzl
SangKwon Jeong
Piotr Karocki
Sanghoon Lee
Daozhuang Lin

An Ping
R. K. Rannow
Dorothy Stanley
Eugene Stoudenmire
Yu Yuan

When the IEEE SA Standards Board approved this standard on 15 February 2024, it had the following membership:

David J. Law, *Chair*
Vacant Position, *Vice Chair*
Gary Hoffman, *Past Chair*
Alpesh Shah, *Secretary*

Sara R. Biyabani
Ted Burse
Stephen Dukes
Doug Edwards
J. Travis Griffith
Guido R. Hiertz
Ronald W. Hotchkiss
Hao Hu

Yousef Kimiagar
Joseph L. Koepfinger*
Howard Li
Xiaohui Liu
John Haiying Lu
Kevin W. Lu
Hiroshi Mano
Paul Nikolich

Robby Robson
Jon Walter Rosdahl
Mark Siira
Lei Wang
F. Keith Waters
Sha Wei
Philip B. Winston
Don Wright

*Member Emeritus

Introduction

This introduction is not part of IEEE Std 3333.1.2-2024, IEEE Standard for the Perceptual Quality Assessment of Three Dimensional (3D), Ultra-High Definition (UHD) and High Dynamic Range (HDR) Contents.

This standard is specifically written to high-end 3D and UHD service purveyors, 3D, UHD and HDR display makers and the 3D digital cinema industry.

Contents

1. Overview	10
1.1 Scope	10
1.2 Purpose	10
1.3 Word usage	11
2. Normative references	11
3. Definitions, acronyms, and abbreviations	12
3.1 Definitions	12
3.2 Acronyms and abbreviations	13
4. Overview of the standard	14
4.1 Quality assessment for 2D contents	14
4.2 Quality assessment for 3D contents	14
4.3 Quality assessment for ultra-high-definition contents	15
5. Quality assessment for 2D contents	15
5.1 General	15
5.2 Single scale structural similarity	15
5.3 Multi-scale structural similarity	17
5.4 Visual information fidelity for image quality assessment	18
5.5 Motion-based perceptual quality assessment of video (MOVIE)	21
5.6 Natural image quality evaluator (NIQE)	26
5.7 Spatial efficient entropic image and video quality (SpEED-QA)	29
5.8 Objective video quality assessment based on perceptually weighted mean squared error	33
6. Quality assessment for 3D contents	41
6.1 General	41
6.2 Full reference quality assessment of stereoscopic images using disparity-gradient-phase similarity	41
6.3 Reduced reference real-time quality assessment of stereoscopic images/video	44
6.4 Analysis of human 3D perception in stereo image	48
6.5 Spatial quality pooling based on human fine 3D perception	54
6.6 Objective image quality assessment of 3D synthesized views	56
6.7 A 3D subjective quality prediction model based on depth distortion	59
7. Quality of experience assessment for ultra-high-definition contents based on human visual system	63
7.1 General	63
7.2 Blind sharpness prediction for ultra-high-definition videos based on the human visual system	63
7.3 Visual preference assessment on ultra-high-definition images	70
8. The impact of compression and packet losses on 3D perception	74
8.1 General	74
8.2 Description of the proposed 3D video database	75
8.3 Subjective experiments	76
9. A database of stereoscopic images plus depth	80
9.1 Description of the proposed stereoscopic image database plus depth	80
9.2 Subjective experiment	82
Annex A (informative) Bibliography	85

IEEE Standard for the Perceptual Quality Assessment of Three-Dimensional (3D), Ultra-High-Definition (UHD) and High Dynamic Range (HDR) Contents

1. Overview

1.1 Scope

This standard establishes methods for quality assessment of 3D, ultra-high definition (UHD) and high dynamic range (HDR) contents based on physiological mechanisms such as perceptual quality and visual attention. This standard identifies and quantifies the following:

- Causes and visual attention of perceptual quality degradation for 3D, UHD, and HDR image and video contents
- Compression distortion, such as multi-view image and video compression
- Interpolation distortion by intermediate-view rendering, such as 3D, UHD and HDR warping, view synthesis
- Structural distortion, such as bit errors on wireless/wired transmission errors
- Visual attention according to the quality degradation
- Deep learning based model for saliency detection and QoE assessment

Key items are needed to characterize the 3D, UHD, and HDR database in terms of the human visual system. These key factors are constructed in conjunction with the visual factors used to perceptual ~~perceive~~ quality and visual attention.

1.2 Purpose

The purpose of this standard is to define quality metrics for 3D, UHD and HDR quality assessment, and establish guidelines in order for users to entertain contents 3D/UHD/HDR displays and devices without any risks.

The major parameters considered in this standard include viewers' characteristics, visual contents, visual environment and display and devices described in the scope. Although metrics and methods for assessing quality of images and videos on 2 dimensional (2D) displays have been established, there has been little progress in doing so in the field of the 3D/UHD/HDR domain. This is, in part, due to the fact that these quality metrics need to take into account additional factors accrued from the dimension extension and resolution changes.

Since the visual quality is eventually determined by the human eye, this standard will define how each human factor affects the visual quality of 3D, UHD, and HDR contents. This standard provides objective 3D/UHD/HDR image and video quality metrics that are in agreement with subjective human judgments and previously researched in the academy and the industry.

1.3 Word usage

The word *shall* indicates mandatory requirements strictly to be followed in order to conform to the standard and from which no deviation is permitted (*shall equals is required to*).^{6,7}

The word *should* indicates that among several possibilities one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required (*should equals is recommended that*).

The word *may* is used to indicate a course of action permissible within the limits of the standard (*may equals is permitted to*).

The word *can* is used for statements of possibility and capability, whether material, physical, or causal (*can equals is able to*).

⁶ The use of the word *must* is deprecated and cannot be used when stating mandatory requirements; *must* is used only to describe unavoidable situations.

⁷ The use of *will* is deprecated and cannot be used when stating mandatory requirements; *will* is only used in statements of fact.

2. Normative references

The following referenced documents are indispensable for the application of this document (i.e., they must be understood and used, so each referenced document is cited in text and its relationship to this document is explained). For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments or corrigenda) applies.

~~Recommendation~~ ITU-P.910, Subjective video quality assessment methods for multimedia applications.¹ ~~8~~

~~Recommendation ITU-R BT.500-12, Methodology for the Subjective Assessment of the Quality of Television Pictures.~~

~~Recommendation~~ ITU-R BT.500-13, Methodology for the subjective assessment of the quality of television pictures.

~~Recommendation~~ ITU-R BT.2021, Subjective methods for the assessment of stereoscopic 3DTV systems.

3. Definitions, acronyms, and abbreviations

3.1 Definitions

For the purposes of this document, the following terms and definitions apply. The *IEEE Standards Dictionary Online* should be consulted for terms not defined in this clause.² ~~9~~

accommodation: The process by which the vertebrate eye changes optical power to maintain a clear image or focus on an object as its distance varies.

cyclopean image: The image ~~that~~ ~~which~~ is fused and perceived by the human visual system.

disparity: The difference in image location of an object seen by the left and right eyes, resulting from the eyes' horizontal separation (parallax). Disparity map implies the difference by number of pixels.

¹ ITU-T publications are available from the International Telecommunications Union (<https://www.itu.int/> ~~http://www.itu.int/~~).

² *IEEE Standards Dictionary Online* ~~subscription~~ is available at: <http://dictionary.ieee.org>. [An IEEE account is required for access to the dictionary, and one can be created at no charge on the dictionary sign-in page.](#)

fovea: The part of the eye located in the center of the macular area of the retina.

human visual system: The part of the central nervous system that ~~which~~ gives organisms the ability to process visual detail, as well as enabling the formation of several non-image photo-response functions.

mean opinion score: In a subjective quality assessment, the average of all subjects ~~subjects'~~ rating scores of quality of experience.

natural scene statistics: ~~It is concerned with~~ The statistical regularities related to natural scenes.

photoreceptor: A specialized type of cell found in the retina that is capable of visual phototransduction. The great biological importance of photoreceptors is that they convert light (visible electromagnetic radiation) into signals that can stimulate biological processes.

quality assessment: The evaluation of quality of the service or product to determine the performance in relation to set standards.

quality of experience: The degree of delight or annoyance of the user of ~~with~~ 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 given ~~in the light of~~ the user's personality and current state.

retina: The sensitive layer of tissue at the back of the eye where sensory images are received.

saliency: The interested regions on image or video that is the state or quality by which it stands out relative to other areas.

segmentation: The act of segmenting an image into different meaningful regions to understand its contents.

sharpness: The degree of clarity of detail in image or video, which ~~This~~ is measured as granularity ~~of defined by~~ the boundaries between zones of different tones and ~~or~~ colors.

stereoscopic display: A display device capable of inducing depth perception in viewers by means of stereopsis for binocular parallax.

subjective assessment: An assessment of quality or visual discomfort where there is no pre-established objective measure or standard, and is thus based solely on the opinion of the evaluator or a group of observers.

vergence: [The](#) simultaneous movements of both eyes in opposite directions to obtain or maintain single binocular vision.

viewing distance: [The](#) distance between the viewer and display.

visual (dis)comfort: The state of mind that expresses ([un](#) ~~dis~~)satisfaction with the visual environment. In other words, the visual (dis)comfort is a subjective sensation which accompanies physiological changes, in this case the ease (stress) of viewing 3D content. As a subjective sensation, it can be measured by asking the viewer to report its level. The positive (negative) sensation usually increases (decreases) rapidly when the human observer gazes at a comfortable scenario or closes [their](#) ~~his~~ eyes.

3.2 Acronyms and abbreviations

2D	two-dimensional
3D	three-dimensional
3DSwIM	3D synthesized view image-quality metric
CTU	coding tree unit
DFT	discrete Fourier transform
DGP-SIM	disparity-gradient-phase-congruency similarity
DIBR	depth-image-based rendering
DMOS	differential mean opinion score
~~DSCQS~~	~~double stimulus continuous quality scale~~
FR	full reference
HD	high definition
HDR	high dynamic range
~~HEVC~~	~~high efficiency video coding~~
HVS	human visual system
IQA	image-quality assessment

~~LPF~~ ~~low pass filter~~

MOS mean opinion score

MSSIM mean structural similarity

NR no reference

NSS natural scene statistics

PLCC Pearson linear correlation coefficient

PLR packet loss rate

PSNR peak signal-to-noise ratio

QA quality assessment

QoE quality of experience

QP quantization parameter

RGB red, green, and blue channels

RMSE root mean squared error

RR reduced reference

RTP real-time transport protocol

SIQA stereoscopic image-quality [metrics](#) ~~assessment~~

SROCC spearman's rank correlation coefficient

SSIM structural similarity

SVR support vector regression

UHD ultra-high definition

4. ~~An~~ Overview of the standard

4.1 Quality assessment for 2D contents

The field of digital image and video processing deals, in large part, with signals that are meant to convey reproductions of visual information for human consumption, and many image and video processing systems, such as those for acquisition, compression, restoration, enhancement and reproduction, etc., operate solely on these visual reproductions. These systems typically involve tradeoffs between resources and the visual quality of the output. In order to make these tradeoffs there needs to be a way of measuring the quality of images or videos that come from a system running under a given configuration. The obvious way of measuring quality is to solicit the opinion of human observers. However, such subjective evaluations are not only cumbersome and expensive, but they also cannot be incorporated into automatic systems that adjust themselves in real-time based on the feedback of output quality. The goal of quality assessment (QA) research is, therefore, to design algorithms for objective evaluation of quality in a way that is consistent with subjective human evaluation. Such QA methods would prove invaluable for testing, optimizing, benchmarking, and monitoring applications.

4.2 Quality assessment for 3D contents

The separated perception of left and right views of stereoscopic 3D stimuli induces various binocular- relevant perception processes in the human visual system, which results in significantly different determination of the visual quality of the stimuli of 3D stimuli from that of 2D stimuli. Indeed, existing perceptual image-quality-assessment models designed for 2D images generally fail to predict quality scores of 3D images accurately. To analyze and predict the degree of perceptual quality of a 3D content, it is necessary to understand the contents in terms of spatial and temporal characteristics that are based on existing psychophysical and statistical models of binocular visual perception. In addition, a perceptual quality-prediction model of ~~depth-image-based-rendering~~ (DIBR-) synthesized views is suggested. Towards ~~Toward~~ this goal, specific artifacts related to DIBR systems need to be considered.

Mostly, the subjective assessment is inherited from what has been traditionally done for 2D subjective assessment as defined by ITU-P.910, ITU-R BT.500-13, and ITU-R BT.2021.¹⁰ ³ However, it is doubtful whether the result of subjective assessment conducted using the 2D subjective assessment methods de-fined in these standards ~~their results~~ are reliable enough to be used as the reference because the viewing environment is quite different from 2D due to intensive immersion of a user wearing the glasses in a dark lighting. Hence, to perform the 3D image and video assessments, contents ~~new~~ databases including depth information shall be required to cover the mechanisms ~~characteristics~~ of human perception, display mechanism, and viewing environment including display specification, viewing distance, and illumination attributes ~~so on~~.

4.3 Quality assessment for ultra-high-definition contents

Generally, the UHD systems have much higher resolution₇ as well as a much larger field of view₇ than HD systems. Therefore ~~As a consequence~~, it is expected that viewers experience more realistic contents on UHD displays. This fosters the study of quality-of-experience issues in UHD systems. In particular, sharpness and preference assessments stand out for the reason that one of the most critical factors discriminating the UHD and HD experiences is the perceived pixel densities by the viewers.

Many post-processing techniques have been studied in order to satisfy the viewers' expectation. Classically, it is believed that the users enjoy ~~have~~ better ~~enjoyment~~ when the visual contents are sharp and have high contrast. However, when the sharpness and/or contrast is too high, the contents become unnatural and difficult to be perceived. As a result, the search for an appropriate sharpness and contrast enhancement level is crucial and practical.

5. Quality assessment for 2D contents

5.1 General

A huge number of photos and videos are captured every day and finding ways to assess and control the quality of the digital photos is quite challenging. A problem of great interest is to find a way to evaluate the perceptual quality of the visual contents.

¹⁰ ³ Information on references can be found in Clause 2.

5.2 Single scale structural similarity

Let $x = \{x_i | i = 1, 2, \dots, N\}$ and $y = \{y_i | i = 1, 2, \dots, N\}$ be two discrete non-negative signals that have been aligned with each other (e.g., two image patches extracted from the same spatial location from two images being compared, respectively), and let μ_x , σ_x^2 , and σ_{xy}^2 be the mean of x , the variance of x , and the covariance of x and y , respectively. Approximately, μ_x and σ_x can be viewed as estimates of the luminance and contrast of x , and σ_{xy} measures the tendency of x and y to vary together, thus an indication of structural similarity. In Wang et al. [B36], the luminance, contrast and structure comparison measures were given as follows:¹¹

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (1)$$

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

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (3)$$

(New)

where

C_1 , C_2 , and C_3 are small constants given by $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, and

$$C_3 = \frac{C_2}{2}, \text{ respectively}$$

L is the dynamic range of the pixel values ($L = 255$ for 8 bits/pixel gray scale images)

$K_1 \ll 1$ and $K_2 \ll 1$ are two scalar constants

The general form of the Structural SIMilarity (SSIM) index between signal x and y is defined as:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (4)$$

¹¹ ⁴The numbers in brackets correspond to those of the bibliography in Annex A.

where

α, β and γ are parameters to define the relative importance of the three components. Specifically, set $\alpha = \beta = \gamma$, and the resulting SSIM index is given by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

(New)

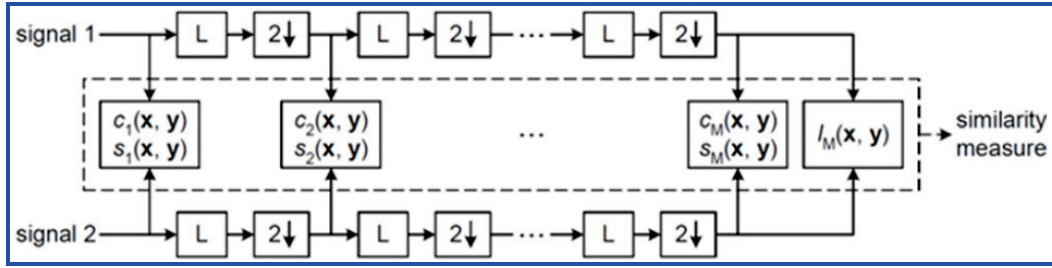
Equation (5) satisfies the following conditions:

- symmetry: $SSIM(x, y) = SSIM(y, x)$
- boundedness: $SSIM(x, y) \leq 1$
- unique maximum: $SSIM(x, y) = 1$ if and only if $x = y$

The drawback of setting $C_1 = C_2 = 0$ is that when the denominator of (5) is close to 0, the resulting measurement becomes unstable. This problem has been solved successfully by adding the two small constants C_1 and C_2 (calculated by setting $K_1 = 0.01$ and $K_2 = 0.03$, respectively).

The SSIM indexing algorithm for image quality assessment is applied using a sliding window approach. The window moves pixel-by-pixel across the whole image space. At each step, the SSIM index is calculated within the local window. If one of the images being compared is considered to have perfect quality, then the resulting SSIM index map can be viewed as the quality map of the other (distorted) image. Instead of using an 8×8 square window, a smooth windowing approach is used for local statistics to avoid “blocking artifacts” in the quality map. Finally, a mean SSIM index of the quality map is used to evaluate the overall image quality.

5.3 Multi-scale structural similarity



(New)

Figure 1 — Multi-scale structural similarity measurement system.

L: low-pass filtering; 2↓: downsampling by 2

The perceivability of image details depends on the sampling density of the image signal, the distance from the image plane to the observer, and the perceptual capability of the observer's visual system. In practice, the subjective evaluation of a given image varies when these factors vary. A single-scale method as described in 5.2 may be appropriate only for specific settings. A multi-scale method is a convenient way to incorporate image details at different resolutions.

This standard proposes a multi-scale SSIM method for image quality assessment whose system diagram is illustrated in Figure 1. Taking the reference and distorted image signals as the input, the system iteratively applies a low-pass filter and downsamples the filtered image by a factor of 2. The original image is indexed as Scale 1, and the highest scale as Scale M, which is obtained after M-1 iterations. At the j-th scale, the contrast comparison [see Equation (2)] and the structure comparison [Equation (3)] are calculated and denoted as $c_j(x, y)$ and $s_j(x, y)$, respectively. The luminance comparison [Equation (1)] is computed only at Scale M and is denoted as $l_M(x, y)$. The overall SSIM evaluation is obtained by combining the measurement at different scales using the following:

$$SSIM(x, y) = [l_M(x, y)]^{\alpha_M} \prod_{j=1}^M [c_j(x, y)]^{\beta} [s_j(x, y)]^{\gamma} \quad (6)$$

(New)

Similar to Equation (4), the exponents α_M , β_j , and γ_j are used to adjust the relative importance of different components. This multiscale SSIM index definition satisfies the three conditions given in 5.2. It also includes the single-scale method as a special case. In particular, a single-scale implementation for Scale M applies the iterative filtering and downsampling procedure up to Scale M and only the exponents α_M , β_M , and γ_M are given nonzero values. To simplify parameter selection, let $\alpha_j = \beta_j = \gamma_j$ for all j 's. In addition, normalize the cross-scale settings such that $\sum_{j=1}^M \gamma_j = 1$. This makes different parameter settings (including all single-scale and multi-scale settings) comparable. The remaining job is to determine the relative values across different scales. Conceptually, this should be related to the contrast sensitivity function (CSF) of the HVS, which states that the human visual sensitivity peaks at middle frequencies (around 4 cycles per degree of visual angle) and decreases along the high- and low-frequency directions. However, the CSF cannot be directly used to derive the parameters in the system because it is typically measured at the visibility threshold level using simplified stimuli (sinusoids), but our purpose is to compare the quality of complex structured images at visible distortion levels.

5.4 Visual information fidelity for image quality assessment

5.4.1 Source model

The NSS model that is used is the GSM model in the wavelet domain. It is convenient to deal with one subband of the wavelet decomposition at this point and later generalize this for multiple subbands. A GSM is a random field (RF) that can be expressed as a product of two independent RFs. That is, a GSM $C = \{\vec{C}_i | i \in I\}$, where I denotes the set of spatial indices for the RF, can be expressed as follows:

$$C = S \times U = \{S_i \times \vec{U}_i | i \in I\} \quad (7)$$

(New)

where

$$S = \{S_i | i \in I\}$$

is an RF of positive scalars

$$U = \{\vec{U}_i | i \in I\}$$

is a Gaussian vector RF with mean zero and covariance CU

$\vec{C}_i$ and $\vec{U}_i$ are M-dimensional vectors, and assume that for the RF U , $\vec{U}_i$ is independent of $\vec{U}_j, \forall i \neq j$. this standard, model each subband of a scale-space-orientation wavelet decomposition (such as the steerable pyramid) of an image as a GSM RF. Partition the subband coefficients into nonoverlapping blocks of M coefficients each, and model block i as the vector $\vec{C}_i$.

5.4.2 Distortion model

The purpose of a distortion model is to describe how the statistics of an image are disturbed by a generic distortion operator. The distortion model that was chosen provides important functionality while being mathematically tractable and computationally simple. It is a signal attenuation and additive noise model in the wavelet domain:

$$D = GC + V = \left\{ g_i \vec{C}_i + \vec{V}_i \mid i \in I \right\} \quad (8)$$

(New)

where

C denotes the RF from a subband in the reference signal

$$D = \left\{ \vec{D}_i \mid i \in I \right\}$$

denotes the RF from the corresponding subband from the test (distorted) signal

$$G = \left\{ g_i \mid i \in I \right\}$$

is a deterministic scalar gain field

$$V = \left\{ \vec{V}_i \mid i \in I \right\}$$

is a stationary additive zero-mean Gaussian noise RF with variance $C_V = \sigma_v^2 I$

The RF V is white, and is independent of S and U . Constrain the field G to be slowly varying.

5.4.3 Human visual system

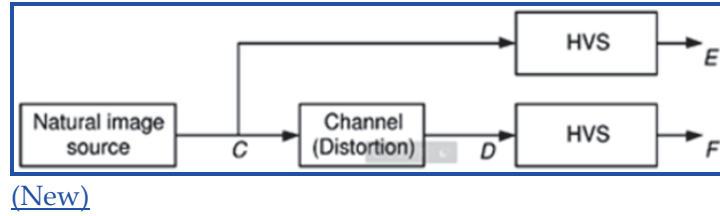


Figure 2 — Mutual information between C and E quantifies the information that the brain could ideally extract from the reference image, whereas the mutual information between C and F quantifies the corresponding information that could be extracted from the test image

In this standard, the HVS is approached as a “distortion channel” that imposes limits on how much information could flow through it. Although one could model different components of the HVS using psychophysical data, the purpose of HVS model in the information fidelity setup is to quantify the uncertainty that the HVS adds to the signal that flows through it. As a matter of analytical and computational simplicity, and more importantly to ease the dependency of the overall algorithm on viewing configuration information, all sources of HVS uncertainty are lumped into one additive noise component that serves as a distortion baseline in comparison to which the distortion added by the distortion channel can be evaluated. This lumped HVS distortion is called visual noise, and is modeled as a stationary, zero mean, additive white Gaussian noise model in the wavelet domain. Thus, the HVS noise in the wavelet domain is modeled as stationary RFs $N = \{\vec{N}_i | i \in I\}$ and

$N' = \{\vec{N}'_i | i \in I\}$, where $\vec{N}_i$ and $\vec{N}'_i$ are zero-mean uncorrelated $I(\vec{C}^N, \vec{D}^N | S^N = s^N)$ gaussian with the same dimensionality as $\vec{C}_i$.

$$E = C = N \quad (9)$$

$$F = D = N' \quad (10)$$

where

E and F denote the visual signal at the output of the HVS model from the reference and the test images in one subband, respectively, from which the brain extracts cognitive information (Figure 2)

The RFs N and N' are assumed to be independent of U , S , and V . Model the covariance of N and N' as follows:

$$\mathbf{C}_N = \mathbf{C}_{N'} = \sigma_n^2 \mathbf{I} \quad (11)$$

(New)

where

σ_n^2 is an HVS model parameter (variance of the visual noise)

5.4.4 Visual information fidelity criterion (IFC)

With the source, distortion, and HVS models as described in 5.4.3, the VIF criterion that we propose can be derived. Let $\vec{C}^N = (\vec{C}_1, \vec{C}_2, \dots, \vec{C}_N)$ denote N elements from C . Let S^N , $\vec{D}^N$, $\vec{E}^N$, and $\vec{F}^N$ be correspondingly defined. In this subclause, it is assumed that the model parameters G , σ_v^2 , and σ_n^2 are known.

The mutual information $I(\vec{C}^N, \vec{E}^N)$ quantifies the amount of information that can be extracted from the output of the HVS by the brain when the test image is being viewed. However, the key concern is the quality of a particular reference-test image pair, rather than the average quality of the ensemble of images as they pass through the distortion channel. It is, therefore, reasonable to tune the natural scene model to a specific reference image by treating $I(\vec{C}^N, \vec{D}^N | S^N = s^N)$ instead of $I(\vec{C}^N, \vec{D}^N)$, where S^N denotes a realization of for a particular reference image. The realization S^N could be thought of as “model parameters” for the associated reference image.

For the reference image, analyze $I(\vec{C}^N, \vec{E}^N | S^N = s^N)$, where S^N denotes a realization of S^N . In this standard, $I(\vec{C}^N, \vec{E}^N | S^N = s^N)$ is denoted as $I(\vec{C}^N, \vec{E}^N | s^N)$ for brevity. With the stated assumptions on C and the distortion model (8), the natural scene model can be written as:

$$I(\vec{C}^N, \vec{E}^N | s^N) = \sum_{j=1}^N \sum_{i=1}^N I(\vec{C}_i, \vec{E}_j | \vec{C}^{i-1}, \vec{E}^{j-1}, s^N) \quad (12)$$

$$= \sum_{i=1}^N I(\vec{C}_i, \vec{E}_i | s_i) \quad (13)$$

$$= \sum_{i=1}^N \left(h(\vec{C}_i + \vec{N}_i | s_i) - h(\vec{N}_i | s_i) \right) \quad (14)$$

$$= \frac{1}{2} \sum_{i=1}^N \log_2 \left(\frac{|s_i^2 \mathbf{C}_U + \sigma_n^2 \mathbf{I}|}{|\sigma_n^2 \mathbf{I}|} \right) \quad (15)$$

(New)

Equation (12) can be derived from chain rule and Equation (13) from the conditional independence of C and N given, $|\cdot|$ denotes the determinant.

Similarly, the following can be shown for the test image:

$$\begin{aligned} I(\vec{C}^N, \vec{F}^N | s^N) &= \sum_{i=1}^N \left(h(g_i \vec{C}_i + \vec{V}_i + \vec{N}_i | s_i) - h(\vec{V}_i + \vec{N}_i | s_i) \right) \\ &= \frac{1}{2} \sum_{i=1}^N \log_2 \left(\frac{|g_i^2 s_i^2 \mathbf{C}_U + (\sigma_v^2 + \sigma_n^2) \mathbf{I}|}{|(\sigma_v^2 + \sigma_n^2) \mathbf{I}|} \right) \end{aligned} \quad (16)$$

(New)

Since $\mathbf{C}_U$ is symmetric, it can be factored as $\mathbf{C}_U = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$, where $\mathbf{Q}$ is an orthonormal matrix, and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues λ_k . One can use this matrix factorization to show the following:

$$I(\vec{C}^N, \vec{E}^N | s^N) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M \log_2 \left(1 + \frac{s_i^2 \lambda_k}{\sigma_n^2} \right) \quad (17)$$

$$I(\vec{C}^N, \vec{F}^N | s^N) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M \log_2 \left(1 + \frac{g_i^2 s_i^2 \lambda_k}{\sigma_v^2 + \sigma_n^2} \right) \quad (18)$$

(New)

Also, Equation (17) and Equation (18) only consider a single subband. Incorporating multiple subbands is achievable by assuming independence between them in terms of both RFs and distortion model parameters. Thus, the VIF proposed in this standard is given by:

$$VIF = \frac{\sum_{j \in \text{subbands}} I(\vec{C}^{N,j}, \vec{F}^{N,j} | s^{N,j})}{\sum_{j \in \text{subbands}} I(\vec{C}^{N,j}, \vec{E}^{N,j} | s^{N,j})} \quad (19)$$

(New)

where

the sum is over the subbands of interest

$\vec{C}^{N,j}$ represents N elements of the RF C_j that describes the coefficients from subband j , etc.

5.5 Motion-based perceptual quality assessment of video (MOVIE)

5.5.1 Spatial MOVIE

Let $r(i)$ and $d(i)$ denote the reference and distorted videos respectively, where i represents a pixel location in the video. The reference and distorted videos are passed through the Gabor filter bank to obtain bandpass filtered videos. Denote the Gabor filtered reference video by $\tilde{f}(i, k)$ and the Gabor filtered distorted video by $\tilde{g}(i, k)$, where $k = 1, 2, \dots, K$ indexes the filters in the Gabor filterbank. Specifically, let $k = 1, 2, \dots, K/P$ correspond to the finest scale, $k = \frac{K}{P} + 1, \dots, \frac{2K}{P}$ the second finest scale and so on. All quality computations begin locally, using local windows B of coefficients extracted from each of the Gabor sub-bands, where the window B spans N pixels. Consider a pixel location i_0 . Let $f(k)$ be a vector of dimension N , where $f(k)$ is composed of the complex magnitude of N elements of $\tilde{f}(i, k)$ spanned by the window B centered on i_0 . The Gabor coefficients are $\tilde{f}(i, k)$ complex, but the vectors $f(k)$ are real and denote the Gabor channel amplitude response. The dependence on the spatio-temporal location i has been dropped for notational convenience by considering a specific location i_0 . If the window B is specified by a set of relative indices, then $f(k) = \{\tilde{f}(i_0 + m, k), m \in B\}$. Similar definition applies for $g(k)$. To index each element of $f(k)$, use the notation $f(k) = [f_1(k), f_2(k), \dots, f_N(k)]^T$.

Define the following spatial error from each subband response:

$$E_s(i_0, k) = \frac{1}{2} \frac{1}{N} \sum_{n=1}^N \left(\frac{f_n(k) - g_n(k)}{M(k) + C_1} \right)^2 \quad (20)$$

$$M(k) = \max \left(\sqrt{\frac{1}{N} \sum_{n=1}^N |f_n(k)|^2}, \sqrt{\frac{1}{N} \sum_{n=1}^N |g_n(k)|^2} \right) \quad (21)$$

(New)

The Gaussian filter responds to the mean intensity or the dc component of the two images. A spatial error index can be defined using the output of the Gaussian filter operating at dc . Let $f(dc)$ and $g(dc)$ denote a vector of dimension N extracted at i_0 from the output of the Gaussian filter operating on the reference and test videos respectively using the same window B . The variables $f(dc)$ and $g(dc)$ are low-pass filtered versions of the two videos. First, remove the effect of the mean intensity from each video before error computation since this acts as a bias to the low frequencies present in the reference and distorted images that are captured by the Gaussian filter. Estimate the mean as the average of the Gaussian filtered output:

$$\mu_f = \frac{1}{N} \sum_{n=1}^N f_n(dc), \mu_g = \frac{1}{N} \sum_{n=1}^N g_n(dc) \quad (22)$$

(New)

An error index for the dc sub-band is then computed in a similar fashion as the Gabor sub-bands:

$$E_{dc}(i_0) = \frac{1}{2} \frac{1}{N} \sum_{n=1}^N \left(\frac{|f_n(dc) - \mu_f| - |g_n(dc) - \mu_g|}{M_{dc} + C_2} \right)^2 \quad (23)$$

(New)

$$M_{dc} = \max \left(\sqrt{\frac{1}{N} \sum_{n=1}^N |f_n(dc) - \mu_f|^2}, \sqrt{\frac{1}{N} \sum_{n=1}^N |g_n(dc) - \mu_g|^2} \right) \quad (24)$$

(New)

The constants C_1 and C_2 in Equation (20) and Equation (21) are added to prevent instability when the denominators are small and are chosen to be $C_1 = 0.1$, $C_2 = 1$. The spatial error indices computed from all the Gabor sub-bands and the Gaussian sub-band can then be pooled to obtain an error index for location i_0 using the following:

$$E_S(i_0) = \frac{\sum_{k=1}^K E_S(i_0, k) + E_{dc}(i_0)}{K+1} \quad (25)$$

(New)

Finally, convert the error index to a quality index at location i_0 using the following:

$$Q_S(i_0) = 1 - E_S(i_0) \quad (26)$$

5.5.2 Temporal MOVIE

First, the analysis focuses on how translational motion manifests itself in the frequency domain. Let $a(x, y)$ denote an image patch and let $A(u, v)$ denote its Fourier transform. Assuming that this patch undergoes translation with a velocity $[\lambda, \phi]$ where λ and ϕ denote velocities along the x and y directions respectively, the resulting video sequence is given by $b(x, y, t) = a(x - \lambda t, y - \phi t)$. Then $B(u, v, w)$, the Fourier transform of $b(x, y, t)$, lies entirely within a plane in the frequency domain. Moreover, the magnitudes of the spatial frequencies do not change but are simply sheared:

$$B(u, v, w) = \begin{cases} \Lambda(u, v) & \text{if } \lambda u + \phi v + w = 0 \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

(New)

Let λ be a vector of dimension N , where λ is composed of N elements of the horizontal component of the flow field of the reference sequence spanned by the window B centered on i_0 . Similarly, ϕ represents the vertical component of flow. Then, using Equation (27), the spectrum of the reference video lies along:

$$\lambda_n u + \phi_n v + w = 0, n = 1, 2, \dots, N \quad (28)$$

Define a sequence of distance vectors $\delta(k)$, $k = 1, 2, \dots, K$ of dimension N . Each element of this vector denotes the distance of the center frequency of the k -th filter from the plane containing the spectrum of the reference video in a window centered on i_0 extracted using B . Let $U_0(k) = [u_0(k), v_0(k), w_0(k)]$, $k = 1, 2, \dots, K$ represent the center frequencies of all the Gabor filters. Then, $\delta(k)$ represents the perpendicular distance of a point from a plane defined by Equation (28) in a 3-dimensional space and is given by:

$$\delta_n(k) = \left| \frac{\lambda_n u_0(k) + \phi_n v_0(k) + w_o(k)}{\sqrt{\lambda_n^2 + \phi_n^2 + 1}} \right|, n = 1, 2, \dots, N \quad (29)$$

(New)

Based on distances in Equation (28) a set of weights based on these distances. The objective is to assign the filters that intersect the spectral plane to have the maximum weight of all filters. The distance of the center frequencies of these filters from the spectral plane is the minimum of all filters. First, define $\alpha'(k)$, $k = 1, 2, \dots, K$ using:

$$\alpha'_n k = \frac{\rho(k) - \delta_n(k)}{\rho(k)} \quad (30)$$

(New)

where

$\rho(k)$ denotes the radius of the sphere along which the center frequency of the k-th filter lies in the frequency domain

To achieve excitatory and inhibitory weights, all weights at each scale are shifted to have a zero-mean. Finally, to make the weights insensitive to the filter geometry that was chosen, normalize them so that the maximum weight is 1. This ensures that the maximum weight remains 1 irrespective of whether the spectral plane exactly intersects the center frequencies of the Gabor filters. We hence have a weight vector $\alpha(k)$, $k = 1, 2, \dots, K$ with elements:

$$\alpha_n(k) = \frac{\alpha'_n(k) - \mu_\alpha}{\max_{k=1, 2, \dots, \frac{K}{P}} [\alpha'_n(k) - \mu_\alpha]}, k = 1, 2, \dots, \frac{K}{P} \quad (31)$$

(New)

where

$$\mu_\alpha = \frac{\sum_{k=1}^{\frac{K}{P}} \alpha'_n(k)}{\frac{K}{P}}$$

(New)

Similar definitions apply for other scales.

Motion tuned responses from the reference and distorted video sequences may be constructed using these weights. Define N -vectors v^r and v^d using:

$$v_n^r = \frac{(f_n(dc) - \mu_f)^2 + \sum_{k=1}^K \alpha_n(k) f_n(k)^2}{(f_n(dc) - \mu_f)^2 + \sum_{k=1}^K f_n(k)^2 + C3} \quad (32)$$

$$v_n^d = \frac{(g_n(dc) - \mu_g)^2 + \sum_{k=1}^K \alpha(k) g_n(k)^2}{(g_n(dc) - \mu_g)^2 + \sum_{k=1}^K g_n(k)^2 + C3} \quad (33)$$

(New)

Define a temporal error index using Equation (34).

$$E_T(i_0) = \frac{1}{N} \sum_{n=1}^N (v_n^r - v_n^d)^2 \quad (34)$$

(New)

The error index in Equation (34) is also exactly 0 when the reference and test images are identical. Finally, convert the error index into a quality index using Equation (35).

$$Q_T(i_0) = 1 - E_T(i_0) \quad (35)$$

5.5.3 MOVIE index

Define frame level quality indices for spatial MOVIE using the ratio of the standard deviation to the mean of the spatial MOVIE scores for that frame. Similar definition applies for temporal MOVIE.

$$FQ_S(t_j) = \frac{\sigma_{Q_S(x,y,t_j)}}{\mu_{Q_S(x,y,t_j)}}, FQ_T(t_j) = \frac{\sigma_{Q_T(x,y,t_j)}}{\mu_{Q_T(x,y,t_j)}} \quad (36)$$

(New)

The spatial MOVIE index is then defined as the average of these frame level descriptors.

$$\text{SpatialMOVIE} = \frac{1}{\tau} \sum_{j=1}^{\tau} FQ_S(t_j) \quad (37)$$

(New)

The range of values of the temporal MOVIE scores is smaller than that of the spatial scores, due to the large divisive normalization in Equation (14) and Equation (15). To offset this effect, use the square root of the temporal scores.

$$\text{TemporalMovie} = \sqrt{\frac{1}{\tau} \sum_{j=1}^{\tau} FQ_T(t_j)} \quad (38)$$

(New)

The MOVIE index is defined as:

$$\text{MOVIE} = \text{SpatialMOVIE} \times \text{TemporalMOVIE} \quad (39)$$

5.6 Natural image quality evaluator (NIOE)

5.6.1 Spatial domain natural scene statistic (NSS)

The classical spatial NSS model that is used begins by preprocessing the image by processes of local mean removal and divisive normalization:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + 1} \quad (40)$$

(New)

where

$$i \in \{1, 2, \dots, M\} \text{ and } j \in \{1, 2, \dots, N\} \quad \text{are spatial indices}$$

M and N are the image dimensions

$$\mu(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} I(i+k, j+l) \quad (41)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} [I(i+k, j+l) - \mu(i, j)]^2} \quad (42)$$

(New)

Estimate the local mean and contrast, respectively, where $w = \{w_k | k = -K, \dots, K, l = -L, \dots, L\}$ is a 2D circularly symmetric Gaussian weighting function sampled out to 3 standard deviations ($K = L = 3$) and rescaled to unit volume.

5.6.2 Patch selection

Once the image coefficients of Equation (40) are computed, the image is partitioned into $P \times P$ patches. Specific NSS features are then computed from the coefficients of each patch. However, only a subset of the patches are used.

The variance field of Equation (42) has been largely ignored in the past in NSS-based image analysis, but it is a rich source of structural image information that can be used to quantify local image sharpness. Letting the $P \times P$ -sized patches be indexed $b = 1, 2, \dots, B$, a direct approach is to compute the average local deviation field of each patch indexed b :

$$\delta(b) = \sum_{(i,j) \in \text{patch}b} \sigma(i,j) \quad (43)$$

(New)

where

δ denotes local activity/sharpness

Once the sharpness of each patch is found, those having a suprathreshold sharpness $\delta > T$ are selected. The threshold T is picked to be a fraction p of the peak patch sharpness over the image. In our experiments, we used the nominal value $p = 0.75$.

5.6.3 Characterizing image patches

The generalized Gaussian distribution (GGD) with zero mean is given by:

$$f(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \quad (44)$$

(New)

where

$\Gamma(\cdot)$ is the gamma function $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt, a > 0$

The parameters of the GGD (α, β) , can be reliably estimated using the moment-matching based approach proposed. The signs of the transformed image coefficients of Equation (40) have been observed to follow a fairly regular structure. However, distortions disturb this correlation structure (see Mittal, Moorthy, and Bovik [B25]). This deviation can be captured by analyzing the sample distribution of the products of pairs of adjacent coefficients computed along horizontal, vertical and diagonal orientations: $\hat{I}(i, j)\hat{I}(i, j+1)$,

$$\hat{I}(i, j)\hat{I}(i+1, j), \quad \hat{I}(i, j)\hat{I}(i+1, j+1) \quad \text{and} \quad \hat{I}(i, j)\hat{I}(i+1, j-1) \quad \text{for } i \in \{1, 2 \dots M\} \quad \text{and} \quad j \in \{1, 2 \dots N\}.$$

The products of neighboring coefficients are well-modeled as following a zero-mode asymmetric generalized Gaussian distribution (AGGD):

$$f(x; \gamma, \beta_l, \beta_r) = \begin{cases} \frac{\gamma}{(\beta_l + \beta_r) \Gamma\left(\frac{1}{\gamma}\right) \exp\left(-\left(\frac{-x}{\beta_l}\right)^\gamma\right)} \forall x \leq 0 \\ \frac{\gamma}{(\beta_l + \beta_r) \Gamma\left(\frac{1}{\gamma}\right) \exp\left(-\left(\frac{-x}{\beta_r}\right)^\gamma\right)} \forall x \geq 0 \end{cases} \quad (45)$$

(New)

The parameters of the AGGD $(\gamma, \beta_l, \beta_r)$ can be efficiently estimated using the moment-matching based approach. The mean of the distribution is also useful:

$$\eta = (\beta_r - \beta_l) \frac{\Gamma\left(\frac{2}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma}\right)} \quad (46)$$

(New)

By extracting estimates along the four orientations, 16 parameters are arrived at yielding 18 overall. All features are computed at two scales to capture multiscale behavior, by low pass filtering and downsampling by a factor of 2, yielding a set of 36 features.