IEEE STANDARDS ASSOCIATION

IEEE Standard for Quality of Experience (QoE) and Visual-Comfort Assessments of Three-Dimensional (3D) Contents Based on Psychophysical Studies

IEEE Computer Society

Sponsored by the
Standards Activities Board

IEEE
3 Park Avenue
New York, NY 10016-5997
USA

IEEE Std 3333.1.1™-2015

IEEE Standard for Quality of Experience (QoE) and Visual-Comfort Assessments of Three-Dimensional (3D) Contents Based on Psychophysical Studies

Sponsor

Standards Activities Board
of the
IEEE Computer Society

Approved 26 March 2015

IEEE-SA Standards Board

Abstract: As the demand and supply for 3D technologies grows, the development of accurate quality-assessment techniques shall be used to develop the 3D display device and signal-processing engine industries. The underlying principles and statistical characteristics of 3D contents based on the human visual system (HVS) are described in this standard. In addition, a reliable 3D subjective assessment methodology that covers the characteristics of human perception, display mechanism, and the viewing environment is introduced in this standard.

Keywords: accommodation and vergence conflict, foveation, human visual system, HVS, IEEE 3333.1.1™, QoE, quality assessment, quality of experience, saliency detection, stereoscopic, stereoscopic display, subjective assessment, visual (dis)comfort, visual contents analysis

---

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

PDF: ISBN 978-0-7381-9773-9 STD20263
Print: ISBN 978-0-7381-9774-6 STDPD20263

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page, appear in all standards and may be found under the heading "Important Notice" or "Important Notices and Disclaimers Concerning IEEE Standards Documents."

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

IEEE Standards documents (standards, recommended practices, and guides), both full-use and trial-use, are developed within IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association ("IEEE-SA") Standards Board. IEEE ("the Institute") develops its standards through a consensus development process, approved by the American National Standards Institute ("ANSI"), which brings together volunteers representing varied viewpoints and interests to achieve the final product. Volunteers are not necessarily members of the Institute 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 does not warrant or represent the accuracy or content of the material contained in its standards, and expressly disclaims all warranties (express, implied and statutory) not included in this or any other document relating to the standard, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; and quality, accuracy, effectiveness, currency, or completeness of material. In addition, IEEE disclaims any and all conditions relating to: results; and workmanlike effort. 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 his or her 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: PROCUREMENT OF 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 development process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English version published by IEEE should be considered the approved IEEE standard.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE-SA Standards Board Operations Manual 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. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that his or her views should be considered the personal views of that individual rather than the formal position of IEEE.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE. However, IEEE does not provide 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 Standards Coordinating Committees 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 revisions to an IEEE standard is welcome to join the relevant IEEE working group.

Comments on standards should be submitted to the following address:

Secretary, IEEE-SA Standards Board
445 Hoes Lane
Piscataway, NJ 08854 USA

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 imply 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.

Copyrights

Photocopies

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 ten years. When a document is more than ten 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 the IEEE-SA Website at http://ieeexplore.ieee.org/xpl/standards.jsp or contact IEEE at the address listed previously. For more information about the IEEE-SA or IEEE's standards development process, visit the IEEE-SA Website at http://standards.ieee.org.

Errata

Errata, if any, for all IEEE standards can be accessed on the IEEE-SA Website at the following URL: http://standards.ieee.org/findstds/errata/index.html. Users are encouraged to check this URL for errata periodically.

Patents

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 http://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.

Participants

At the time this IEEE standard was completed, the 3D Human Factors Working Group had the following membership:

Sanghoon Lee, *Chair*

Sungho Cho
Ricky Christanto
Seunghan Han
Huychan Jung
Sangkwon Jung
Seongyong Kim
Taewan Kim
JaeKeun Kwak
Patrick Le Callet
Jaeun Lee
Jongmuk Lee
Sanggeun Lee
Jing Li
An Ping

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

Ping An
Susan Burgess
Juan Carreon
Keith Chow
P. Eastman
Randall Groves
Werner Hoelzl
Noriyuki Ikeuchi
Akio Iso
Piotr Karocki
Jongyoo Kim
JaeKeun Kwak
Sanghoon Lee
Young Lae Moon
Charles Ngethe
Annette Reilly
Eugene Stoudenmire
Marcy Stutzman
Min Zhang
Daidi Zhong

When the IEEE-SA Standards Board approved this standard on 26 March 2015, it had the following membership:

John Kulick, *Chair*
Jon Walter Rosdahl, *Vice Chair*
Richard H. Hulett, *Past Chair*
Konstantinos Karachalios, *Secretary*

Masayuki Ariyoshi
Ted Burse
Stephen Dukes
Jean-Philippe Faure
J. Travis Griffith
Gary Hoffman
Michael Janezic
Joseph L. Koepfinger*
David J. Law
Hung Ling
Andrew Myles
T. W. Olsen
Glenn Parsons
Ronald C. Petersen
Annette D. Reilly
Stephen J. Shellhammer
Adrian P. Stephens
Yatin Trivedi
Phillip Winston
Don Wright
Yu Yuan
Daidi Zhong

*Member Emeritus

Patrick Gibbons
IEEE-SA Content Production and Management

Soo Kim
IEEE-SA Operational Program Management

Introduction

This introduction is not part of IEEE Std 3333.1.1™-2015, IEEE Standard for Quality of Experience (QoE) and Visual-Comfort Assessments of Three-Dimensional (3D) Contents Based on Psychophysical Studies.

This standard establishes methods of visual discomfort and quality of experience (QoE) assessments of three-dimensional (3D) contents based on psychophysical studies. These key factors are constructed in conjunction with the visual factors used to quantify discomfort and QoE degradation.

Contents

IEEE Standard for Quality of Experience (QoE) and Visual-Comfort Assessments of Three-Dimensional (3D) Contents Based on Psychophysical Studies

IMPORTANT NOTICE: IEEE Standards documents are not intended to ensure safety, security, health, or environmental protection, or ensure against interference with or from other devices or networks. Implementers of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, and interference protection practices and all applicable laws and regulations.

This IEEE document is made available for use subject to important notices and legal disclaimers. These notices and disclaimers appear in all publications containing this document and may be found under the heading "Important Notice" or "Important Notices and Disclaimers Concerning IEEE Documents." They can also be obtained on request from IEEE or viewed at http://standards.ieee.org/IPR/disclaimers.html.

1. Scope

This standard establishes methods for visual saliency prediction, visual contents analysis, and subjective assessment for quantifying the visual discomfort and quality of experience (QoE) of three-dimensional (3D) image and video.

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.

ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures. [1]

[1] ITU publications are available from Publication Sales, International Telecommunication Union, Place des Nations, 1211 Geneva 20, Switzerland (http://www.itu.int/).

ITU-R BT.1438: Subjective assessment of stereoscopic television pictures.

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

ITU-T P.910: Subjective video quality assessment methods for multimedia applications.

Steroscopic (3D Imaging) Database.[2]

3. Definitions, abbreviations, and acronyms

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. [3]

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.

angular disparity: The angle differences in image location of an object resulting from the eye movements by both human eyes.

auto-stereoscopic display: A display device capable of providing depth perception to human viewers without the use of glasses.

ciliary muscle: The ring of striated smooth muscles in the human eye's vascular layer that controls accommodation according to viewing distance.

comfortable viewing zone: The range of object or viewer position where a viewer can watch a scene without any discomfort.

crosstalk: The incomplete isolation of the left and right image that causes one image leak into the other.

crystalline lens: The lens of a human eye composed of fibers that come from epithelial cells.

diplopia: The subjective complaint of seeing two images when viewing one object.

disparity contrast: The absolute difference between values in a target disparity value with neighbor disparity values within the same window.

disparity gradient: The absolute values of the derivate to the disparity map on the x and y axes.

extra-ocular muscles: The six muscles (in each eye) that are expanded or contracted for eye movements.

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

horopter: The locus of points in space that yield single vision.

[2] Available at http://grouper.ieee.org/groups/3dhf
[3] *IEEE Standards Dictionary Online* subscription is available at:
http://www.ieee.org/portal/innovate/products/standard/standards_dictionary.html.

human visual system: The part of the central nervous system that gives organisms the ability to process visual detail and enables the formation of several non-image photo response functions.

lab color space: Color-opponent space with dimension lightness and two-color opponent dimensions (a and b) based on nonlinearly compressed CIE xyz color-space coordinates .

luminance contrast: The absolute difference between values in a target luminance value with neighbor luminance values within the same window.

luminance gradient: The absolute values of the derivate to the disparity map on the x and y axes.

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

negative disparity: When the angular disparity value is smaller than zero because the corresponding object is located in front of the 3D display.

Panum's fusional area: The small region on either side of the horopter where objects will still be seen as fused. This region in space corresponding to an area on the retina is Panum's fusional area or Panum's area.

photoreceptor: The retinal cells responsible for the detection of light.

positive disparity: When the angular disparity value is larger than zero because the corresponding object is located behind the 3D display.

probability mass function: In statistics, this is the function that gives the probability that a discrete random variable is exactly equal to some value.

quality assessment: The evaluation of a 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 an application or service resulting from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in light of the user's personality and current state.

retina: The sensitive layer of tissue at the back of the eye that receives sensory images.

shear distortion: A different kind of distortion caused when a viewer moves.

skewness: In statistics, this is the degree to which a statistical distribution is not in equilibrium around the average, an exactly symmetrical distribution having a value of zero.

stereoscopic display: A device capable of experiencing depth perception to 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.

support vector machine: In machine learning, this is a supervised learning model with associated learning algorithms.

system crosstalk: The degree of the leaking image from the other image that is determined by the display itself.

vergence: Simultaneous movements of both eyes in opposite directions to obtain or maintain single binocular vision clearly.

viewing distance: The distance between the viewer and display.

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

visual cortex: The part of the cerebral cortex responsible for processing visual information.

visual fatigue: A decrease in performance of the human vision system due to a physiological change. This change shall be diagnosed by medical personnel using various symptoms that can either be subjectively reported by the subject or objectively measured. Visual fatigue is usually considered an effect that declines more slowly after removal of the uncomfortable stimulus.

visual saliency: The state or quality by which interested regions on image or video stand out relative to the other regions.

wavelet transform: In mathematics, this is a representation of a square-integral function by a certain orthonormal series generated by a wavelet.

3.2 Abbreviations and acronyms

2D	two dimensional
3D	three dimensional
ARD	adaptive rectangular design
ASD	adaptive square design
CIE	International Commission on Illumination
CVZ	comfortable viewing zone
FPC	full paired comparison
GGD	generalized Gaussian distribution
HRC	hypothetical reference circuit
HVS	human visual system
MICSQ	multimodal interactive continuous scoring of quality
ORD	optimized rectangular design
OSD	optimized square design
PLCC	Pearson linear correlation coefficient

PMF	probability mass function
QoE	quality of experience
SRC	source video contents
SROCC	Spearman's rank correlation coefficient
SSCQE	single stimulus continuous quality evaluation
SVR	support vector regression

4. Ergonomic requirements and recommendations

4.1 General

Unlike in two-dimensional (2D) video, the ocular adjustment to 3D depth can induce neurological symptoms, such as visual discomfort and headache, and 3D distortions that cause quality degradation. Understanding these problems involves several intricate visual factors that can only be probed by investigating the reliable spatial and temporal features in 3D contents and by using a reliable subjective testing methodology.

4.2 Visual saliency prediction

When viewing a 3D image and video, some scenes have only a few regions that cause visual discomfort and low QoE states. In these cases, the viewer tends to pay more attention to this region, and the rest of the image is projected onto the retina with a low density of photoreceptors. Therefore, to develop reliable human factor models and a quality assessment metric, a concrete visual saliency algorithm shall be developed.

4.3 Visual contents analysis

To analyze and predict the degree of the QoE or visual comfort when viewing 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 3D visual perception.

4.4 Subjective assessment

Most subjective assessment is inherited from what has been traditionally done for 2D subjective assessment as defined by ITU-T P.910, ITU-R BT.500-13, and ITU-R BT.2021. However, it is doubtful whether these results are reliable enough to be used as references because the viewing environment is quite different from 2D due to the intensive immersion of a user wearing the glasses in a the dark. Hence, to perform the subjective 3D image and video assessments, a novel interface shall be designed that covers the characteristics of the human perception, display mechanism, viewing environment, and so on.

5. Visual saliency prediction

5.1 General

An important ingredient in further improving 3D video processing technologies is the effort to incorporate better models of 3D perception. Among these, saliency detection, or the automated discovery of points of high visual interest, conspicuity, or task relevance, is a challenging problem. This clause describes visual saliency prediction method by considering human visual system (HVS) characteristics.

5.2 Human visual system

To capture human factors or predict the visual discomfort of 3D contents, the characteristics of the HVS such as a retina and fovea shall be considered.

The fovea is responsible for sharp central vision, which is needed in human beings for reading, watching television or movies, driving, and any activity for which visual detail is of primary importance. As shown in Figure 1, photo-receptors possess non-uniform spatial distribution with the highest density at the fovea, and the density decreases dramatically with distance from the fovea. Here we describe how the region with the highest saliency (obtained as described above) falls on the fovea and, hence, has the highest sensitivity/resolution.

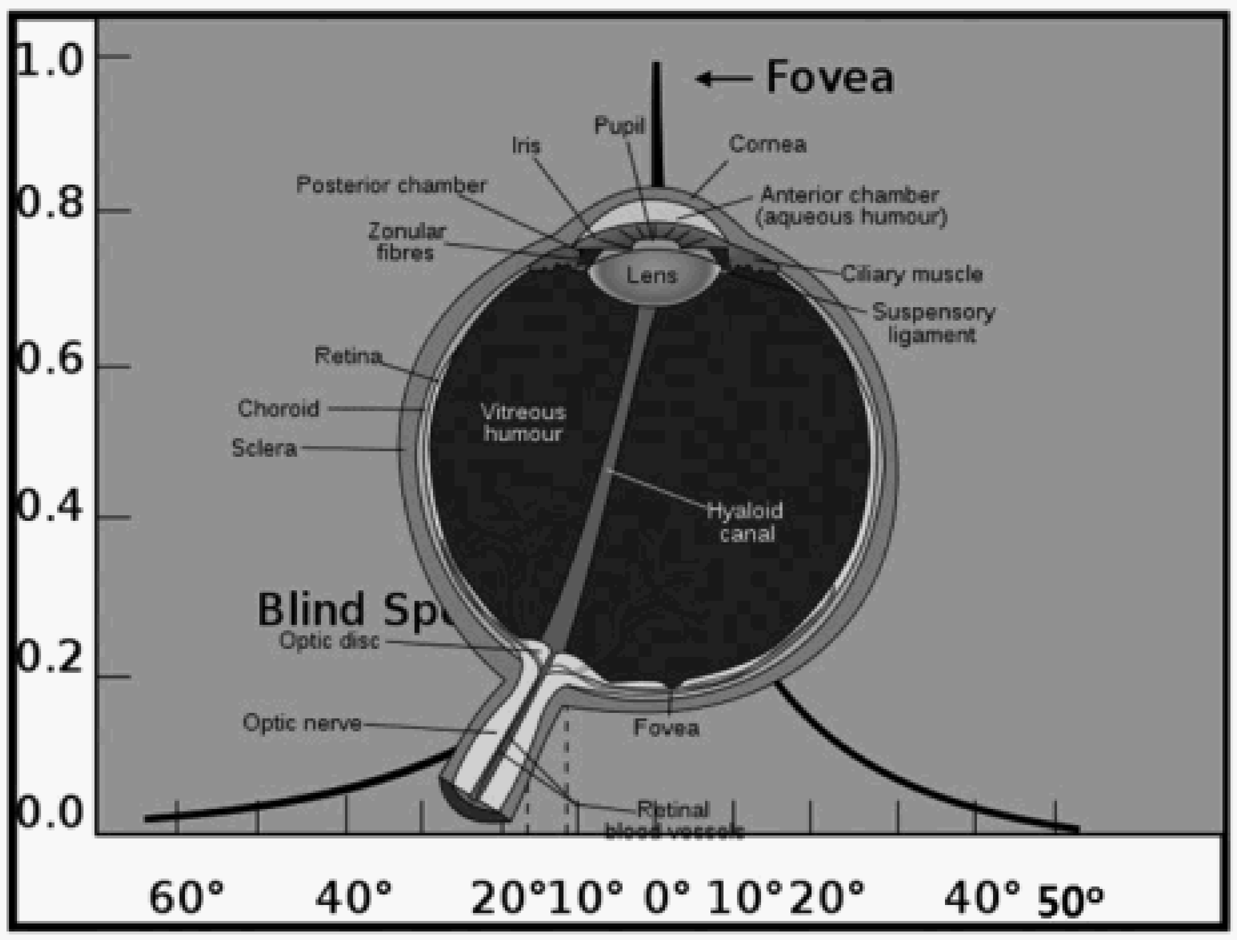


Figure 1—Eye structure as a function of eccentricity

The foveation model used in this standard shall be represented as

$$f\{V,(x,y)\} = \min\left[\frac{e\ln\left(\frac{1}{CT_0}\right)}{\alpha\left\{e+\arctan\left(\frac{d}{VW}\right)\right\}}, \frac{\pi VW}{360}\right] \tag{1}$$

where

d	is the pixel distance between point (x, y) and fixation point (x_f, y_f)
m	is the screen magnification factor
$\arctan\left(\frac{d}{VW}\right)$	is the eccentricity
W	is the width of the display
V	is the viewing distance

The fitting parameters in this model have been estimated yielding the following:

- a spatial frequency decay constant $\alpha = 0.106$
- half-resolution eccentricity constant $e = 2.3$
- a minimum contrast threshold $\frac{1}{CT_0} = \frac{1}{64}$.

5.3 Saliency prediction

5.3.1 2D saliency features

This standard considers four 2D visual saliency features: luminance, color, size, and compactness.

The luminance contrast and gradient of stereoscopically fixated patches are generally higher than in randomly selected patches. Thus these features shall be represented as

$$\begin{cases} W_{lc}^{(x,y)} = 1, \text{if } \frac{C_l^{(x,y)}}{\hat{C}_l} > 1 \\ W_{lc}^{(x,y)} = \frac{C_l^{(x,y)}}{\hat{C}_l}, \text{otherwise} \end{cases} \tag{2}$$

$$\begin{cases} W_{\lg}^{(x,y)} = 1, \text{if } \frac{G_l^{(x,y)}}{\hat{G}_l} > 1 \\ W_{\lg}^{(x,y)} = \frac{G_l^{(x,y)}}{\hat{G}_l}, \text{otherwise} \end{cases} \tag{3}$$

where

$C_l^{(x,y)}$ ($G_l^{(x,y)}$) is the luminance contrast (gradient) at (x, y) in the original frame

$\hat{C}_l$ ($\hat{G}_l$) is the mean of the luminance contrast (gradient) of the original frame

$W_{lc}^{(x,y)}$ ($W_{\lg}^{(x,y)}$) is the luminance contrast (gradient) saliency weight at (x, y) in the original frame

Using these saliency weights, the luminance saliency weight shall be defined by

$$W_l^{(x,y)} = w_{lc} W_{lc}^{(x,y)} + w_{\mathrm{lg}} W_{\mathrm{lg}}^{(x,y)} \tag{4}$$

where

$W_l^{(x,y)}$ is the luminance saliency weight at (x, y) in the original frame.

There is a correlation between luminance gradient and luminance contrast; these two values can be combined by weighted sum. The weights of the luminance gradient and luminance contrast are w_{lc} and w_{lg}, respectively. They increase (decrease) when the location of the original frame contains higher (lower) luminance contrast and gradient values. Figure 2 depicts the visual saliency according to luminance.

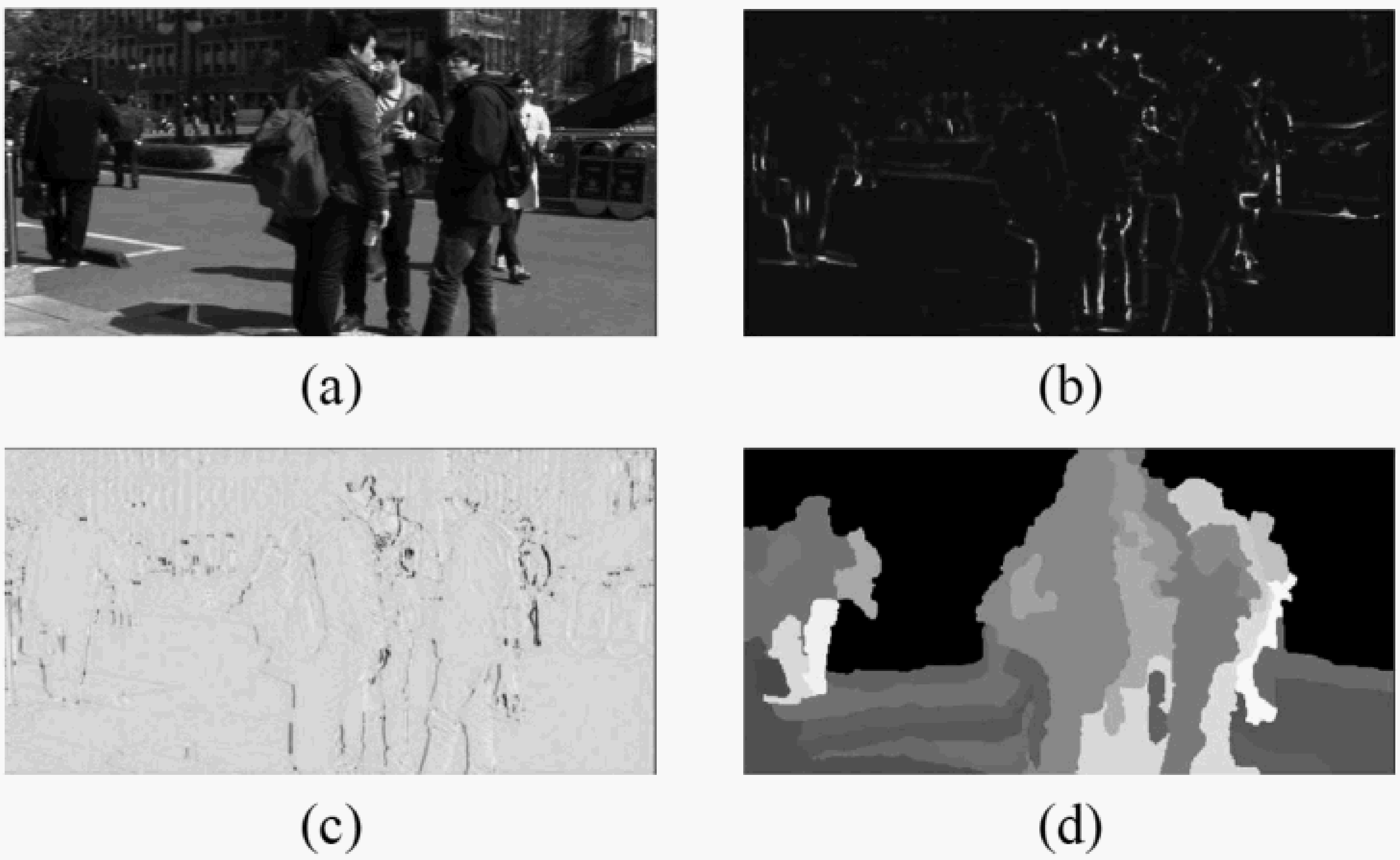


Figure 2—Luminance feature for visual saliency: (a) original left image, (b) luminance contrast map, (c) luminance gradient map, and (d) saliency map

Color information in the 3D image/video is highly related to the visual salient region. Accordingly, the color contrast and gradient features are investigated to capture the local rate of color changes and calculate the visual saliency according to color. In addition, these are computed after converting the images into the perceptually uniform CIE-Lab color space. The visual saliency according to color shall be represented as

$$\begin{cases} W_{cc}^{(x,y)} = 1, \text{ if } \dfrac{C_c^{(x,y)}}{\hat{C}_c} > 1 \\ W_{cc}^{(x,y)} = \dfrac{C_c^{(x,y)}}{\hat{C}_c}, \text{otherwise} \end{cases} \tag{5}$$

$$\begin{cases} W_{cg}^{(x,y)} = 1, \text{if } \dfrac{G_c^{(x,y)}}{\hat{G}_c} > 1 \\ W_{cg}^{(x,y)} = \dfrac{G_c^{(x,y)}}{\hat{G}_c}, \text{otherwise} \end{cases} \tag{6}$$

where

$C_c^{(x,y)}$ ($G_c^{(x,y)}$) is the color contrast (gradient) at (x, y) in the original frame
$\hat{C}_c$ ($\hat{G}_c$) is the mean of the color contrast (gradient) of the original frame
$W_{cc}^{(x,y)}$ ($W_{cg}^{(x,y)}$) is the color contrast (gradient) saliency weight at (x, y) in the original frame

Using these saliency weights, the color saliency weight shall be defined

$$W_c^{(x,y)} = w_{cc} W_{cc}^{(x,y)} + w_{cg} W_{cg}^{(x,y)} \tag{7}$$

where

$W_c^{(x,y)}$ is the color saliency weight at (x, y) in the original frame.

There is a correlation between color gradient and color contrast; these two values can be combined by weighted sum. The weights of the color gradient and color contrast are w_{cc} and w_{cg} , respectively. They increase (decrease) when the location of the original frame contains higher (lower) color contrast and gradient values.

Finally, the relationship between visual saliency and the size and compactness of objects are described in this standard. Generally, humans perceive, recognize, and more frequently fixate larger (and often nearer) objects on the image/videos. Based on this observation, this standard defines the relative object size as the ratio of the number of pixels of an object to the number of pixels in the image and introduces a simple threshold value on suitable object size. Moreover, humans adapt to variations in wider objects more easily than narrower objects. Based on this observation, this standard also defines the compactness of a salient region as the mean distance between the central point and other points inside the salient region. Figure 3 shows the visual saliency according to compactness.

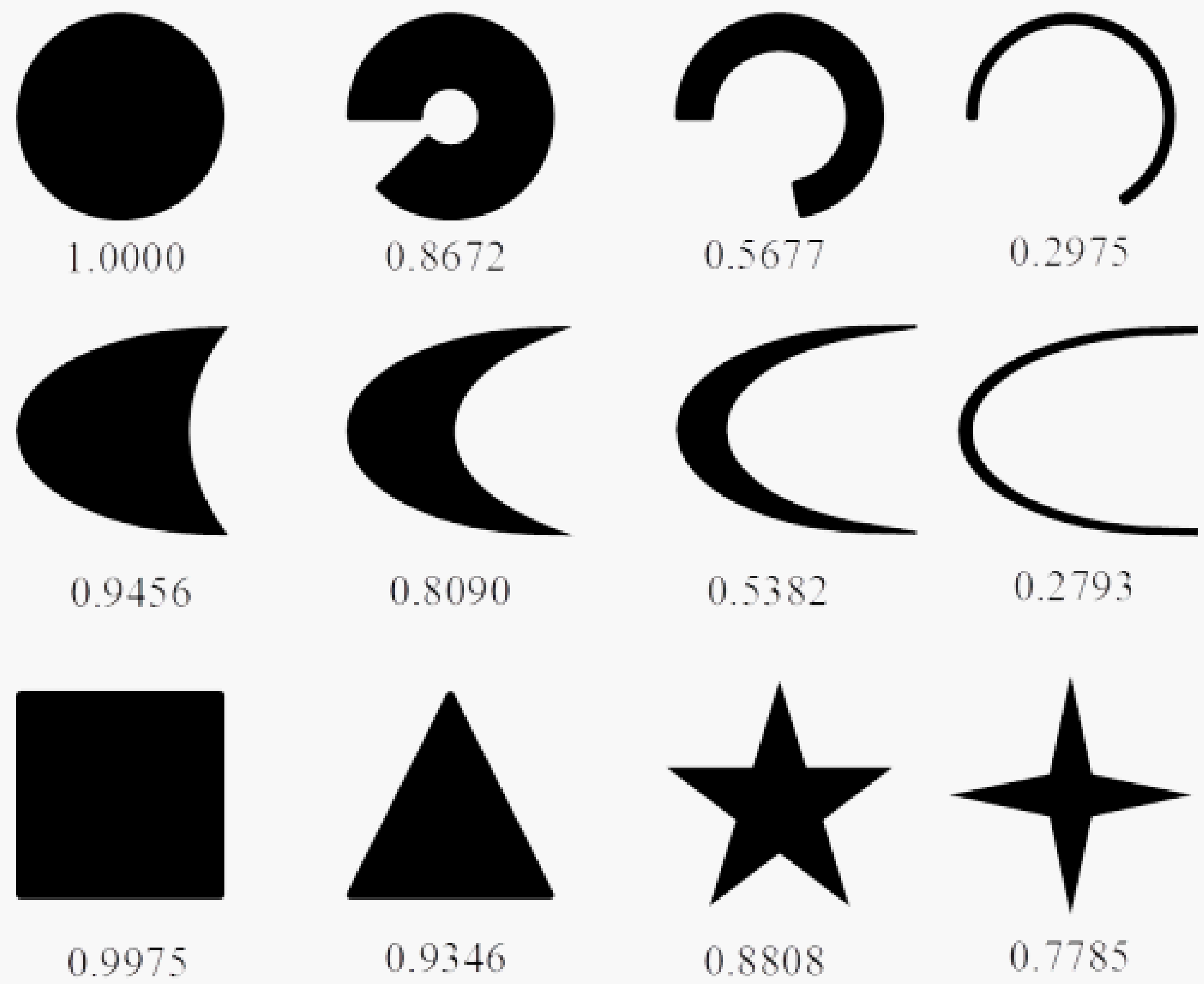


Figure 3—Compactness feature of visual saliency

5.3.2 3D saliency features

Two 3D saliency features shall be considered: depth discontinuities and visual discomfort.

Depth discontinuities are important factors in defining visual saliency. In particular, the disparity contrast and gradient are highly related to 3D visual fixations. Fixated disparity contrast and gradient are generally lower than random location.

In addition, viewers tend to fixate away from large disparity gradients and contrast, preferring instead smooth depth regions.

These attentional characteristics as part of visual saliency according to depth discontinuities shall be measured as follows:

$$\begin{cases} W_{dc}^{(x,y)} = 1, \text{if } \dfrac{C_d^{(x,y)}}{\hat{C}_d} > 1 \\ W_{lc}^{(x,y)} = \dfrac{C_d^{(x,y)}}{\hat{C}_d}, \text{otherwise} \end{cases} \tag{8}$$

$$\begin{cases} W_{dg}^{(x,y)} = 1, \text{if } \dfrac{G_d^{(x,y)}}{\hat{G}_d} > 1 \\ W_{dg}^{(x,y)} = \dfrac{G_d^{(x,y)}}{\hat{G}_d}, \text{otherwise} \end{cases} \tag{9}$$

where

$C_d^{(x,y)}$ ($G_d^{(x,y)}$) is the disparity contrast (gradient) at (x, y) in the original frame

$\hat{C}_d$ ($\hat{G}_d$) is the mean of the disparity contrast (gradient) of the original frame

$W_{dc}^{(x,y)}$ ($W_{dg}^{(x,y)}$) is the disparity contrast (gradient) saliency weight at (x, y) in the original frame

Using above saliency weights, the depth discontinuity saliency weight shall be defined by

$$W_{dd}^{(x,y)} = w_{dc} W_{dc}^{(x,y)} + w_{dg} W_{dg}^{(x,y)} \tag{10}$$

where

$W_{dd}^{(x,y)}$ is the depth discontinuity saliency weight at (x, y) in the original frame.

There is a correlation between disparity gradient and luminance contrast; these two values may be combined by weighted sum. The weights of the disparity gradient and luminance contrast are w_{dc} and w_{dg}, respectively. They decrease (increase) when the location of the original frame contains higher (lower) disparity contrast and gradient values. Figure 4 depicts the visual saliency according to depth discontinuities.

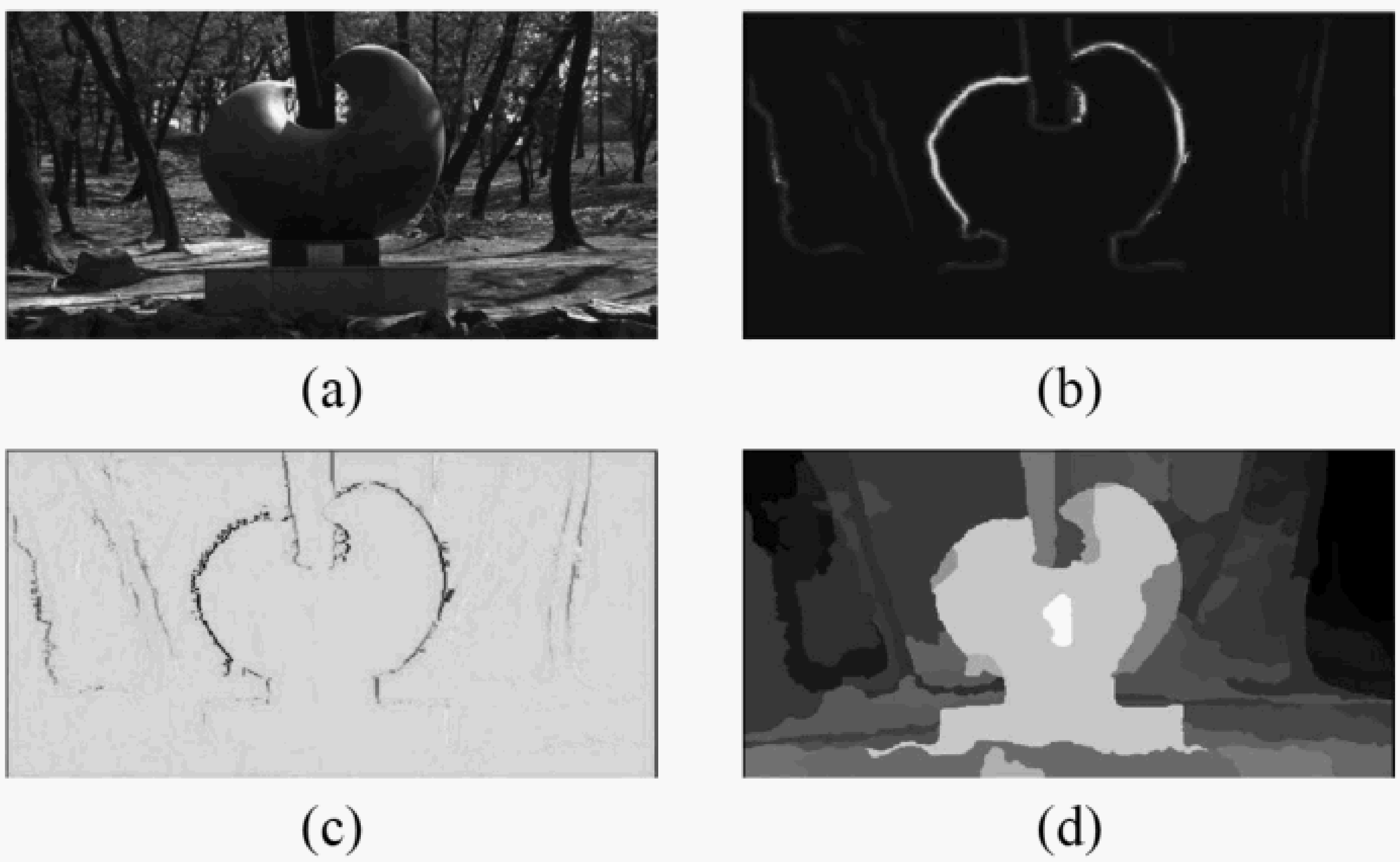


Figure 4—Depth discontinuities for visual saliency: (a) original image, (b) disparity contrast map, (c) disparity gradient map, (d) and saliency map

Disparity of an object and its size affect whether fixating on it causes visual discomfort when viewing 3D content. Human fixations tended to land on salient objects when disparity lies within a comfortable viewing zone (CVZ) (within –1 in ~ 1 in of angular disparity). By contrast, when disparity does not lie within a CVZ, the fixation distribution becomes diffused. Using these attention characteristics, the 3D saliency features according to visual discomfort in terms of disparity and a CVZ shall be measured as

$$\begin{cases} W_{vd}^{(x,y)} = \dfrac{-1}{D^{(x,y)}} \times \eta_D, \text{if } D^{(x,y)} \leq -1^\circ \\ W_{vd}^{(x,y)} = \eta_D, \quad \text{if } -1^\circ < D^{(x,y)} \leq 1^\circ \\ W_{vd}^{(x,y)} = \dfrac{1}{D^{(x,y)}} \times \eta_D, \ \text{if } D^{(x,y)} \geq 1^\circ \end{cases} \tag{11}$$

$$\eta_D = 1 - \frac{D^{(x,y)} - D_N^M}{D_F^M - D_N^M} \tag{12}$$

where

$W_{vd}^{(x,y)}$ is the visual discomfort saliency weight expressed as a function of the angular disparity $D^{(x,y)}$ at (x, y) in the original frame.

Denote η_D as an index $(0 \leq \eta_D \leq 1)$ on D_N^M (D_F^M), which is the nearest (farthest) angular disparity of the original frame $\left(D_N^M \leq D^{(x,y)} \leq D_F^M\right)$. The value of $W_{vd}^{(x,y)}$ takes its maximum value when $D^{(x,y)}$ is in the CVZ. However, if the disparities are out of the CVZ, $W_{vd}^{(x,y)}$ is decreased in proportion to the magnitude of disparity. Figure 5 shows the visual saliency for various disparity cases.

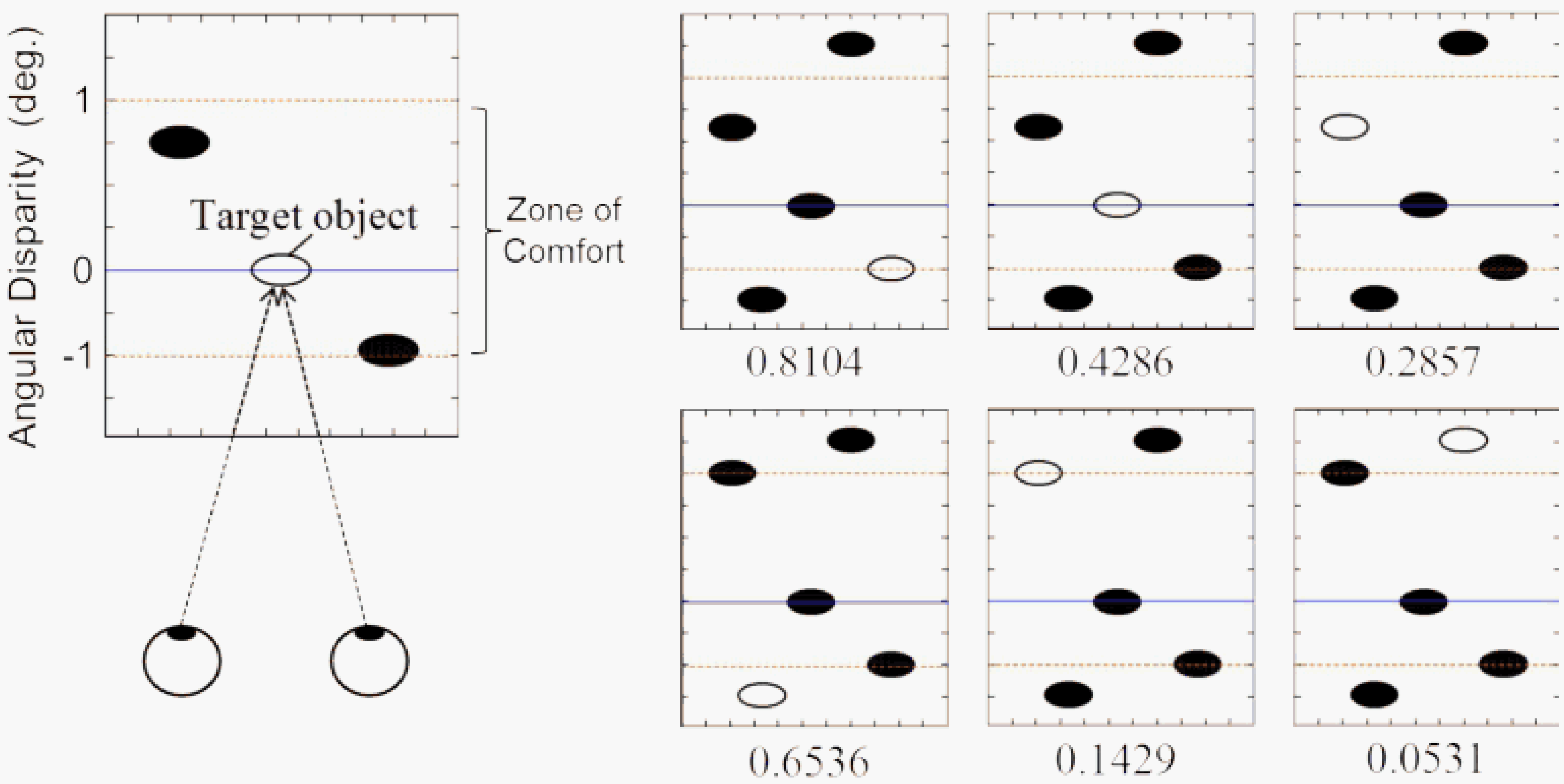


Figure 5—Visual saliency for various disparity cases (top-down views)

5.3.3 Visual saliency prediction map

The weighted sum method of the 2D features and 3D features shall be used in order to calculate the final visual saliency prediction map as

$$R_S^{(x,y)} = w_{2D-l} W_l^{(x,y)} + w_{2D-c} W_c^{(x,y)} + w_{3D-dd} W_{dd}^{(x,y)} + w_{3D-vd} W_{vd}^{(x,y)} \quad (13)$$

where

w_{2D-l}	is the weight of luminance feature
w_{2D-c}	is the weight of color feature
w_{3D-dd}	is the weight of depth discontinuities,
w_{3D-vd}	is the weight of visual discomfort as the 3D saliency feature

Using Equation (13), the final visual saliency prediction map $R_S^{(x,y)}$ $\left(0 \le R_S^{(x,y)} \le 1\right)$ shall be procured. Figure 6 shows the final saliency prediction map. Regions having low saliency weight, such as roads, ground, or sky, are excluded from the visual saliency prediction maps, as shown in Figure 6. In addition, objects that have high luminance/color gradient or contrast have relatively large weights, and the objects that are wide and dense with few or no depth discontinuities tend to have higher saliency weights. The eye-tracking results are closely similar to the final visual saliency prediction maps shown in Figure 6.

(a) (b) (c) (d) (e) (f)

Figure 6—Visual saliency prediction maps
(a) and (d): original images
(b) and (e): visual saliency prediction maps
(c) and (f) eye-tracker suits

6. Visual-contents analysis

6.1 General

Based on the human perception model, to predict the level of visual discomfort that would be experienced when viewing a 3D image, several important spatial features (excessive disparity, accommodation and vergence conflict, and distribution of disparity) derived from a statistical analysis of disparity are presented in this clause. Moreover, we develop a new framework of visual activity for quantitatively analyzing 3D video.

6.2 The human visual system

In order to place a fixated object on the corresponding retinal points by changing the gaze distance, vergence must involve the simultaneous rotation of both eyes in opposite directions. In a natural viewing environment, accommodation and vergence, the two important oculomotor mechanisms, exhibit dynamic properties while maintaining a mutual response via interoperation for comfortable viewing. However, when viewing a 3D image on a flat stereoscopic display, discrepancies occur between the degree of accommodation to achieve a sharp image for a given amount of vergence, which causes perceptual confusion and conflicts in the visual control system. Hence, mutual interactive processes between accommodation and vergence are investigated in Figure 7.

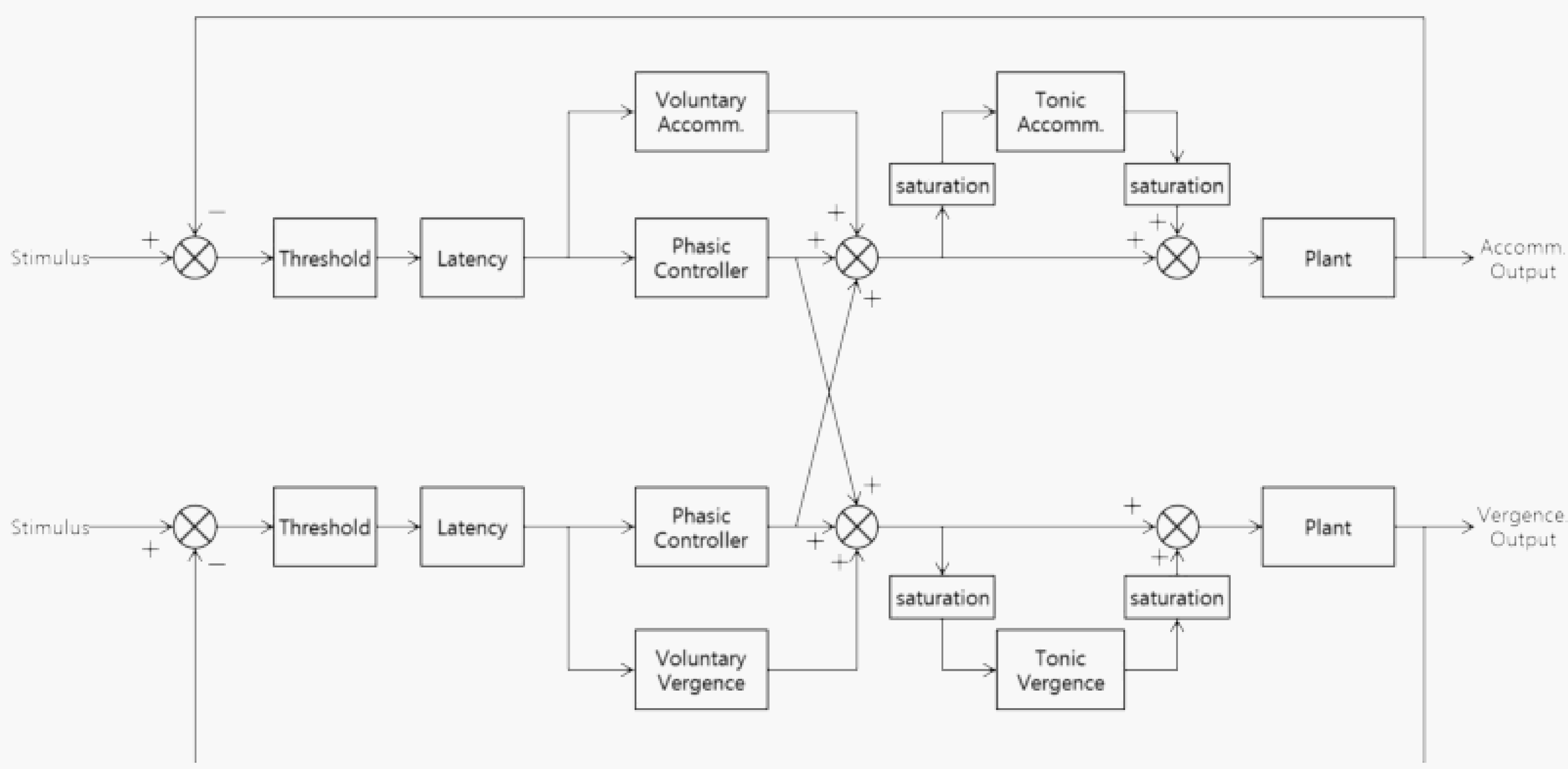
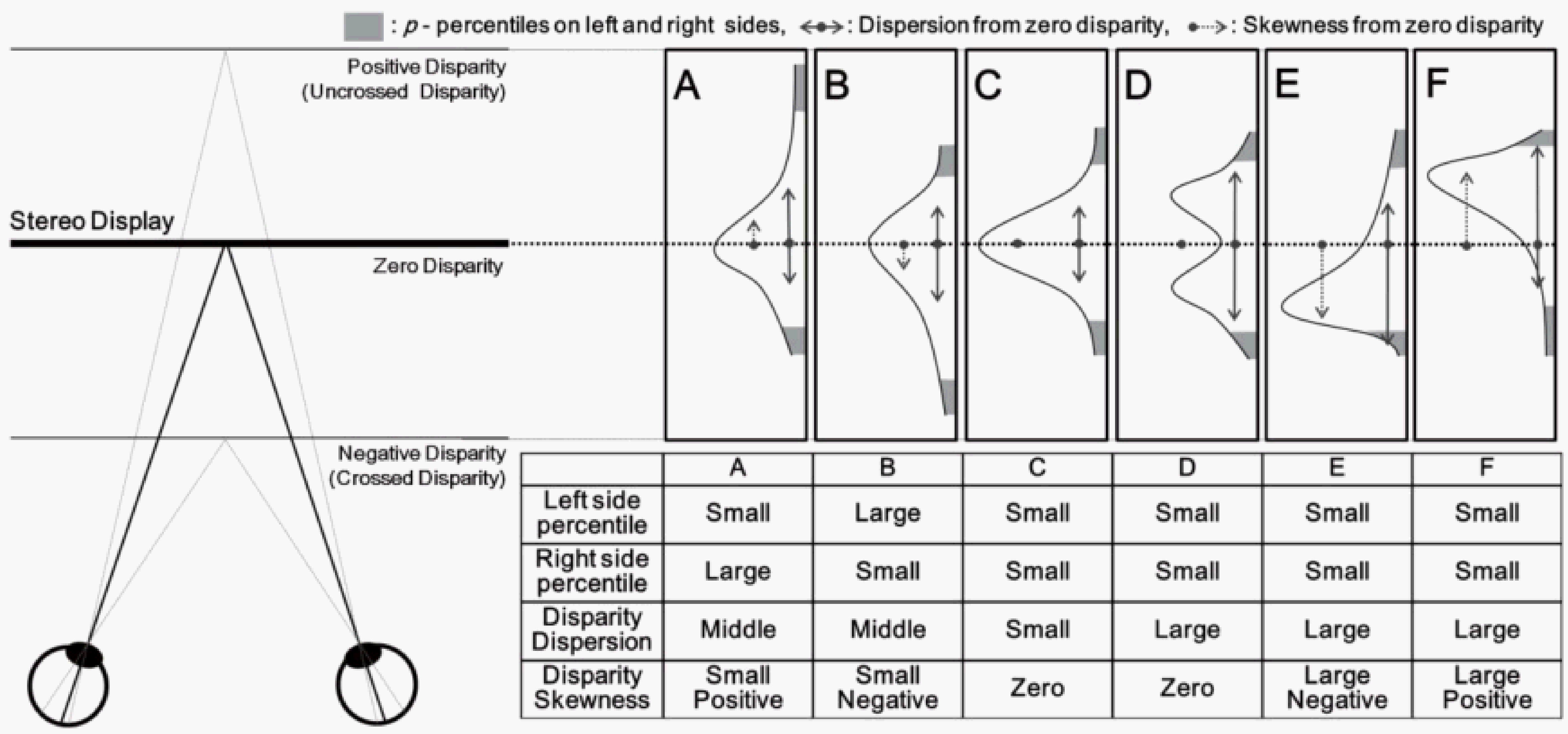


Figure 7—Dual interaction model with fast and slow components

Figure 7 shows that the interoperation between accommodation and vergence shall be modeled by two dual parallel feedback control systems where each system is composed of a controller, a plant, and related components. The controller, in the physiological sense, refers to the neurological mechanism that generates neurological signals from blur and disparity to drive accommodation and vergence. The controller generates an appropriate signal to drive the plant that produces the desired action. For accommodation, ciliary muscles and the crystalline lens shall be represented as the plant. For vergence, the lateral and medial recti of extraocular muscles and the eyeball shall be represented as the plant. The feedback loop shall allow the phasic response to be replaced by the adaptable tonic response. Therefore, when humans view a certain object for a very long time, the accommodation and vergence responses exhibit adaptation due to the tonic components.

6.3 Spatial characteristics of 3D contents

To investigate the spatial characteristic of 3D contents, this standard describes three spatial characteristics of visual discomfort when viewing a 3D scene: excessive disparity, accommodation and vergence conflict, and distribution of disparity. Based on the consideration of the spatial characteristics, several features from the statistics of the disparity map shall be extracted. A 3D image shall have zero, negative, or positive parallax at the points appearing on the screen itself, behind the screen, or in front of the screen, respectively. Disparity maps shall have a variety of distributions, for Examples A through F in Figure 8, which are obtained along the depth axis. The distributions exhibit the disparity characteristics of a 3D image well, so that the statistics of the distribution are able to be excellent features for the causative factors of visual discomfort.



	A	B	C	D	E	F
Left side percentile	Small	Large	Small	Small	Small	Small
Right side percentile	Large	Small	Small	Small	Small	Small
Disparity Dispersion	Middle	Middle	Small	Large	Large	Large
Disparity Skewness	Small Positive	Small Negative	Zero	Zero	Large Negative	Large Positive

Figure 8—Examples of disparity distribution and comparisons of statistical features

6.3.1 Excessive disparity

The excessive disparities may appear at one or both ends of distribution of disparity. Especially, in Figure 8, the distributions of Examples A and B have excessive disparities toward positive and negative, respectively. They suffer from more discomfort caused by excessive disparity than the others. The values of both ends of distribution shall be a good feature for the excessive disparity. Thus we shall adopt means of p^{th}-percentiles on the both left and right sides of the distribution, f_1 and f_2, respectively, as features:

$$f_1 = \frac{1}{d_{\max}} \cdot \left(\frac{1}{N_P^l} \sum_{n < N \cdot p/100} d(n) \right) \tag{14}$$

$$f_2 = \frac{1}{d_{\max}} \cdot \left(\frac{1}{N_P^r} \sum_{n > N \cdot (100-p)/100} d(n) \right) \tag{15}$$

where

- N is the total number of disparity values
- N_P^l is the numbers of p^{th}-percentile on left side (p shall be a small value, such as 5% or 10%)
- N_P^r is the numbers of p^{th}-percentile on right side (p shall be a small value, such as 5% or 10%)
- $d(n)$ is the n^{th} disparity among the disparity values sorted in rank order
- $d_{\max}$ is the maximum perceptible disparity

Mere minimum and maximum values are not adopted because the disparity map is not perfect and has errors resulting in extreme values. Hence, the means of p^{th}-percentile on both left and right sides, which represent the excessive disparity, shall be used to extract the excessive disparity features.

6.3.2 Accommodation and vergence conflict

The accommodation and vergence conflict occurs due to the deviation of distance intended for vergence eye movement from that for accommodation fixed on the screen. Most non-zero disparities compel vergence eye movement, which causes visual discomfort, but sufficiently small disparities within a range of depth of focus may not induce visual discomfort. There is a certain tendency that the more dispersive the distribution of disparity is from zero, the more the accommodation vergence conflict is likely to occur. Thus, for the feature of accommodation and vergence conflict, f_3, a modified standard deviation inserting zero instead of the mean shall be measured:

$$f_3 = \frac{1}{d_{\max}} \sqrt{\frac{1}{N} \sum_n d(n)^2} \tag{16}$$

where

if $f_3 > 1$, set $f_3 = 1$

The distributions of Examples C and D in Figure 8 have the similar means but totally different dispersion from the zero disparity, which implies that the 3D image of Example D shall induce more severe accommodation and vergence conflict than that of Example C.

6.3.3 Distribution of disparity

The distributions of Examples E and F in Figure 8 have similar dispersions but different skewness of the disparity distributions. The negative disparities tend to induce more visual discomfort than positive disparities. Hence, a modified skewness to deal with the influence of parallax distribution, f_4, shall be defined:

$$f_4 = \frac{\sum_n d(n)}{\sum_n |d(n)|} \tag{17}$$

If a distribution of disparity is more concentrated on the negative (or positive) side of zero disparity, f_4 becomes closer to –1 (or 1). The sign and magnitude of f_4 shall be an indicator of disparity skewness from zero disparity. In the case of Examples C and D in Figure 8, the disparities are symmetrically distributed around the zero disparity, so that f_4 becomes zero and there is no influence of disparity skewness.

6.3.4 Visual discomfort prediction based on spatial characteristics

The visual discomfort score shall be predicted by using a regression tool that maps the feature vectors (f_1, f_2, f_3, f_4) to the associated visual (dis)comfort score. The test and training sets are listed in the IEEE-SA stereo image database with the corresponding subjective mean opinion scores. To apply these features, a support vector regression (SVR) method may be used, which has been shown to perform well on high-dimensional regression problems and is successfully utilized in the conventional no-reference quality assessment methods, as shown in Figure 9.

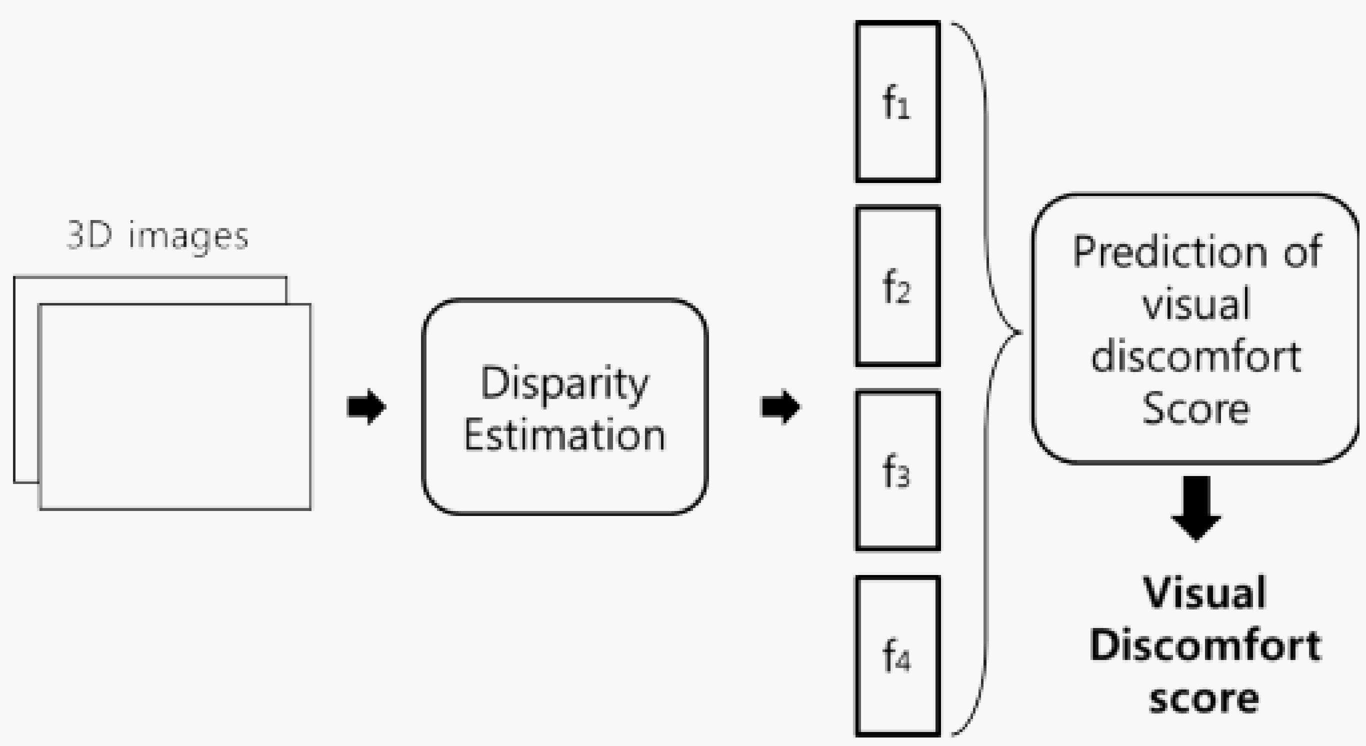


Figure 9—Framework of the visual discomfort predictor based on the spatial characteristics

6.4 Temporal characteristics of 3D contents

6.4.1 General

Once objects in 3D content are formed in the voluminous 3D space, the statistical distributions of those objects shall be measured with respect to the object disparity over the spatial domain and its movement over the temporal domain in the 3D space to analyze the temporal characteristics of 3D video.

6.4.2 Disparity activity

When large disparity variations occur in 3D video, neural metabolic rates are increased, leading to anomalous binocular percepts and consequent visual discomfort. Thus, disparity activity shall be defined as follows.

Let $\mathbf{M}_D$ denote the disparity map of the 3D scene so that each point of $\mathbf{M}_D$ corresponds to the depth projection of each voxel. Then

$$\mathbf{M}_D(t) = \{M_D(x_1, x_2, t) | 1 \le x_1 \le o_w, 1 \le x_2 \le o_h\} \tag{18}$$

where

$$M_D(x_1, x_2, t) = \arg\min_d |o_L(x_1, x_2, t) - o_R(x_1 + d, x_2, t)| \tag{19}$$

Here $\mathbf{M}_D(t)$ and $M_D(x_1, x_2, t)$ denote M_D at time t and its pixel value at (x_1, x_2), respectively. $o_L(x_1, x_2, t)$ and $o_R(x_1, x_2, t)$ denote pixel values at (x_1, x_2) at time t on the original left and right image, respectively; and o_w and o_h denote the width and height, of the original left (or right) image. In Equation (18), d is the disparity at pixel coordinate (x_1, x_2), which can be estimated using a suitable stereo matching algorithm.

As a way of analyzing the natural scene statistics, this standard counts the number of wavelet coefficients according to magnitude. After taking the wavelet transform, the probability mass function (PMF) of the wavelet coefficients histogram shall be counted. Let $\mathbf{p}_i(t)$ be the PMF of the i^{th} subband in time t. Let $\mathbf{B}$ be the set of bins of the histogram with regards to the wavelet coefficients, and $p_i(j,t)$ be the PMF of the j^{th} bin. Then, $\mathbf{p}_i(t)$ shall be written by

$$\mathbf{p}_i(t) = \{p_i(j,t) | \forall j \in \mathbf{B}\}. \tag{20}$$

If $\mathbf{W}_i(\cdot)$ denotes the wavelet coefficient matrix of the i^{th} subband over the wavelet domain, and $\overline{B}$ is its interval between bins, the values of wavelet coefficients $\mathbf{W}_i(\cdot)$ belonging to the j^{th} bin are in the range of $j - \overline{B}/2 \le W_i(M_M(t))(w,h) \le j + \overline{B}/2$. Using the definitions, $p_i(j,t)$ shall be represented as

$$p_i(j,t) = \sum_{h=1}^{i_h} \sum_{w=1}^{i_w} \Lambda_j(w,h,t) / i_h i_w \tag{21}$$

where

$$\Lambda_j(w,h,t) = \begin{cases} 1, & \text{if } j - \overline{B}/2 \le W_i(M_D(t))(w,h) \le j + \overline{B}/2 \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

where

i_w is the width and height of the i th subband
i_h is the width and height of the i th subband

Then, the wavelet coefficients shall be fitted into a statistical model, and its shape parameter shall be found. Denote the GGD fitting to $p_i(t)$ as

$$\overline{GGD}\left(p_i(t): \mu_i(t), \sigma_i(t)^2, \gamma_i(t)\right) \tag{23}$$

where

$\overline{GGD}$ is the fitting function to the generalized Gaussian distribution
$\mu_i(t)$ is the fitted mean of $p_i(t)$
$\sigma_i(t)^2$ is the fitted variance of $p_i(t)$
$\gamma_i(t)$ is the fitted shape parameter of $p_i(t)$

The disparity activity of the i th subband in time t shall be obtained by

$$A_{i,D}(t) = \frac{\gamma_{\max} - \gamma_i(t)}{\gamma_{\max} - \gamma_{\min}} \tag{24}$$

where

$\gamma_{\min}$ is the minimum shape parameter
$\gamma_{\max}$ is the maximum shape parameter
$\gamma_i(t)$ is the shape parameter of the i th subband in time t

Hence, the disparity activity shall be written by

$$A_D = \sum_t \left(\sum_i A_{i,D}(t) / I\right) / T \tag{25}$$

where

T is the time duration
I is the number of subbands

6.4.3 Motion activity

$\mathbf{M}_M$ denotes the motion map of spatially occupied voxels at the x, y, and z axes. For each direction x, y, or z, the mean value of location changes over a period of time shall be obtained by

$$\mathbf{M}_M(t) = \sum_{\tau=t-T_M+1}^{t} \frac{\partial o(x(\tau))}{\partial t} \Big/ T_M \tag{26}$$

where

$\mathbf{M}_M(t)$ is $\mathbf{M}_M$ in time t,
T_M is the time interval for extracting the motion map,
$\mathbf{M}_M(t)$ is the partial differentiation (motion) of voxels in the original space between time $t-T_M+1$ and t.

To assess the motion activity, the motion map $\mathbf{M}_M$ is applied to the PMF like calculating the disparity activity.

$$\Lambda_j(w,h,t) = \begin{cases} 1, & \text{if } j-\overline{B}/2 \le W_i(M_D(t))(w,h) \le j+\overline{B}/2 \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

where

i_w is the width and height of the i^{th} subband
i_h is the width and height of the i^{th} subband

Then, the wavelet coefficients shall be fitted into a statistical model and its shape parameter shall be found. Denote the GGD fitting to $p_i(t)$ as

$$\overline{GGD}\left(p_i(t) : \mu_i(t), \sigma_i(t)^2, \gamma_i(t)\right) \tag{28}$$

where

$\overline{GGD}$ is the fitting function to the generalized Gaussian distribution
$\mu_i(t)$ is the fitted mean of $p_i(t)$
$\sigma_i(t)^2$ is the fitted variance of $p_i(t)$
$\gamma_i(t)$ is the fitted shape parameter of $p_i(t)$

The motion activity of the i^{th} subband in time t shall be obtained by

$$A_{i,M}(t) = \frac{\gamma_{\max} - \gamma_i(t)}{\gamma_{\max} - \gamma_{\min}} \tag{29}$$

where

$\gamma_{\min}$ is the minimum shape parameter
$\gamma_{\max}$ is the maximum shape parameter
$\gamma_i(t)$ is the shape parameter of the i^{th} subband in time t

Hence, the motion activity shall be written by

$$A_M = \sum_t \left(\sum_i A_{i,M}(t) / I \right) / T \tag{30}$$

where

T is the time duration
I is the number of subbands

6.4.4 Color activity

To obtain a coherent color map, low-pass and median filtering operations are applied to decrease textural variations, allowing the color components to be observed more clearly. The color map for each color component of Cb and Cr in $YCbCr$ space shall be obtained by

$$M_C(t) = F_M\left(F_L\left(o_L(\mathbf{x}(t))\big|_C\right)\right), \quad C \in \{Cb, Cr\} \tag{31}$$

where

$M_C(t)$ is M_C as a function of time t
F_M is median filtering
F_L is low pass filtering
$o_L(\cdot)\big|_C$ is the color component (Cb and Cr) of the left image

To assess the color activity, the color map M_C shall be applied to the PMF like calculating the disparity activity.

$$\Lambda_j(w,h,t) = \begin{cases} 1, & \text{if } j - \overline{B}/2 \le W_i(M_C(t))(w,h) \le j + \overline{B}/2 \\ 0, & \text{otherwise} \end{cases} \tag{32}$$

where

i_w is the width and height of the i th subband
i_h is the width and height of the i th subband

Then, the wavelet coefficients shall be fitted into a statistical model and its shape parameter shall be found. Denote the GGD fitting to $p_i(t)$ as

$$\overline{GGD}\left(p_i(t) : \mu_i(t), \sigma_i(t)^2, \gamma_i(t)\right) \tag{33}$$

where

$\overline{GGD}$ is the fitting function to the generalized Gaussian distribution
$\mu_i(t)$ is the fitted mean of $p_i(t)$

$\sigma_i(t)^2$ is the fitted variance of $p_i(t)$
$\gamma_i(t)$ is the fitted shape parameter of $p_i(t)$

The color activity of the i^{th} subband in time t shall be obtained by

$$A_{i,C}(t) = \frac{\gamma_{\max} - \gamma_i(t)}{\gamma_{\max} - \gamma_{\min}} \tag{34}$$

where

$\gamma_{\min}$ is the minimum shape parameter
$\gamma_{\max}$ is the maximum shape parameter
$\gamma_i(t)$ is the shape parameter of the i^{th} subband in time t

Hence, the color activity shall be written by

$$A_C = \sum_t \left(\sum_i A_{i,C}(t) / I \right) / T \tag{35}$$

where

T is the time duration
I is the number of subbands

6.4.5 QoE prediction based on temporal characteristics

The 3D video is analyzed, and statistical features related to disparity, motion, and color are extracted. The obtained feature maps are evaluated in the wavelet domain, and the activities (labelled M_D, M_M, and M_C, respectively) are then obtained by Equation (18) through Equation (35). Combining the features produces the 3D visual activity measure.

Finally, the overall 3D visual activity index shall be obtained by

$$A = \sum_k w_k A_k, \quad k \in \{D, M, C\} \tag{36}$$

where

w_k is a parameter that is used to adjust the relative importance of the three factors.

The 3D visual activity shall be used to measure the QoE when viewing a 3D video. Highly "active" 3D content often leads to increased visual discomfort; therefore, the use of a perceptual activity measure shall be used to predict the QoE. A subjective study was conducted to evaluate the relation between the human perception of QoE and the 3D visual activity measure. The video sequences used in the subjective test were drawn from the IEEE-SA stereo image database 0. To combine the three factors with appropriate weights w_k ($k \in \{D, M, C\}$), an SVR was also applied. The weights computed turned out to be $w_D = 0.3557$,

$w_M = 0.4471$, and $w_C = 0.1961$. The computed Pearson linear correlation coefficient (PLCC) (0.7499) and Spearman's rank correlation coefficient (SROCC) (0.7344) indicate that 3D visual activity functions quite reasonably well as a predictor of QoE [B13].

7. Subjective assessment

7.1 General

A new methodology that we call multimodal interactive continuous scoring of quality (MICSQ) for continuous assessment method is introduced in this clause. In addition, the optimal paired comparison method is also investigated in this clause.

7.2 Display device for subjective assessment

The results of subjective assessment shall be strongly affected by the performance of display, especially when viewing 3D contents. It is important to note that system crosstalk is independent of the content and it is determined by the display itself (system crosstalk).

The system crosstalk shall be defined by

$$\text{system crosstalk} = \frac{\text{percentage part of the right-eye image leaked to the left-eye position}}{\text{percentage part of the left-eye image observed at the left-eye position}} \quad (37)$$

Then, the value of system crosstalk for subjective 3D image and video assessments shall be lower than 5%. Preferably, when the value of system crosstalk is lower than 2%, high reliability of the subjective assessment results shall be achieved.

7.3 Continuous assessment methodology

7.3.1 General

Continuous assessment methodology has been widely used in most previous studies because of the time-variant 3D content characteristics, such as derivatives of motion and disparity, as shown in ITU-R BT.1438. Conventionally, single stimulus continuous quality evaluation (SSCQE) has been used to assess presence, depth, and naturalness in 2D and 3D videos. However, it is migrated from subjective 2D video quality assessment, and the accuracy of this methodology has not been measured empirically in 3D viewing environments. Thus, a new reliable subjective assessment methodology for the continuous evaluation in 3D contents named multimodal interactive continuous scoring of quality (MICSQ) is investigated in this standard.

7.3.2 Multimodal interactive continuous scoring of quality (MICSQ)

7.3.2.1 Interaction process of MICSQ

MICSQ is based on the interaction process among the 3D display, the assessment tool, and subject(s). It is activated by the device-interaction and human-interaction processes as follows.

The device interaction process of MICSQ shall be conducted by using the interaction process between the server and assessment tool in real-time as shown in Figure 10. The two important roles of the server are playing a test sequence and storing the rating score obtained by subjects. To maintain the long-period synchronization, the tablet shall send the duration time from the beginning to the end of the assessment. However, when the subject stops the assessment task by mistake, the device interaction process shall make the assessment task cease and restart from the beginning of the test sequence.

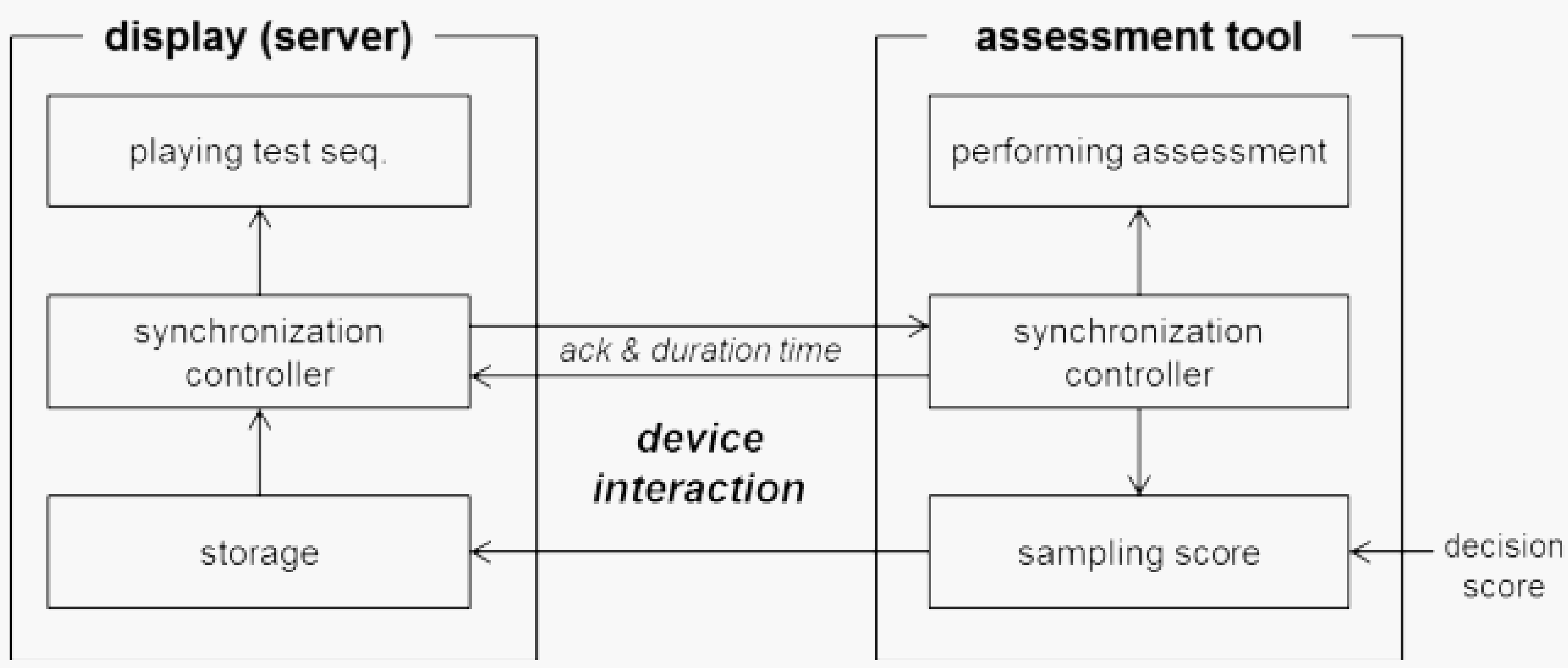


Figure 10—Device interaction of MICSQ

The human interaction process of MICSQ provides a main difference from conventional processes by using a tablet. The concentration loss interrupts an accurate task of continuous assessment, and the immersion to 3D video also obtains unreliable assessment results. To resolve these problems, the assessment tool for MICSQ makes sight, hearing, and touch stimulus to subjects consistently during the assessment, as shown in Figure 11. The tablet shall lead the subjects to prevent loss of concentration by using the periodic flickering, beep sound, and vibration. In order to reduce the immersion problem, the tablet shall adjust the cycle timing of the flickering, beep sound, and vibration randomly and announce the rating score at every second. By announcing the score, the subjects shall perceive their own rating score unwittingly, even if the subjects are fully absorbed in watching test sequence. In particular, although the standard has not been officially discussed yet, the haptic interface, which is a tactile feedback technology between the subject and the tablet, is very appropriate as a human interaction process. Arranged items of the assistant device against the concentration loss and immersion problems are summarized in Table 1.

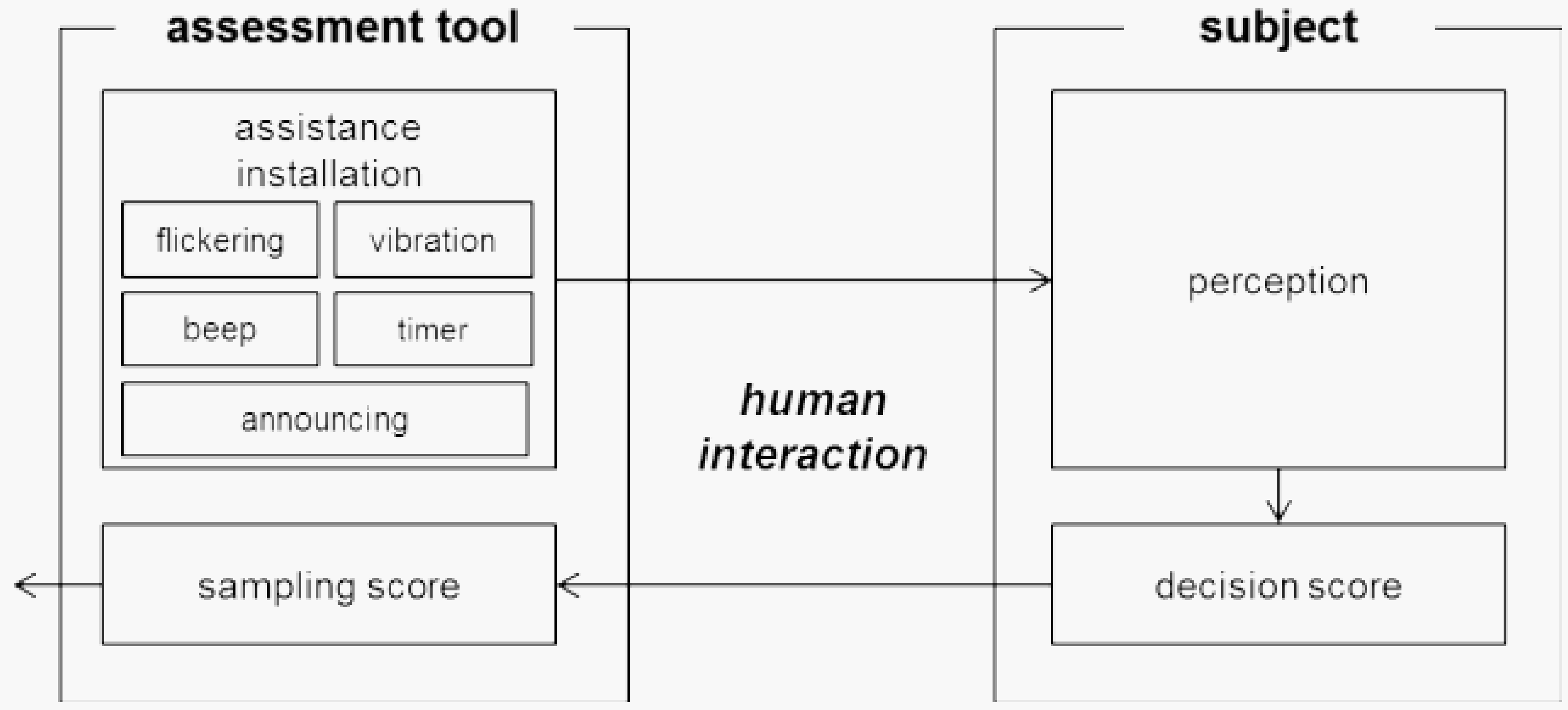


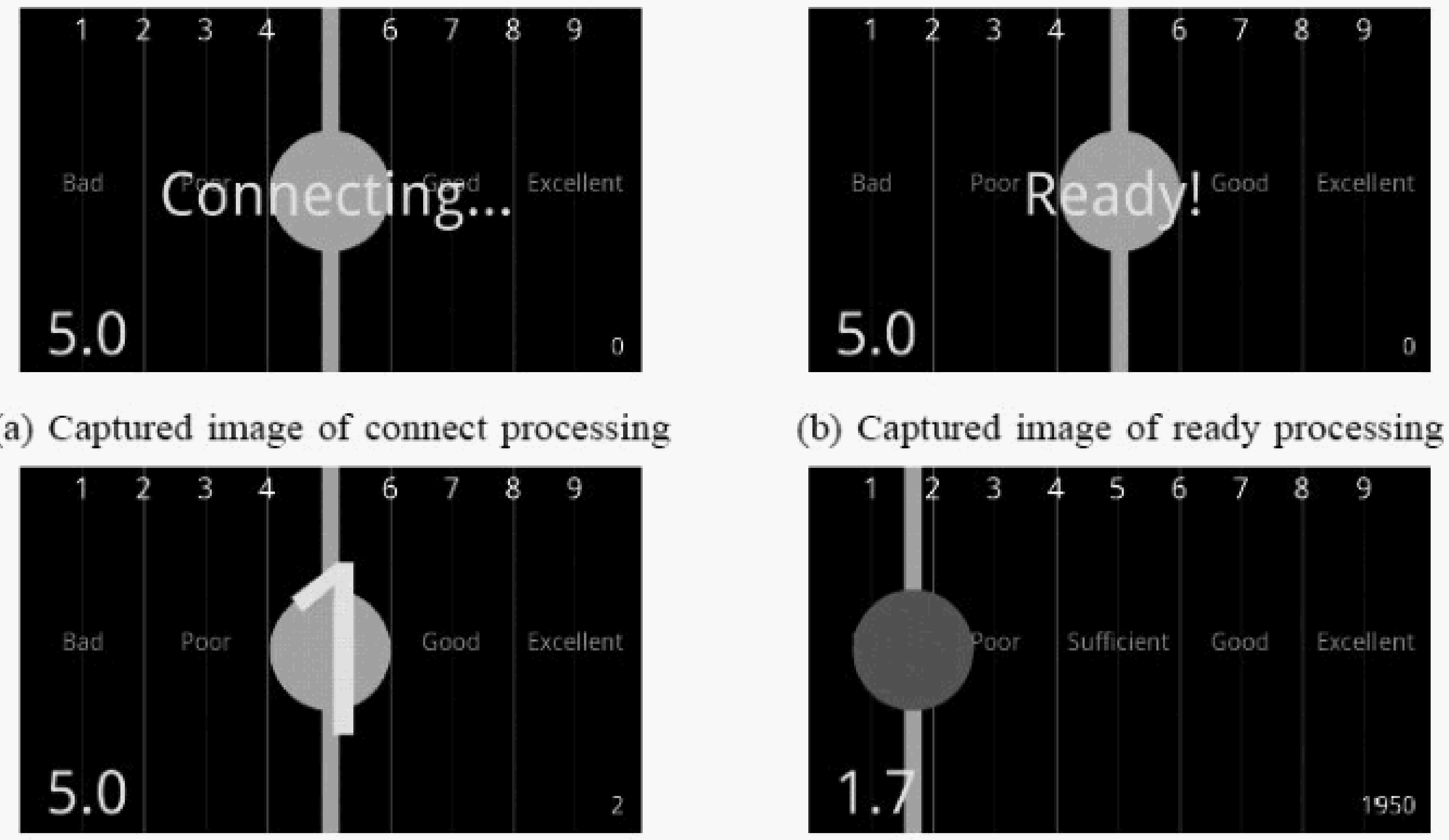
Figure 11—Human interaction of MICSQ

Table 1—Multimodal cues of MICSQ

<table>
<tr><th rowspan="3"></th><th colspan="4">Multimodal cues</th></tr>
<tr><th colspan="2">Aural</th><th rowspan="2">Tactile</th><th rowspan="2">Visual</th></tr>
<tr><th>Announcement</th><th>Beep</th></tr>
<tr><td>Concentration loss problem</td><td colspan="2">periodic beeping</td><td colspan="2">periodic (vibration/flickering)</td></tr>
<tr><td rowspan="2">Immersion problem</td><td>—</td><td>short-term sound</td><td colspan="2">short-term (vibration/flickering)</td></tr>
<tr><td>wrong situation</td><td>short-term sound</td><td colspan="2">continuous (vibration/flickering)</td></tr>
</table>

7.3.2.2 Interface of MICSQ

The position of the slider in MICSQ shall be adjusted along a zero to ten score (it can be adjusted) with the adjective terms [bad], [poor], [sufficient], [good], and [excellent], following the ITU recommendation. The current rating is displayed on the lower left corner of the screen, and the processing time in milliseconds is displayed on the lower right corner of screen, as shown in Figure 12. The score shall vary continuously when moving the subject's finger to left or right on the screen of the tablet.



(a) Captured image of connect processing

(b) Captured image of ready processing

(c) Captured image of count processing

(d) Captured image of assess processing

Figure 12—Client's user interface of MICSQ

7.3.3 Subjective assessment using MICSQ

7.3.3.1 Geometry of MICSQ

MICSQ shall require participants to provide a real-time subjective rating for visual quality, discomfort, fatigue, visibility, or sharpness by using a tablet-pc, as shown in Figure 13.

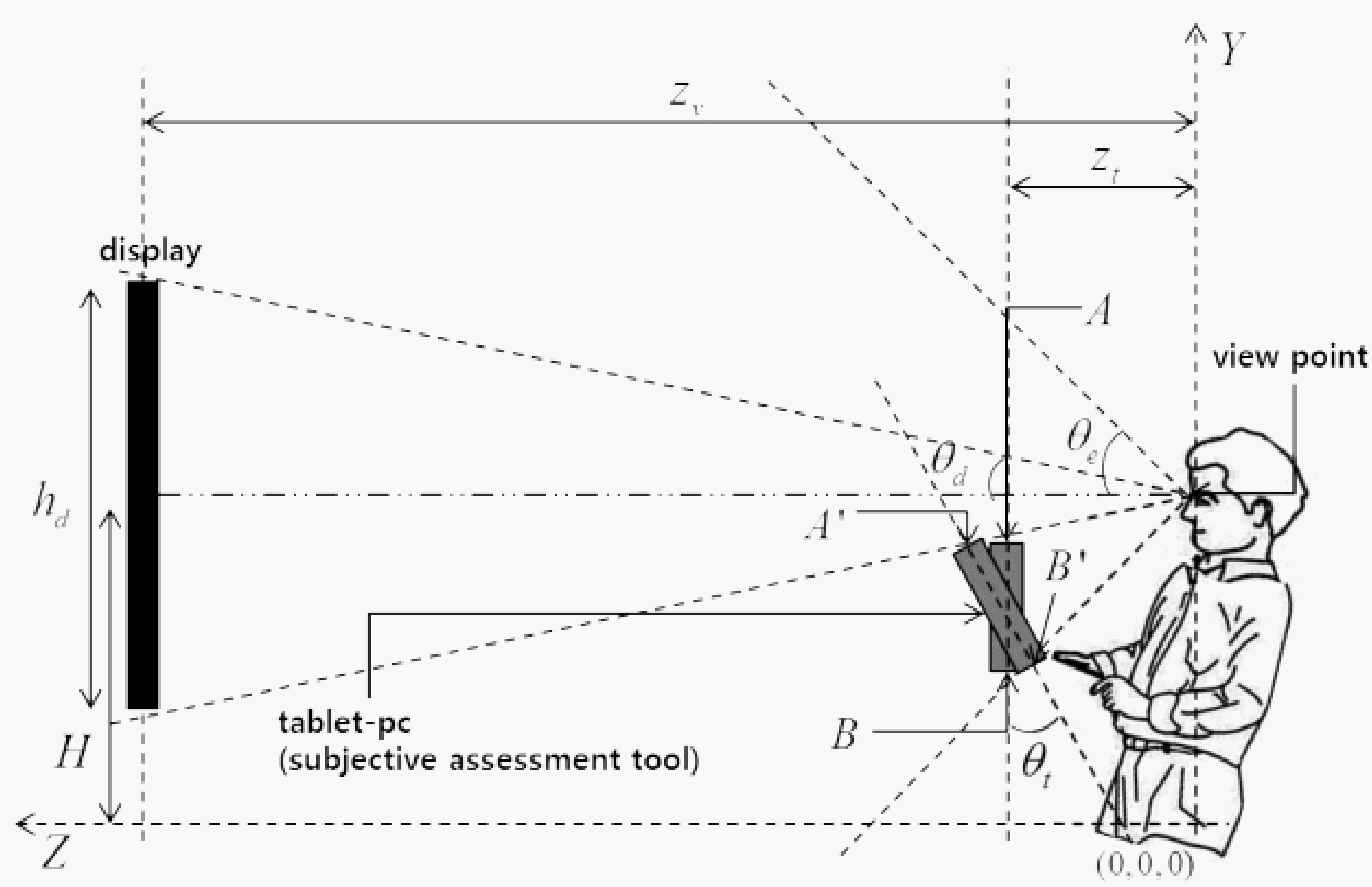


Figure 13—Geometry of MICSQ

Notate the subject's viewing angle of vertical direction as $2\theta_e$ and the angle among the subject's eye over the top and bottom of the display as $2\theta_d$. The line between the center of the display and the subject's eye is parallel with the floor with the height H . The angle between the horizontal line and the top and bottom of the display is

$$\theta_d = \arctan\left(\frac{h_d}{2z_v}\right) \tag{38}$$

As shown in Figure 13, the coordinate of an intersecting point A between the subject's sight and the top of the tablet is $(X,Y,Z)=\left(0, H-\frac{h_d z_t}{2z_v}, z_t\right)$. Thus, the point B is calculated by $\left(0, H-\frac{h_d z_t}{2z_v}-h_t, z_t\right)$ where h_t is the height of the tablet while satisfying

$$H-\frac{h_d z_t}{2z_v}-h_t \geq H - z_t \tan\theta_e . \tag{39}$$

However, if the subjective assessment is conducted by setting the tablet to be perpendicular to the floor, subjects shall not perceive the tablet screen with a uniform angle from the eyes. Thus, it is better to lean the tablet toward the display. Considering this, this standard shall modify the state of the tablet by leaning backward to the angle of θ_t, so that the modified coordinates of A and B are calculated by $A' = \left(0, h-\frac{h_d z_t}{2z_v}(z_t + w_t \sin\theta_t), z_t + w_t \sin\theta_t\right)$ and $B' = (0, \tan\theta_e(z_t - w_t \sin\theta_t), z_t - w_t \sin\theta_t)$, respectively. Therefore, the angle of θ_t shall be

$$\cos\theta_t \leq \frac{1}{h_t}\left(z_t \tan\theta_e - \frac{h_d z_t}{2z_v}\right). \tag{40}$$

7.3.3.2 Assessment environment

Single-user subjective 3D QoE or visual comfort assessments using MICSQ are performed using 3D displays such as polarized, shutter-glass stereoscopic displays, and autostereoscopic 3D displays. The viewing distance shall be set to three times of the height of the display. In addition, an Android-based tablet-pc shall be needed.

At least 20 people shall participate in the assessment wherein much of them shall not be involved in 3D or image processing research. All subjects shall have a normal or good visual acuity of greater than 1 and a good stereoscopic acuity of less than 60 arc.

7.4 Paired comparison methodology

7.4.1 General

In paired-comparison tests, there are two presentation patterns for the stimulus pair, i.e., the way to display the stimulus pair to observers. The stimuli can either be presented one after the other on a single screen (time-sequential presentation) or they shall be presented on two well calibrated and synchronized screens (time-parallel presentation). Different presentation patterns shall be used for different experimental setups.

The number of observers is defined to make the obtained subjective experimental results reach an acceptable accuracy level. In general, about 40 observers shall be sufficient to obtain reasonable accuracy. In addition, the number of stimuli shall be defined to make the paired comparison test feasible and the presentation order shall be defined to avoid any bias effects from the presentation order of the test stimuli on the observers' judgment.

7.4.2 Optimized rectangular design for paired comparison

Optimized rectangular design (ORD) is a balanced efficient design for paired comparison. "Balanced" means the occurrence frequency of each stimulus under test is identical. "Efficient" means, unlike full-paired comparison, only a subset of pairs are compared with this design. The selection of the pairs fulfills the efficiency criterion if they provide more information in a statistical sense on the estimates of the stimuli than the other pairs.

ORD is used under the constraint that the rank ordering of the stimuli S_1, S_2, ..., S_N, is known or can be estimated, for example, based on pre-test results or prior knowledge. The number of stimuli has to be a divisible number. The basic idea is to arrange the test stimuli into a rectangular matrix according to the rank ordering and then only compare the stimuli which are in the same column or row. In this standard, a new ORD method named optimized square design (OSD) is developed as follows.

Suppose the ordering indices of the stimuli (descending or ascending) are $\mathbf{d} = (d_1, d_2, ..., d_N)$. The rectangular matrix is arranged in such a way that the elements of the vector $\mathbf{d}$ are placed along a spiral as shown in Figure 14, which shall be defined as matrix R_{ORD}.

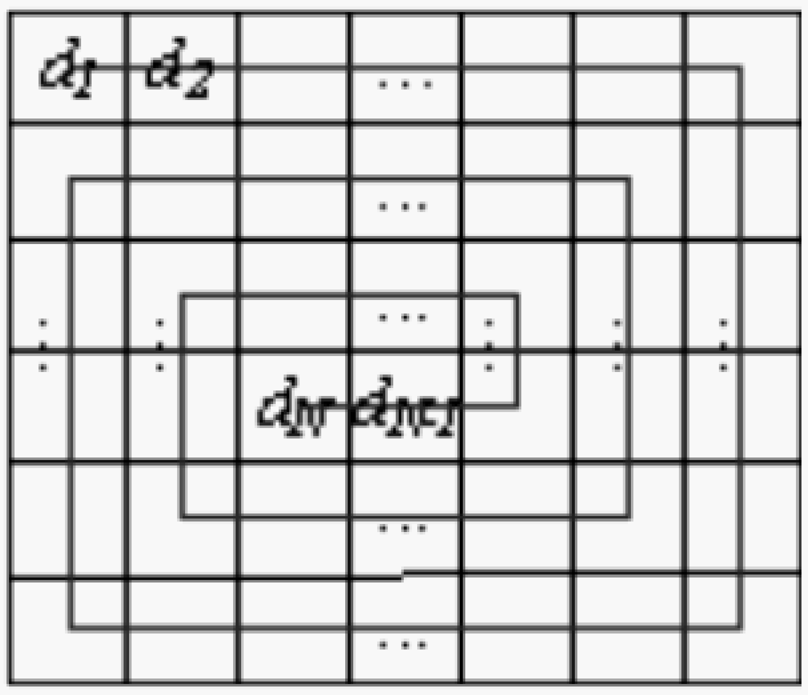


Figure 14—The design of rectangular matrix R_{ORD}

The stimulus pair $\{S_i S_j\}$ is compared if and only if $(i,j) \in C$ $(i,j) \in C$, where C shall be defined as:

$$C = \{(x,y) | p = p' \vee q = q', \text{where } x = r_{pq}, y = r_{p'q'} \text{ in } R_{ORD}\}. \tag{41}$$

$R_{ORD} = (r_{pq})_{t_1 \times t_2}$, r_{pq} is the index of the stimulus in position (p,q). In this design, the matrix R_{ORD} is fixed for all observers.

For better understanding, an example is given here. Supposing there are 12 test stimuli. Based on prior-knowledge, the rank ordering of these stimuli shall be estimated, which is $\mathbf{d} = (2,5,6,1,8,9,3,10,4,11,7,12)$. The R_{ORD} thus shall be designed as follows:

$$R_{ORD} = \begin{bmatrix} 2 & 5 & 6 & 1 \\ 11 & 7 & 12 & 8 \\ 4 & 10 & 3 & 9 \end{bmatrix}. \tag{42}$$

In this way, the adjacent stimulus indices d_i and d_{i+1} are always arranged in the same column or row of the matrix R_{ORD}. For example in this example,
$C = \{(2,5),(2,6),(2,1),(5,6),(5,1),(6,1),(11,7),(11,12),...(2,11),(2,4),(11,4),(5,7),...\}$. In the test, each participant compares the stimulus pairs whose indices belong to the set C, i.e., stimuli $\{S_2 S_5\}$ and $\{S_2 S_6\}$, etc. The number of appearances for each stimulus is five for each participant.

7.4.3 Adaptive rectangular design for paired comparison

Adaptive rectangular design (ARD) is also a balanced and efficient design similar to ORD. This design is used when the rank ordering of the test stimuli is not available. In this case, the test stimuli are arranged into a rectangular matrix according to all previous test observations. For the first observer, the arrangement is random. The matrix is adaptive to all previous observations and updated to each observer. The rule of the comparison is exactly the same with the ORD, i.e., only the stimuli that are in the same column or row of the rectangular matrix are compared.

In this standard, ARD is proposed in the way that the matrix R_{ORD} is updated for each observer. Adaptive square design (ASD) is a special case for ARD, where R_{ORD} is a squared matrix. This adaptive design is used for the conditions that previous estimates are not available. The detailed steps of this design are as follows:

a) For the first observer, the indices of the stimuli are randomly placed in R_{ORD}. The pair comparison experiment is executed, as specified for ORD, only the pairs whose indices are in the same column or row of R_{ORD} are compared.

b) According to all obtained $k-1$ $(k \geq 2)$ observations on the pairs, the paired comparison data shall be converted to scale values by utilizing the Bradley-Terry model or Thurstone-Mosteller model. The rank ordering indices of the stimuli (descending or ascending) $d^{k-1} = (d_1^{k-1}, d_2^{k-1}, ..., d_N^{k-1})$ shall be obtained (d^{k-1} represents the vector of ordering indices after $k-1$ times of observations).

c) For the k^{th} observer $(k \geq 2)$, based on the ordering vector d^{k-1}, the matrix R_{ORD}^k and C^k are constructed as shown in Figure 14, (R_{ORD}^k and C^k represents R_{ORD} and C for the k^{th} observer).

The pair comparison experiment is executed, as specified for ORD, only the pairs whose indices are in the set C^k are compared.

d) Repeat from b), until termination conditions are satisfied (e.g., all observers finished the test or the targeted accuracy based on confidence intervals is obtained).

The following shows an example with 12 stimuli as presented beforehand. As there is no pre-test for the test stimuli, for the first observer, the indices of the stimuli are randomly arranged in the matrix as follows:

$$R_{ORD}^{1} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}. \tag{43}$$

Thus, for the first observer, there are in total $3(\text{row}) \times \frac{4 \times 3}{2} + 4(\text{column}) \times \frac{3 \times 2}{2} = 30$ pairs to compare, i.e., $\{S_1 S_2\}, \{S_1 S_3\}, \{S_1 S_4\}, \{S_2 S_3\}, \ldots \{S_{11} S_{12}\}$. After the first observer's test, the rank ordering of the quality of the stimuli shall be estimated as $d^1 = (3,5,1,6,9,12,2,4,8,7,10,11)$.

For the second observer, the matrix R_{ORD} is arranged according to this rank ordering in the before-mentioned spiral, thus:

$$R_{ORD}^{2} = \begin{bmatrix} 3 & 5 & 1 & 6 \\ 7 & 10 & 11 & 9 \\ 8 & 4 & 2 & 12 \end{bmatrix}. \tag{44}$$

Then, for the third observer, the matrix R_{ORD}^{3} is updated based on all previous two observers' pair-comparison results. This procedure of executing the subjective assessment, calculating the ranking, and rearranging the data into the matrix continues until the test is finished.

7.4.4 Subjective assessment using paired comparison

7.4.4.1 Number of observers and stimuli

Generally, in a paired comparison test using full paired comparison (FPC) method, 10 is the minimum required number of observers. To achieve the same level of accuracy (20 observers in an FPC test), 40 observers are necessary for the ARD or ORD methods.

In addition, the number of stimuli shall be taken into account to avoid any fatigue induced by watching the same video content. Thus, guidelines for the selection of the number of stimuli are proposed:

a) For each observer, if the viewing material is deemed comfortable, then the test time shall be 20 to 40 min intermixed with breaks. If the material is known to contain excessive parallax, and thus known to be potentially uncomfortable, then the duration shall be limited.

b) The number of stimuli shall be estimated based on the test method and the duration of the stimuli. For example, if the duration of the test stimulus is 10 s, and between each sequence shall be a gray image that lasts 3 s, adding 5 s for voting, one pair will require $10 + 3 + 10 + 5 = 28$ s using time-sequential paired comparison, and $10 + 5 + 3 = 18$ s using time-parallel-paired comparison. The targeted number of pairs in the test is 43 to 86 pairs for time-sequential method, and 67 to 134 pairs for the side-by-side method.

c) Generally, in an image/video quality assessment experiment, there are source video contents (SRC) with different types of distortions or different levels of degradations. In this case, only the video sequences of the same content but with different degradations are compared in the test. The number of video contents and the number of degradation types shall be estimated to follow Guideline 1. For example, if 6 SRCs are selected and the planned test duration is approximately 40 min, the time-parallel comparison is selected, then, for each viewer, the maximum number of pairs is 134 which lead to $134/6 \approx 22$ pairs/SRC. According to Figure 15, when mapping 22 to the x-axis, the corresponding y-axis value for the number of hypothetical reference circuits (HRCs) is approximately 10, e.g., the matrix R_{ORD} shall be 2×5 or 2×4 or 3×3.

The selection of the number of SRCs shall take into consideration that observers would not get fatigued, impatient, or annoyed with watching the same content repeatedly, i.e., the number of contents shall be sufficiently large.

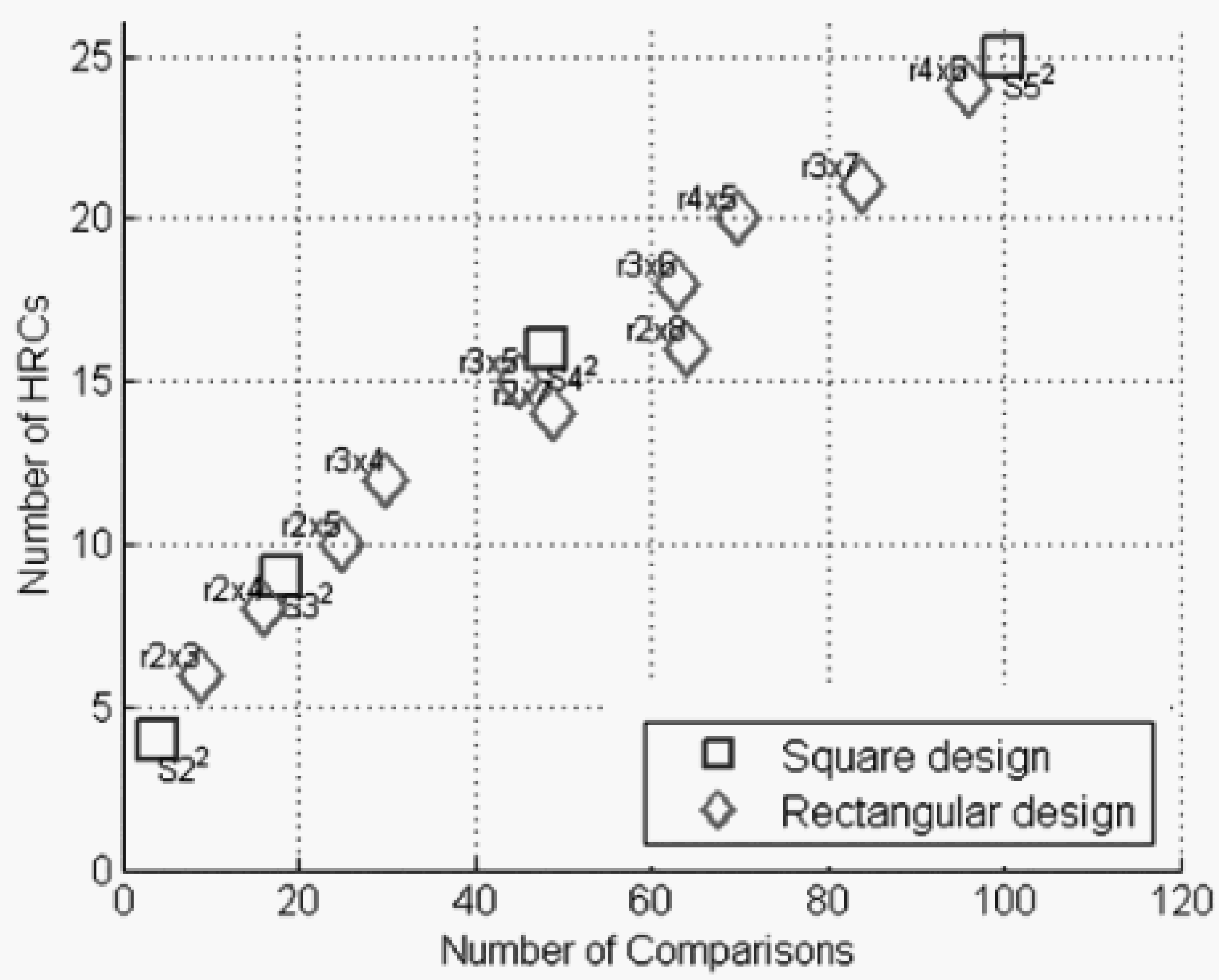


Figure 15—Relationship between the number of HRCs and comparisons in rectangular design

7.4.4.2 Constraints on presentation order

In the process of the stimulus presentation, an imbalance of the randomization of the stimuli shall affect the paired comparison results significantly. Thus, the constraints on the stimulus randomization shall be defined as follows:

a) The presentation of the sequence content shall be random, no observer watches the same content in two consecutive presentations.

b) For each observer, the presentation order for each sequence shall be balanced, i.e., $\{S_A S_X\}$, $\{S_Y S_A\}$. This means for all the pairs that include stimulus S_A, half of the pairs shall show S_A first, the rest shall show S_A second in time-sequential pair comparison presentation. For the condition of time-parallel pair comparison, half of the pairs shall show S_A on the left screen, the rest shall show S_A on the right screen.

For all observers, all the pairs of stimuli shall be displayed in both orders. For example, if one observer watches $\{S_A S_B\}$, there shall be another observer who watches $\{S_B S_A\}$.

Annex A

(informative)

Bibliography

Bibliographical references are resources that provide additional or helpful material but do not need to be understood or used to implement this standard. Reference to these resources is made for informational use only.

[B1] Bradley, R. A., "14 Paired comparisons: Some basic procedures and examples," *Handbook of statistics 4: Nonparametric Methods*, Elsevier, 1985, pp. 299–326. [4]

[B2] Clifford, C. W. G., "Perceptual adaptation: motion parallels orientation," *Trends in Cognitive Sciences*, vol. 6, no. 3, pp. 136–143, March 2002.

[B3] Daly, S. J., R. T. Held, and D. M. Hoffman, "Perceptual issues in stereoscopic signal processing," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pt. 2, pp. 347–361, June 2011. [5]

[B4] Eadie, A. S. and P. J. Carlin, "Evolution of control system models of ocular accommodation, vergence and their interaction," *Medical & Biological Engineering & Computing*, vol. 33, no. 4, pp. 517–524, July 1995.[6]

[B5] Geisler, W. S. and J. S. Perry, "A real-time foveated multiresolution system for low bandwidth video communication," *Proceedings of the SPIE*, vol.3299, July 1998.[7]

[B6] Haigh, S. M., L. Barningham, M. Berntsen, L. V. Coutts, E. S. T. Hobbs, J. Irabor, E. M. Lever, P. Tang and A. J. Wilkins, "Discomfort and the cortical haemodynamic response to coloured gratings," *Vision Research*, vol. 89, pp. 47–53, August 2013.

[B7] ISO 11664-1: 2007, "Colorimetry—Part 1: CIE standard colorimetric observers."[8]

[B8] Kim, H., S. Lee and A. C. Bovik, "Saliency prediction on stereoscopic videos," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1476–1490, April 2014.

[B9] Kim, T., J. Kang, S. Lee and A. C. Bovik, "Multimodal Interactive Continuous Scoring of Subjective 3D Video Quality of Experience," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 387–402, February 2014.

[B10] Lambooij, M., M. Fortuin, I. Heynderickx, and W. IJsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review," Journal of Imaging Science and Technology, vol. 53, no. 3, pp. 30201–30214, May 2009.[9]

[B11] Lambooij, M, W. A. IJsselsteijn, and I. Heynderickx, "Visual discomfort of 3D TV: Assessment methods and modeling," *Displays*, vol. 32, no. 4, pp. 209–218, October 2011.

[B12] Le Callet, P., S. Möller and A. Perkis, eds., "Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)", Lausanne, Switzerland, Version 1.2, March 2013.[10]

[4] Elsevier journals are available at http://www.sciencedirect.com/
[5] IEEE publications are available from The Institute of Electrical and Electronics Engineers, 445 Hoes Lane, Piscataway, NJ 08854, USA (http://standards.ieee.org/).
[6] Springer publications are available at http://link.springer.com/
[7] SPIE publications are available at http://proceedings.spiedigitallibrary.org/
[8] ISO publications are available from the ISO Central Secretariat, 1, ch. de la Voie-Creuse, CP 56, CH-1211 Geneva 20, Switzerland (http://www.iso.org/). ISO publications are also available in the United States from the Sales Department, American National Standards Institute, 25 West 43rd Street, 4th Floor, New York, NY 10036, USA (http://www.ansi.org/).
[9] Available at http://ist.publisher.ingentaconnect.com/content/ist/jist/2009/00000053/00000003/art00001
[10] Available at http://www.qualinet.eu/images/stories/QoE_whitepaper_v1.2.pdf

[B13] Lee, K, A. K. Moorthy, S. Lee and A. C. Bovik, “3D Visual activity assessment based on natural scene statistics,” *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 450–465, January 2014.

[B14] Li, J., M. Barkowsky, and P. Le Callet, “Analysis and improvement of a paired comparison method in the application of 3DTV subjective experiment,” IEEE International Conference on Image Processing (IEEE ICIP 2012), Orlando, USA, Oct., 2012.

[B15] Li, J., M. Barkowsky, and P. Le Callet, “Boosting Paired Comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs,” *Proceedings of the SPIE Electronic Imaging, Stereoscopic Displays and Applications*, pp. 1–12, 2013.

[B16] Li, J., M. Barkowsky, and P. Le Callet, “Subjective assessment methodology for Preference of Experience in 3DTV,” IEEE Image, Video and Multidimensional Signal Processing Workshop (IVMSP 2013), Seoul, Korea, June, 2013.

[B17] Recommended standard procedures for the clinical measurement and specification of visual acuity, Report of working group 39. *Adv. Ophthalmol.*, 1980.

[B18] Schor, C. M., “A dynamic model of cross-coupling between accommodation and convergence: Simulations of step and frequency responses,” *Optometry and Vision Sciences*, vol. 69, no. 4, pp. 258–269, 1992.[11]

[B19] Tam, W. J., F. Speranza, S. Yano, K. Shimono, and H. Ono, “Stereoscopic 3D-TV: Visual comfort,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pt. 2, pp. 335–346, June 2011.

[B20] Thurstone, L. L., “A law of comparative judgment,” *Psychological Review*, vol. 34, no. 4, p. 273–286, 1927.

[B21] Wang, J, M. P. DaSilva, P. LeCallet, and V. Ricordel, “Computational model of stereoscopic 3D visual saliency,” *IEEE Transactions on Image Processing*, vol. 22, no.6, pp. 2151–2165, June 2013.

[B22] Wilkins, A., I. Nimmo-Smith, A. Tait, C. McManus, S. Della Sala, A. Tilley, K. Arnold, M. Barrie and S. Scott, “A neurological basis for visual discomfort,” *Brain*, vol. 107, no. 4, pp. 989–1017, 1984.[12]

[B23] Yang, S. N., T. Schlieski, B. Selmins, S. C. Cooper, R. A. Doherty, P. J. Corriveau, and J. E. Sheedy, “Stereoscopic viewing and reported perceived immersion and symptoms,” *Optometry and Vision Sciences*, vol. 89, no. 7, pp. 1068–1080, July 2012.

[B24] Yano, S., S. Ide, T. Mitsuhashi, and H. Thwaites, “A study of visual fatigue and visual comfort for 3D HDTV/HDTV images,” *Displays*, vol. 23, no. 4, pp. 191–201, 2002.

[11] Available at http://journals.lww.com/optvissci/Abstract/1992/04000/A_Dynamic_Model_of_Cross_Coupling_Between.2.aspx
[12] Oxford University Press publications are available at http://www.oxfordjournals.org/en/index.html

Consensus

WE BUILD IT.

Connect with us on:

Facebook: https://www.facebook.com/ieeesa

Twitter: @ieeesa

LinkedIn: http://www.linkedin.com/groups/IEEESA-Official-IEEE-Standards-Association-1791118

IEEE-SA Standards Insight blog: http://standardsinsight.com

YouTube: IEEE-SA Channel

IEEE
standards.ieee.org
Phone: +1 732 981 0060 Fax: +1 732 562 1571