Multimodal Interactive Continuous Scoring of Subjective 3D Video Quality of Experience

Taewan Kim, Jiwoo Kang, Sanghoon Lee, *Senior Member, IEEE*, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*—People experience a variety of 3D visual programs, such as 3D cinema, 3D TV and 3D games, making it necessary to deploy reliable methodologies for predicting each viewer's subjective experience. We propose a new methodology that we call *multimodal interactive continuous scoring of quality* (MICSQ). MICSQ is composed of a *device interaction process* between the 3D display and a separate device (PC, tablet, etc.) used as an assessment tool, and a *human interaction process* between the subject(s) and the separate device. The scoring process is multimodal, using aural and tactile cues to help engage and focus the subject(s) on their tasks by enhancing neuroplasticity. Recorded human responses to 3D visualizations obtained via MICSQ correlate highly with measurements of spatial and temporal activity in the 3D video content. We have also found that 3D quality of experience (QoE) assessment results obtained using MICSQ are more reliable over a wide dynamic range of content than obtained by the conventional single stimulus continuous quality evaluation (SSCQE) protocol. Moreover, the wireless device interaction process makes it possible for multiple subjects to assess 3D QoE simultaneously in a large space such as a movie theater, at different viewing angles and distances. We conducted a series of interesting 3D experiments showing the accuracy and versatility of the new system, while yielding new findings on visual comfort in terms of disparity, motion and an interesting relation between the naturalness and depth of field (DOF) of a stereo camera.**

***Index Terms*—Multimodal interactive continuous scoring of quality (MICSQ), 3D quality of experience (QoE), subjective assessment, visual comfort evaluation, interactive continuous subjective quality assessment, empirical 3D distortion.**

I. Introduction

RECENT successful technological developments in 3D displays and 3D image processing have led to an explosion in consumer demand for 3D content [1]–[5]. Many high-profile Hollywood 3D cinematic experiences have been produced using sophisticated 3D video capture and stereography techniques [6], and the first live 3D NFL football game was broadcasted this past year on BSkyB. This has brought to the front issues such as monitoring 3D video quality and acquiring 3D content that is comfortable to view [2]–[4], [7], [8]. Unlike 2D video, the ocular adjustment to 3D depth can induce neurological symptoms such as visual fatigue and headache, as well as 3D distortions that cause quality degradation [1], [2], [9]. Understanding these problems involves a number of intricate visual factors which can only be probed using a sophisticated subjective testing methodology for measuring the quality of the 3D visual experience [10].

A significant factor which strongly affects the 3D quality of experience (QoE) is *visual comfort*. The term visual comfort is used to express the subjective feeling of a physical state of ease (or lack thereof), which can vary over time, associated with the watching of 3D video content [1], [2]. Conflicts between vergence and accommodation that can occur on stereoscopic video appear to rarely affect our experience when watching the real world of natural disparity fields. Understanding how 3D visual comfort is affected by 3D video is a topic of intense study in engineering [4], [5], [11], ophthalmology [7], [12] and neurology [9], [13].

Because of the complex characteristics of the human visual system and individual differences, human viewers experience 3D QoE in diverse ways [1], [10], [14]. It is difficult to design a generic objective metric for 3D QoE prediction. Successful objective 3D image/video quality assessment (I/VQA) methods have not yet been demonstrated, in the sense that none delivers significantly better performance than simply applying monocular IQA algorithms to each image in the stereopair, then combining the results. Subjective interfaces for capturing 3D human quality assessment have largely been inherited from what has traditionally been done in 2D. However, the experimental 3D viewing environment is quite different from that in 2D due to the immersion experience of a user wearing 3D glasses, or the angle-dependent peculiarities currently perceived when viewing autostereoscopic displays.

As early as 1992, the authors of [15] made clear the need for new methods to evaluate 3D content. Since 2000, international standardization activities have accelerated on 3D display, human factors [16] and 3D image safety [17]. Yet, specific subjective 3D QoE assessment environments currently in use have some important drawbacks. To understand this, we will review currently used methods. The important and widely used absolute category rating (ACR) method dictates the collection of perceived quality scores of 3D content on a discrete 5-category

Manuscript received November 26, 2012; revised April 03, 2013 and July 30, 2013; accepted October 08, 2013. Date of publication November 25, 2013; date of current version January 15, 2014. This work was supported by the Technology Innovation Program (10042402, Standardization of 3D human factors and 3D medical application services) funded by the Ministry of Knowledge Economy (MKE) of Korea and the MSIP (Ministry of Science, ICT&Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2013-H0401-13-1003) supervised by the NIPA (National IT Industry Promotion Agency). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yiannis Andreopoulos.

T. Kim, J. Kang, and S. Lee (corresponding author) are with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: enoughrice21@yonsei.ac.kr, jwkang@yonsei.ac.kr; slee@yonsei.ac.kr).

Alan C. Bovik is with the Laboratory for Image and Video Engineering (LIVE), Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2292592

rating scale [18]. Continuous assessment methods where the subject rates the 3D video content continuously over time have been widely used because of the strong time-varying characteristics of 3D content [3], [7], [14], [19]–[21]. Single stimulus continuous quality evaluation (SSCQE) protocols to assess such attributes as 3D presence, depth and naturalness in stereoscopic video are also commonly used. SSCQE has also been used to assess visual discomfort levels experienced when viewing 3D videos generated via the 2D-plus-depth format [3].

The methods used to conduct 3D SSCQE studies have largely migrated intact from 2D. As a consequence, there exist limitations in the direct application of 2D SSCQE methods for assessing 3D video QoE, especially in regards to stably capturing relevant 3D quality attributes. Perhaps more so when viewing 3D than 2D, subjects tend to become deeply absorbed when viewing an ostensibly entertaining 3D video, thereby becoming distracted from the assessment process (commonly referred to as the immersion problem) [22]. This loss of concentration may be amplified during long assessment periods, leading to inaccuracies and asynchrony between the test sequence and the human response. Slowed reaction speeds can also cause degradations in the reliability of the QoE assessments.

Towards addressing these limitations, we propose a new 3D subjective methodology that we term *multimodal interactive continuous scoring of quality* (MICSQ) for 3D QoE assessment. The goal of MICSQ is to minimize distractions between viewing and assessment. As shown in Fig. 1(a), current interfaces for 3D SSCQE are ‘one-way’, since the subject decides and records scores on the same display as the video. MICSQ breaks this process into separate interactions: 1) between the tablet used as an assessment tool and the 3D display (*device interaction*) using a wireless network protocol, and 2) between the tablet and the subject (*human interaction*), using haptic and audition cues to enhance the recording of subjective results. It is also possible for multiple users to perform subjective assessment simultaneously.

This flexibility makes it possible to deploy this new subjective methodology in expanded and diverse visual experiments, such as measuring the degree of visual discomfort experienced by multiple simultaneous viewers having different placements relative to the display. The advantages of MICSQ are

- **High reliability**: Compared to conventional SSCQE, it enables a better, less distracting visual display (as verified by the evaluation of statistical confidence levels) without a distracting visual quality scale.
- **Multi-user assessment**: Using a wireless protocol and multiple tablets, it is possible to collect simultaneous subjective scores from multiple subjects viewing the same 3D content in a large space such as a movie theater.

In addition to validating its reliability, we demonstrate the efficacy and unique strengths of MICSQ in a variety of 3D experiments that address several critical factors relevant to 3D image and video perception:

- **Comfortable viewing zone (CVZ)**: The CVZ shrinks with increased time spent viewing uncomfortable stimuli.
- **Speed and direction of depth motion**: Visual comfort is affected differently by the speed of motion in depth depending on the (forward or backward) direction, even in the CVZ.

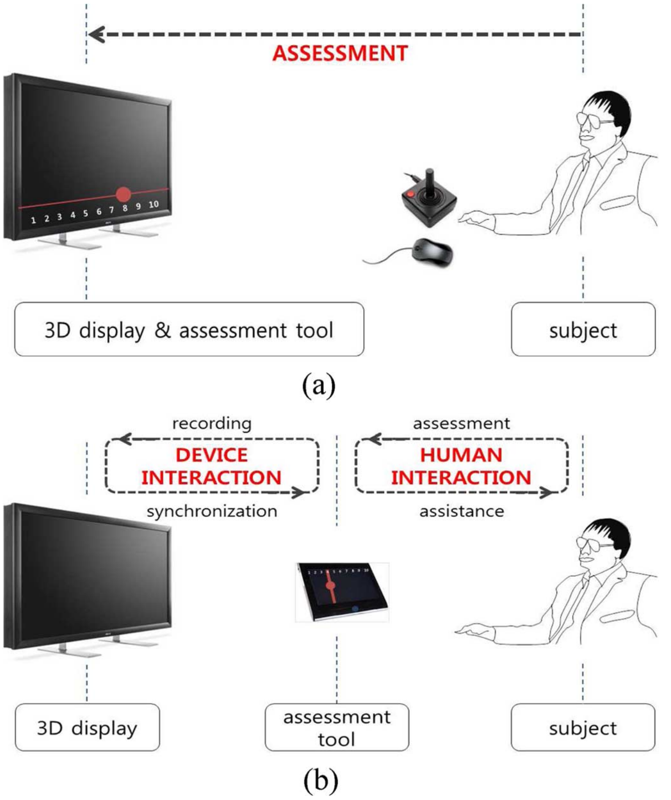


Fig. 1. Conceptual diagram of conventional and proposed MICSQ methodology for subjective 3D video QoE assessment. (a) Subject assesses 3D QoE using a slider (e.g., mouse or stick) [‘one-way’ framework]. (b) Subject(s) assess 3D QoE interactively using a tablet(s) to score the video that is viewed on a separate display [mutual ‘two-way’ framework]. (a) Conventional methodology. (b) Proposed methodology.

- **Artificial DOF on naturalness**: The naturalness of 3D video is affected by the parameters of the DOF of the camera relative to the diameter of the cornea.

The remainder of this paper is organized as follows. In Section II, we describe the design of MICSQ including the geometric parameters involved with single and multiple subjects. In Section III, we measure and compare the achievable accuracy of subjective assessment using MICSQ and SSCQE as a function of the 3D spatial and temporal complexities of the test sequences being viewed. Furthermore, we conduct a number of relevant experiments based on the reliability of MICSQ in Section IV. Finally, concluding remarks are given in Section V. The experimental set-up is summarized in the Appendix.

II. MICSQ Methodology for Subjective 3D QoE Assessment

Subjective assessment of 3D video can be viewed as a process of explorative study, involving psychophysical measurement and scaling, and questionnaires. Traditionally, explorative studies on 3D display have been conducted by gathering focus group opinions following the viewing of a test sequence [1]. These methods are broadly the same as methods described in 2D recommendation documents [23]. In the following, we propose a very different approach.

TABLE I
MULTIMODAL ASSESSMENT TOOL CUES

	Concentration loss Prob.	Immersion Prob.
Aural Cue	periodic beeping	randomized beeping &announcing score
Tactile Cue	periodic vibration	randomized vibration

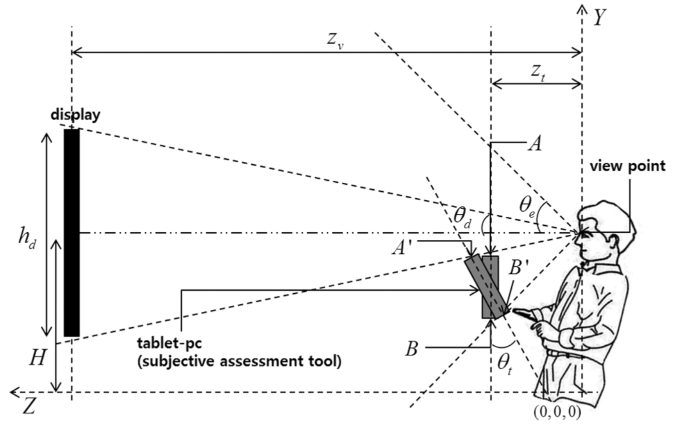


Fig. 2. MICSQ viewing environment for a single subject. The display and tablet screen simultaneously lie within the subject's FOV.

A. Multimodal Interactive Continuous Scoring of Quality (MICSQ)

MICSQ is a new approach to conducting subjective 3D QoE assessment experiments using interactions between the subject, a quality assessment tool, and the 3D display. By using a wireless connection between the 3D display and the (possibly multiple) quality assessment tool(s), single or multiple subjects can simultaneously participate in the same experiment, assessing the same video in a classroom or movie theater environment.

We propose a real-time *device interaction process* between the 3D display server and quality assessment tool, which we currently envision as a tablet. The server controls the play and display of each 3D video and storage of the rating scores given by the subjects.

The *human interaction process* introduces the new idea of a tablet-based multi-subject interface. We also propose multimodal protocols to deal with the aforementioned problems of concentration loss, which may lead to degradation or interruption of a continuous assessment task, and the related immersion problem. To address these problems, the assessment tool delivers audio and tactile cues to the subjects throughout the task; these cues are tabulated in Table I. For example, the tablet may prevent loss of concentration by supplying periodic beeping and vibration reminders until the end of the assessment. A variety of tactile and aural cues have been proposed and discussed in previous studies [24]–[29]. The physical and cognitive interactions between the responses to visual, haptic and aural cues tend to mutually boost the efficacy of each modality [27]. In [28], the authors created a multimodal device for visually impaired users to explore maps by offering a combination of audio and tactile stimuli. The authors found that users were better able to perform exploration tasks using these cues and assert that the combined use of aural and vibrational cues is crucial for the successful exploration of an image. Tactile and audio cues are particularly effective at enhancing perceptual ability when viewing fully immersive systems using large or head-mount displays [30].

Thus, to address the immersion problem, the tablet may adjust the timing of the audio and tactile cues randomly, and could periodically announce the rating score during the assessment task. By audibly announcing the score every second, the subjects remain aware of the scores they are delivering, even if they are fully absorbed in watching the video. Evidence for neuroplasticity of the motor system suggests that haptic and aural assistance is a good way to enhance task performance. While much work remains to be done regarding determining the degree to which multimodal cues can enhance subject performance in visual tasks, the available evidence regarding neuroplastic enhancement by multimodal activity suggests that such approaches may prove to be highly effective [24]. In particular, the haptic (tactile) feedback technology between the subject and the tablet has the potential to minimally intrude upon the subject's awareness while supplying cues for guidance, control, and distraction reduction. Multimodal interfaces that incorporate haptic and aural feedback can enhance the user experience by combining multiple synergistic information cues [31], with the potential to yield subjective task results that are more reliable, complete, and robust as compared to conventional methodologies.

B. Geometry of MICSQ

The proposed geometric layout of subject, display and tablet is depicted in Fig. 2. The subject views the display at a distance z_v while the tablet is located in front of the subject at a distance z_t. The viewing distance is assumed to be about $z_v = 3h_d$, where h_d is the height of the stereoscopic display (in agreement with ITU recommendation [23]). An aim of this environment is to allow the display and tablet screen to simultaneously and comfortably fall within the subject's field of view (FOV).

1) Single-Subject Geometry: Denote the subject's viewing angle in the vertical direction by $2\theta_e$, and the angle subtended by the subject's eye and the top and bottom of the display by $2\theta_d$. The line between the center of the display and the subject's eye (optical axis when fixated at screen center) is parallel with the floor at height H. The angle between this horizontal line and the top and bottom of the display is

$$\theta_d = \arctan\left(\frac{h_d}{2z_v}\right). \tag{1}$$

As shown in Fig. 2, the location of an intersecting point A between the subject's line of sight and the top of the tablet is $(X, Y, Z) = (0, H - \frac{h_d z_t}{2z_v}, z_t)$. Therefore, $B = (0, H - \frac{h_d z_t}{2z_v} - h_t, z_t)$ where h_t is the height of the tablet and

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

Based on this condition, the distance from subject to tablet is $z_t \geq h_t / (\tan\theta_e - \frac{h_d}{2z_v})$. To find a favorable position of the tablet for a given individual, z_t could be located in the range $z_t < l \cos\theta_d$ where l is the arm length of the viewer. Then

$$\frac{h_t}{\left(\tan\theta_e - \frac{h_d}{2z_v}\right)} \leq z_t \leq l \cos\theta_d. \tag{3}$$

If the subjective assessment is conducted with the tablet oriented perpendicular to the floor, the subjects may not perceive

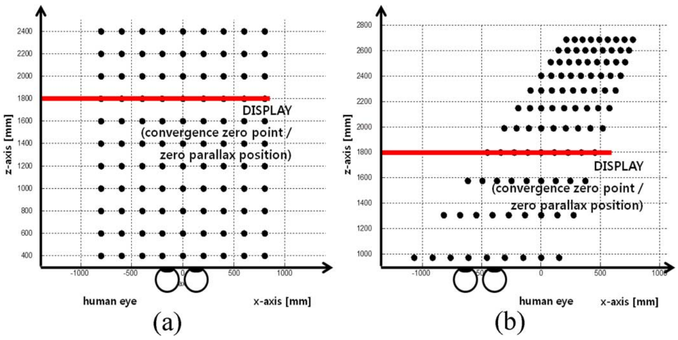


Fig. 3. For a given viewing distance $z_v = 1800$ mm, two grid shapes are used to demonstrate shear distortion as a function of viewing position. The pixel resolution unit is mm, and the interval of two adjacent pixels is 200 mm in the stereoscopic display along the x- and z-axes. (a) Viewing position assumed at the center (0, 0, 0). (b) Viewing position is located at 0.5 m to the left side $(-500, 0, 0)$.

the tablet screen at a uniform angle, thereby producing undesirable foreshortening and display artifacts. Thus, modify the angle of the tablet by the amount θ_t, yielding modified coordinates $A' = \left(0, h - \frac{h_d}{2z_v}(z_t + w_t \sin\theta_t), z_t + w_t \sin\theta_t\right)$ and $B' = (0, \tan\theta_e(z_t - w_t \sin\theta_t), z_t - w_t \sin\theta_t)$, respectively. Then the angle θ_t satisfies

$$\cos\theta_t \leq \frac{1}{h_t}\left(z_t \tan\theta_e - \frac{h_d z_t}{2z_v}\right). \quad (4)$$

Thus, the position of the tablet may be optimized within the ranges of z_t and θ_t, as shown in (3) and (4). If the conditions above are satisfied, then it is possible to perceive the test sequence and the rating score at the same time.

2) Multiple-Subjects Geometry: One of the merits of MICSQ is that multiple subjects can participate in the assessment task simultaneously in a large space. For multi-view autostereoscopic display experiments, multi-user assessment could be performed by arranging the subjects to sit within a pre-defined viewing zone, while satisfying the conditions in Section IIB.1 above. If the viewing angle is increased, the reduction of luminance may become severe, and other impairments may arise, such as ‘shear distortion’ [1], [4]. Moreover, viewers may experience different degrees of motion parallax when watching a multiview autostereoscopic display, depending on the viewing angle. To obtain valid subjective results, it is necessary to establish an appropriate viewing region where no such distortions occur.

Fig. 3 depicts two different appearances of the grid with units of 200 mm along the *x*- and *z*-axes. In Fig. 3(a), the grid pattern when the subject is at the center is shown without distortion. When the subject shifts 0.5 m to the left $[(x, y, z) = (-500, 0, 0)]$, the viewing angle is changed. In this case, the perceived grid pattern is distorted as shown in Fig. 3(b). Therefore, the subject will perceive unnatural depth as well as distorted object shapes leading to errors in subjective assessment. Thus, it is necessary to determine optimal viewing positions, where shear distortions are not observed.

We conducted this experiment by having subjects view a 3D test image projected to be seen in front of the center of the stereoscopic display at a viewing distance z_v, as shown in Fig. 4. Detailed descriptions of the viewing environment and the subjects are given in the Appendix. The test image contained a gray square parallelepiped seen as floating in front of a black background. The screen (stereoscopic display) was moved to the left and right at a constant speed of $k = 30$ cm/s ($k = 10$ cm/s) on a rail. We instructed the subjects to fix their eye on the upper surface of the rectangular parallelepiped (square-shaped) during the 10 second display interval. When the subjects began to perceive the shear distortion on the upper surface of the rectangular parallelepiped different from the square, their instruction was to lower their scores. The best viewing region was then obtained in terms of the maximal left and right horizontal displacements in angle of θ_L and θ_R as shown in Fig. 4.

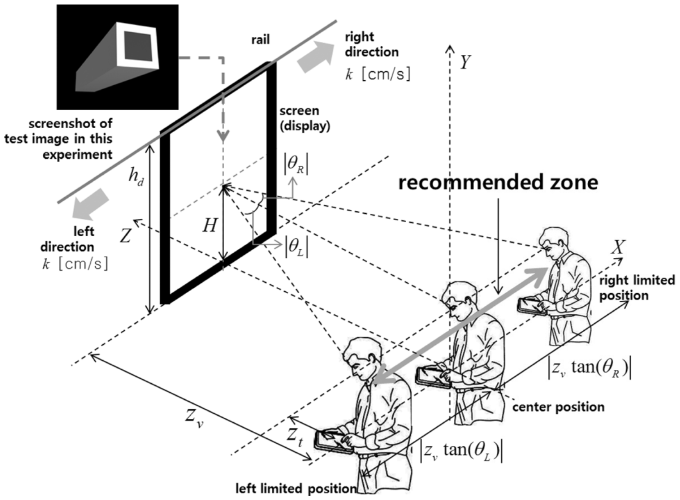
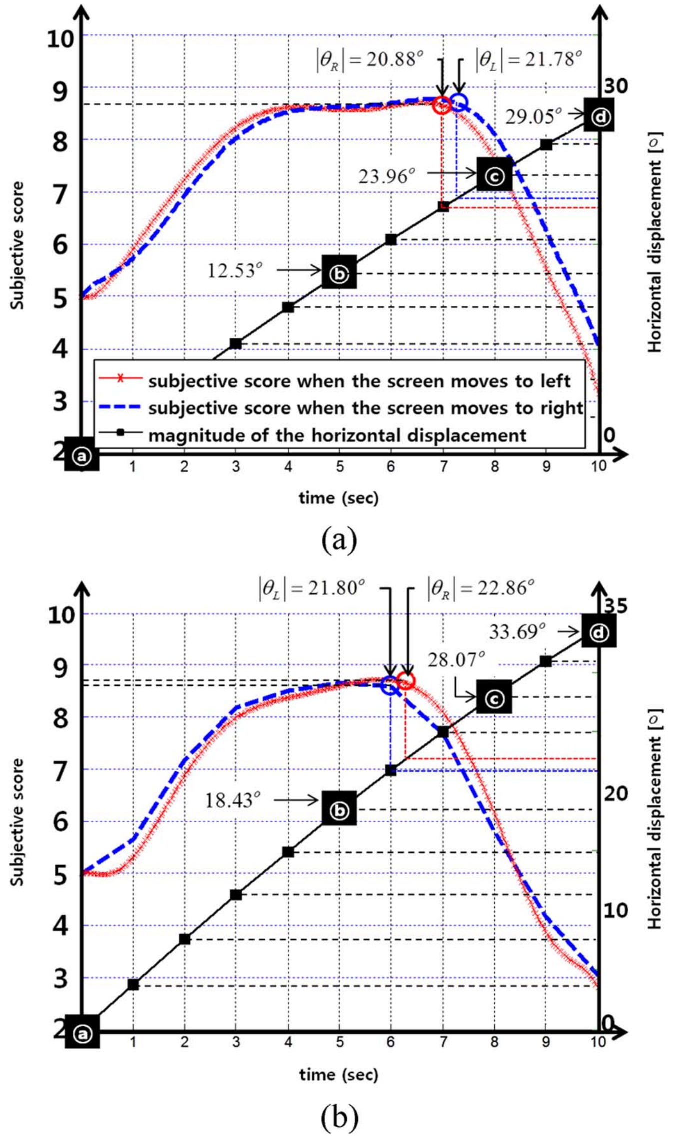


Fig. 4. Recommended viewing zone for multi-user assessment and experimental environment for measuring left and right horizontal displacements (θ_L and θ_R) using MICSQ. Subjects assess the shape of a square while fixing their eye on the upper surface of a square parallelepiped for a period of 10 seconds. During this time, the screen (stereoscopic display) moves to the right from the center along a rail (for example). The subjects should indicate changes in shape (to a non-square parallelepiped) by decrementing their scores. The value of θ_L can then be determined from these recorded scores.

The subjective results for a stereoscopic display and a polarized projector are shown in Figs. 5(a) and (b). The subject scores were initialized to five at time 0 sec. When the stereoscopic display moves to the right or left from the subject, as shown in Table II, the subjective scores were found to begin falling at $\theta_L = 21.78°$ and $\theta_R = 20.88°$ (at around 7 sec.), (likewise, for the polarized projector, $\theta_L = 21.80°$ and $\theta_R = 22.86°$ at around 6 sec.).

Subjective quality assessment also relies on viewing position. In order to study this effect, we performed a subjective visual comfort assessment[1] by fixing the screen position (stereoscopic display) and placing 40 subjects at four different viewing positions $a \sim d$ (position a is the center position as shown in Fig. 4. Subjective scores gradually decrease away from the middle position because the subjects experience visual discomfort due to the rapid variation of disparity, as shown in Figs. 6(a) and (b). Since the initial speed of motion did not appear to noticeably cause visual discomfort, the scores shown in both figures show almost no variation during the first second. The subjective scores at the center (position a) and position b decreased from five to

[1] Forty subjects assessed the visual comfort of the 3D video demonstrated in Section IV-A.2 over 10 second intervals (see Fig. 14). Detailed description of the video is given in Section IV-A.2.



(a)

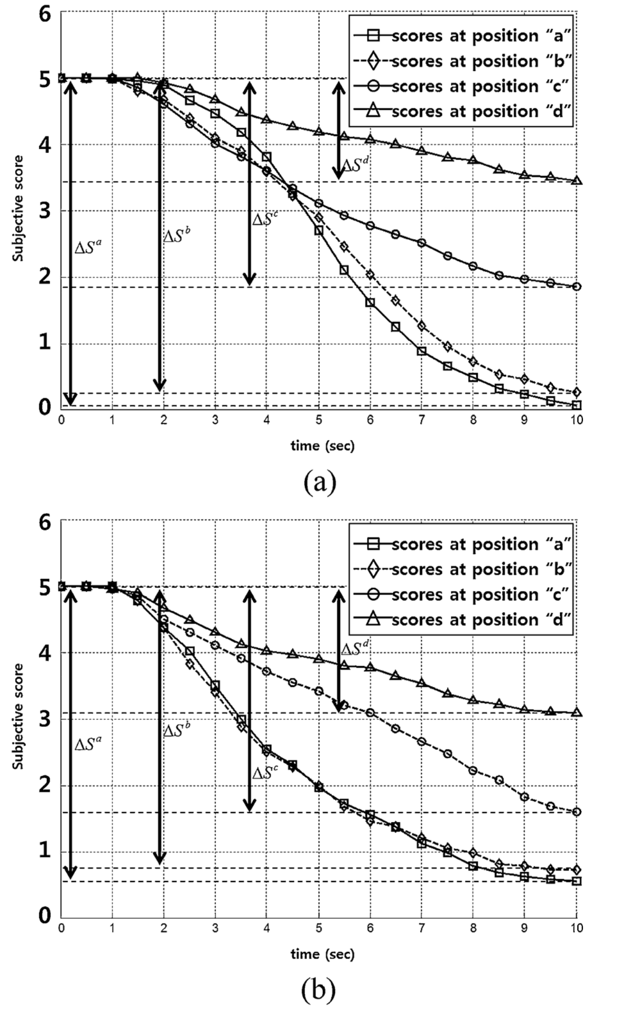
(b)

Fig. 5. Human scores as a function of off-axis viewing angle. When the subjects began to perceive the shear distortion in accordance with the magnitude of the horizontal displacement (black curve), their instruction was to lower their subjective scores (y-axis). (a) Stereoscopic display. (b) Polarized projector.

TABLE II
OPTIMAL VIEWING ZONE FOR MULTI-USER ASSESSMENT IN MICSQ

	Stereoscopic display	Autostereoscopic display	Polarized projector
$\theta_{\mathbf{L}}$	$\lvert\theta_L\rvert \leq 21.78°$	pre-defined viewing zone	$\lvert\theta_L\rvert \leq 21.80°$
$\theta_{\mathbf{R}}$	$\lvert\theta_R\rvert \leq 20.88°$		$\lvert\theta_R\rvert \leq 22.86°$

zero, while the subjective scores at viewing positions c and d only decreased to two and three, respectively. Thus, subjects located outside of the optimal viewing zone experience a very different 3D QoE compared to those at the center position. These experiences may increase or decrease with viewing position but the subjective score at the center should be a basis for the assessment result. Therefore, if the subjective scores at one viewing position are very different from the scores at the center, we may presume that the assessment was performed outside of the optimal viewing zone.



(a)

(b)

Fig. 6. Subjective assessment of the 3D video in Section IV.A.2 at four different viewing positions (in the right direction) ($a \sim d$ in Fig. 5). The viewing positions of a and b fall within the optimal viewing zone in Fig. 5 but the positions of c and d are out of the zone. (a) Stereoscopic display. (b) Polarized projector.

TABLE III
VALUES OF $\triangle S^i$ AND $\triangle S^i/\triangle S^a$ AT EACH VIEWING POSITION

		$\triangle S^i$		$\triangle S^i/\triangle S^a$	
		Left	Right	Left	Right
Stereoscopic display	**a**	4.881	4.937	100%	100%
	b	4.653	4.742	95.33%	96.05%
	c	3.084	3.154	63.18%	63.89%
	d	1.981	1.558	40.59%	31.56%
Polarized projector	**a**	4.503	4.443	100%	100%
	b	4.333	4.276	96.22%	96.24%
	c	3.511	3.402	77.97%	76.57%
	d	2.309	1.905	51.28%	42.88%

In order to compare subjective scores at each viewing position ($b \sim d$) and at the center (a), as shown in Fig. 6, we denote the magnitude of the variation from the beginning to the end of the subjective score at each viewing position i ($i = a, b, c, d$) by $\triangle S^i$. We measure similarity between the value of $\triangle S^a$ and the other values ($\triangle S^b$, $\triangle S^c$, $\triangle S^d$) by computing the ratio between them. As shown in Table III, in both displays, the ratio between the values of $\triangle S^a$ and $\triangle S^d$ is significantly different (by about 41% and 51% in the left direction and by about 32% and 43% in the right direction). Subjective results not given from within

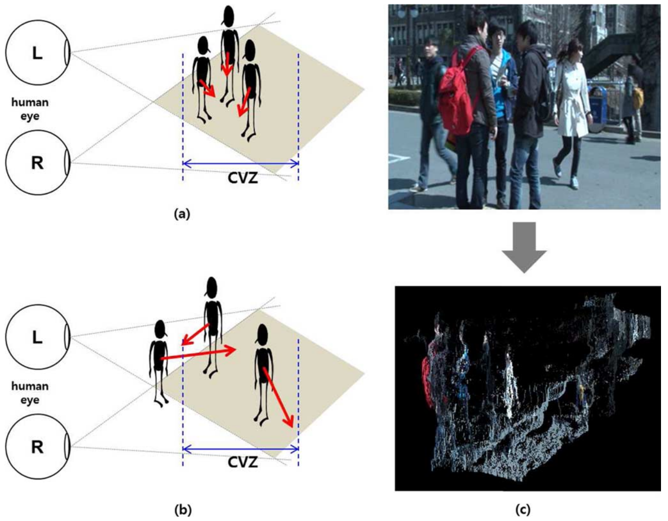


Fig. 7. Spatial and temporal complexity of 3D video: Three persons are standing with (a) low depth distribution and low motion variance, and with (b) high depth distribution and high motion variance. (c) Rendering of the depth distribution.

the recommended viewing zone diverge significantly from those obtained within the recommended viewing zone, as shown in Table III.

III. Reliability of MICSQ Based on the Analysis of Spatio-Temporal Complexity

A. 3D Spatial Complexity

The magnitude and distribution of depth information can be used to characterize 3D spatial complexity. If multiple objects have large disparities, or are distributed throughout an unusually broad disparity range, the stereo fusion process may be rendered incomplete. For example, in Fig. 7(a), three persons are standing very close to one another within the CVZ, so the fusion process should occur easily. By contrast, in Fig. 7(b), the three persons are standing further from each other, so the probability of falling outside of the CVZ may increase, leading to difficulty in fusion. The distribution of depth information can be obtained using depth estimation reference software (DERS) [32], as depicted in Fig. 7(c). Then, the 3D spatial complexity of the nth frame is defined by

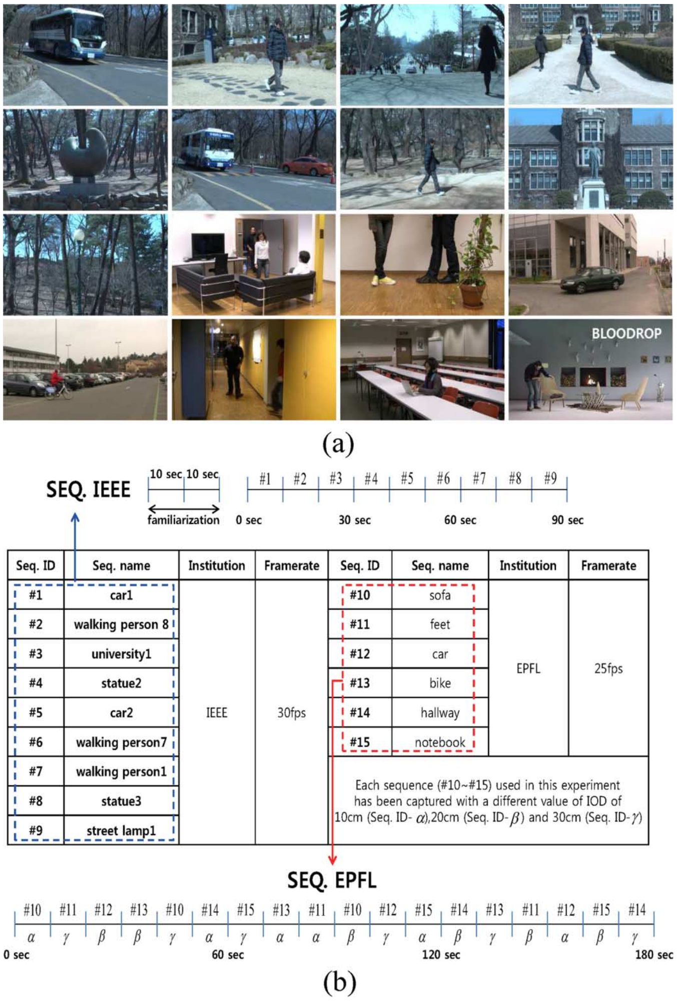
$$f_s(n) = \alpha \cdot \mathrm{mean}_{(u,v)} \left\{ D_s^{(u,v)}(n) \right\} + \beta \cdot \mathrm{std} \left\{ D_s^{(u,v)}(n) \right\} \tag{5}$$

where the mean and standard deviation (std) are computed over all 2D spatial coordinates (u, v) in the nth frame, $D_s^{(u,v)}(n)$ is a screen disparity at each location and α and β weight the relative importance of the mean and variance of the disparity.[2]

B. Temporal Complexity

Temporal complexity also affects visual comfort. It can be quantified by measuring the variability of motion along the x-, y-, and z-axes. If there is a high variability along the direction of multiple moving objects, then viewers may experience increased visual discomfort. Denote a motion vector computed at each location (u, v) in the nth frame by $V^{(u,v)}(n) = (V_x^{(u,v)}(n), V_y^{(u,v)}(n), V_z^{(u,v)}(n))$. Then the velocity along the $x-, y-$, and z-axes is $V_x^{(u,v)}(n)$, $V_y^{(u,v)}(n)$ and $V_z^{(u,v)}(n)$, respectively. Hence, the temporal complexity can be defined by

$$f_t(n) = \mathrm{mean} \left\{ \mathrm{std}\left(V_x^{(u,v)}(n) \right), \mathrm{std}\left(V_y^{(u,v)}(n) \right), \mathrm{std}\left(V_z^{(u,v)}(n) \right) \right\}. \tag{6}$$



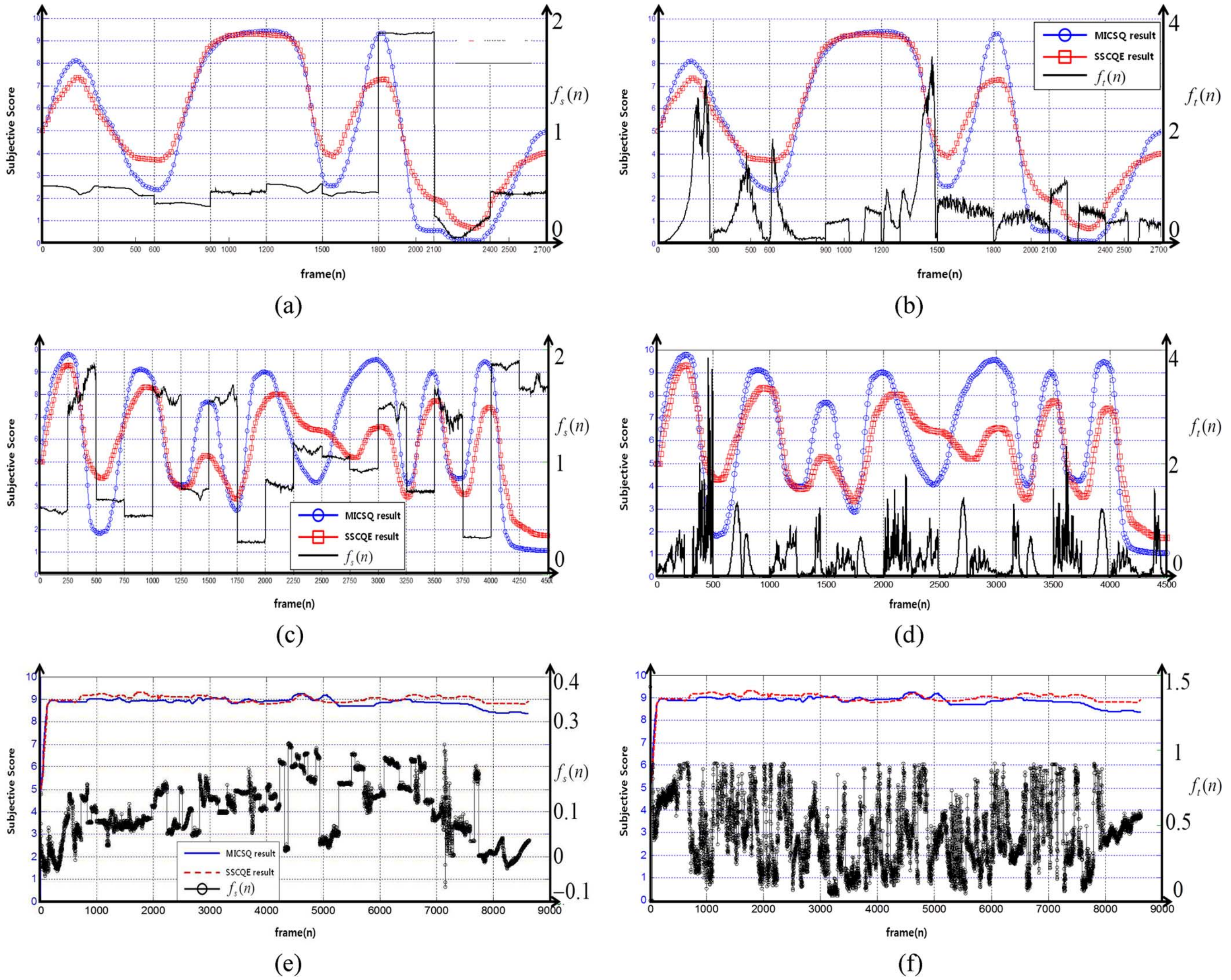
Seq. ID	Seq. name	Institution	Framerate	Seq. ID	Seq. name	Institution	Framerate
#1	car1	IEEE	30fps	#10	sofa	EPFL	25fps
#2	walking person 8			#11	feet		
#3	university1			#12	car		
#4	statue2			#13	bike		
#5	car2			#14	hallway		
#6	walking person7			#15	notebook		
#7	walking person1			Each sequence (#10~#15) used in this experiment has been captured with a different value of IOD of 10cm (Seq. ID- α),20cm (Seq. ID- β) and 30cm (Seq. ID- γ)			
#8	statue3						
#9	street lamp1						

Fig. 8. The two test sequences (“SEQ. IEEE” and “SEQ. EPFL”) composed of test sequences from the IEEE and EPFL 3D video databases and the short movie “BLOODROP”. (a) Screenshots of 3D sequences. (b) Composition orders of “SEQ. IEEE” and “SEQ. EPFL”.

C. Stimuli

We conducted subjective visual comfort assessment experiments on two test sequences which are composed of multiple sequences obtained from 3D video databases made available by IEEE [16] and EPFL [21]. Moreover, one short movie named “BLOODROP” was also used. The sequence “SEQ. IEEE” is composed of nine sequences from the IEEE 3D video database having a frame rate of 30 fps and a length 90 seconds, while the sequence “SEQ. EPFL” is composed of eighteen sequences from the EPFL 3D video database at a frame rate of 25 fps and a length 180 seconds. The sequence “BLOODROP” is a 6 minute

[2] For simplicity, we set $\alpha = \beta = 1$.



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 9. Subjective 3D visual comfort assessment results using both MICSQ and standard SSCQE on “SEQ. IEEE”, “SEQ. EPFL” and “BLOODROP” as a function of 3D spatial complexity $f_s(n)$ and temporal complexity $f_t(n)$. (a) Results of “SEQ. IEEE” with $f_s(n)$. (b) Results of “SEQ. IEEE” with $f_t(n)$. (c) Results of “SEQ. EPFL” with $f_s(n)$. (d) Results of “SEQ. EPFL” with $f_t(n)$. (e) Results of “BLOODROP” with $f_s(n)$. (f) Results of “BLOODROP” with $f_t(n)$.

movie captured at a frame rate of 24 fps. Screenshots from the composed and “BLOODROP” video sequences along with descriptions of the composite sequences “SEQ. IEEE” and “SEQ. EPFL” are described in Fig. 8.

D. Procedure

The subjective assessment experiments were conducted under constant room and background illumination conditions [23]. Fifty-six subjects were asked to assess the degree of visual comfort experienced when viewing the above three stereoscopic videos using both the MICSQ and SSCQE protocols. SSCQE was conducted using a mouse driving a cursor that each of the 40 subjects used to adjust the rating scale, which was overlaid on the same display as the test sequences [33]. Moreover, a touch-screen slider that was not overlaid on the same display was also used by 16 subjects to record the degree of visual comfort [7], [14]. A tablet-pc was used for the slider, and multimodal protocols were not utilized in this study.

In addition, we also conducted an ACR procedure and compared the results with those obtained via MICSQ and SSCQE. The subjective visual comfort assessment experiments using MICSQ, SSCQE and ACR were separated by ten day intervals in order to rest the subjects. Furthermore, to familiarize the subjects with each methodology and with the stimuli, two 3D video sequences were displayed before each assessment series. The MPEG 3D TV sequences were captured at a frame rate of 30 fps and are each 10 seconds in length (‘Lovebird1’ and ‘Newspaper’ [34]). After completing each task, each subject was asked to rate their experiences with MICSQ and SSCQE via a questionnaire, as shown in Fig. 12.

E. MICSQ v.s. SSCQE

Fig. 9 depicts the results of the subjective visual comfort assessment experiments on all sequences plotted against $f_s(n)$ and $f_t(n)$. Clearly, the subjective scores obtained using MICSQ and SSCQE show a high degree of correlation with the dynamics of $f_s(n)$ and $f_t(n)$. Note that the subjective scores of

"SEQ. IEEE" decrease markedly when $f_t(n)$ exceeds 1, as shown in the 150th–600th and 1300th–1600th frame ranges in Fig. 9(b). In other words, the subjects feel more visual discomfort during these periods. Likewise, $f_s(n)$ in "SEQ. IEEE" increases dramatically (up to $+2^\circ$, which is out of the CVZ as a rule-of-thumb $|D_s(n)| < \pm 1^\circ$) [1]–[3], [6] during the 1800th–2100th frames, while the subjective scores noticeably decreased. However, the subjective scores delivered using both MICSQ and SSCQE remained low following the 2100th frame, even when $f_s(n)$ fell to 0°. Because of the dramatic variation in the 3D spatial complexity, the accommodation and vergence processes in the human eye may have struggled to find a new stable state. The subjects may then continue feeling visual discomfort. Alternately, they may have experienced a hysteresis effect of continued dissatisfaction with the QoE even after a return to normal complexity. The subjective scores at times in "SEQ. IEEE", (for example, 900th–1300th frames) are relatively high because both $f_s(n)$ and $f_t(n)$ both fall below 1.

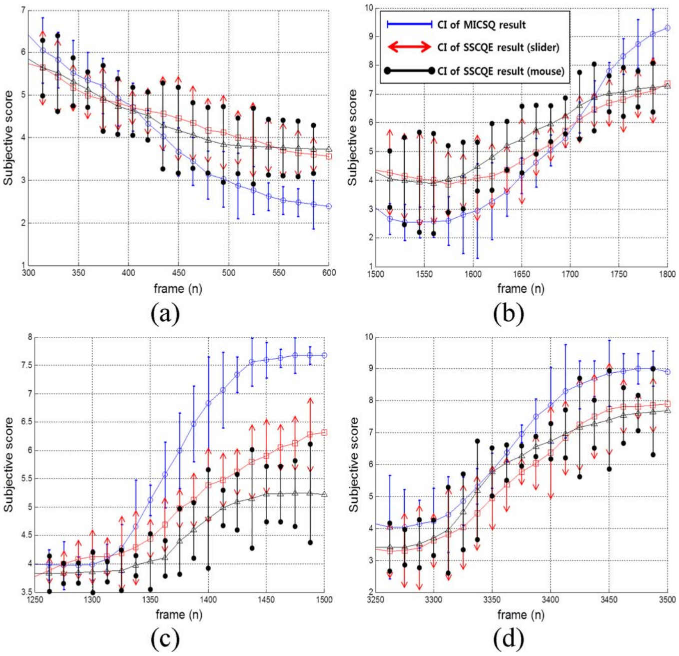
On the other hand, the $f_s(n)$ measurements of "SEQ. EPFL" exhibit higher dynamics than that of "SEQ. IEEE" overall, as shown in Fig. 9(c). The subjective scores obtained using MICSQ and SSCQE decreased in alignment with large values of $f_s(n)$ during the 1000th–1250th, 1500th–1750th, 3000th–3250th, and 4000th–4500th frames, although $f_t(n)$ remained small. The dynamic fluctuation of $f_t(n)$ in "SEQ. EPFL" apparently also influenced the level of reported visual comfort. For example, the reported level of visual comfort decreased during the 2000th–2250th frames where the motion activity was high. Furthermore, the reported visual comfort levels decreased as both $f_s(n)$ and $f_t(n)$ increased during the 250th–500th and 3500th–3750th frame periods.

Unlike the above sequences, the measurements of $f_s(n)$ in "BLOODROP" exhibit larger dynamics (in the range $-0.05^\circ < f_s(n) < 0.25^\circ$) as shown in Fig. 9(e). This range satisfies the CVZ, so that subjective scores obtained using both methodologies exhibit almost no variation over the entire duration. Moreover, the largest measurement of $f_t(n)$ on "BLOODROP" was less than 1, with most of the contribution arising from the temporal complexity along the *x*- and *y*-axes. Therefore, temporal complexity may not have affected visual discomfort when the subjects viewed "BLOODROP", even if $f_t(n)$ exhibited higher dynamics.

Note that the subjective scores obtained using MICSQ and SSCQE varied similarly against $f_s(n)$ and $f_t(n)$. The commercial 3D video (e.g., "BLOODROP") was produced with low spatial and temporal complexities to avoid severe visual discomfort [Figs. 9(e) and (f)] [6]. The use of such a monotonous sequence may prove problematic when attempting to capture discomfort-related characteristics of the 3D visual system. Conversely, the sequences "SEQ. IEEE" and "SEQ. EPFL" were designed to provide clear tests of subjective assessment capability; the reliability of the two methods against human responses did indeed diverge, as we show next.

1) Reliability: When conducting subjective assessment using multiple subjects, the mean opinion score (MOS) is computed as

$$d_k = \frac{\sum_{j=1}^{N} s_{jk}}{N} \tag{7}$$



(a) (b) (c) (d)

Fig. 10. CIs of the (a) 300th–600th, (b) 1500th–1800th frames of "SEQ. IEEE" and the (c) 1250th–1500th, (d) 3250th–3500th frames of "SEQ. EPFL" by using MICSQ, SSCQE-slider and SSCQE-mouse.

where N is the number of subjects and s_{jk} is the score delivered by subject j on the test sequence k. [3] To measure the statistical reliability of the predicted data, we computed confidence intervals (CIs) on the MOS values. Using the MOS of the all subjects, the CI of $(100 \times \alpha)\%$ was computed using the Student's t-distribution

$$\mathrm{CI}_k = t(1 - \alpha/2, N) \cdot \frac{\sigma_k}{\sqrt{N}} \tag{8}$$

where σ_k is the std of a single test condition among the subjects, and $t(1-\alpha/2, N)$ is the t-value with $N-1$ degrees of freedom.[4]

Fig. 10 depicts the CIs of the three subjective methods (MICSQ, conventional SSCQE with a slider (SSCQE-slider) and SSCQE with a mouse (SSCQE-mouse)) at the 300th–600th and 1500th–1800th frames of "SEQ. IEEE" and the 1250th–1500th and 3250th–3500th frames of "SEQ. EPFL". The values of the MOS using SSCQE and MICSQ are quite similar, but the average length of the CI for MICSQ is shorter (approximately 55%) than for SSCQE. However, the difference between the CI of SSCQE-slider and that of SSCQE-mouse is not significant. This means that the subjective scores obtained by SSCQE varied more than those captured using MICSQ. Moreover, this is not due to the change of the assessment tool itself, but mainly due to the multimodal protocols of MICSQ. The scores of the two methods diverged sharply towards the end of the evaluation intervals, as shown in Figs. 10(a) and (c). One possible explanation for this is that decision uncertainty is reduced by neuroplastic adaptation to multimodal information as each video sequence plays out [35]. In addition, to demonstrate the high reliability of MICSQ exhibited on the other sequences, we tabulate the mean length and std of the CIs on

[3] $k = 1$ for "SEQ. IEEE" and $k = 2$ for "SEQ. EPFL".

[4] We set $\alpha = 0.05$ in accordance with a significance level of 95% and $N = 56$ as the number of subjects.

TABLE IV
MEAN LENGTH AND STANDARD DEVIATION OF CIs

	Mean			**Standard deviation**		
	MICSQ	**SSCQE-slider**	**SSCQE-mouse**	**MICSQ**	**SSCQE-slider**	**SSCQE-mouse**
"SEQ. IEEE" ($k = 1$)	1.78	3.04	3.12	0.247	0.466	0.464
"SEQ. EPFL" ($k = 2$)	1.46	2.76	2.88	0.380	0.822	0.512
"BLOODROP"	1.06	1.72	1.34	0.242	0.368	0.255
"Cafeteria"	2.09	3.12	3.36	0.599	0.872	0.989
"Running"	3.08	3.74	3.87	0.958	1.161	1.313
"Street"	1.92	2.78	2.97	0.710	0.883	0.831
Average	**1.90**	**2.89**		**0.523**	**0.745**	

TABLE V
STANDARD DEVIATION AND DYNAMICS OF SUBJECTIVE SCORES

	std($\mathbf{S}_k$)			$\mathbf{V}_k$		
	MICSQ	**SSCQE-slider**	**SSCQE-mouse**	**MICSQ**	**SSCQE-slider**	**SSCQE-mouse**
"SEQ. IEEE" ($k = 1$)	3.147	2.434	2.406	0.234	0.186	0.182
"SEQ. EPFL" ($k = 2$)	2.658	2.096	1.908	0.230	0.094	0.086
"BLOODROP"	0.742	0.598	0.506			
"Cafeteria"	2.140	1.652	1.398			
"Running"	4.056	3.360	2.980			
"Street"	3.331	3.078	2.844			
Average	**2.679**	**2.105**		**0.232**	**0.137**	

the six sequences ("BLOODROP", "Cafeteria", "Running" and "Street"[5] in [16] and the two "SEQ. IEEE" and "SEQ. EPFL" sequences) in Table IV. The mean length and std of the CIs from MICSQ are smaller than those of SSCQE, implying that subjective assessment using MICSQ is more reliable than using SSCQE.

2) Dynamic Representation of Perception: In the process of deciding a proper 3D QoE score, a large variety of information sources, such as binocular disparity, are used as input to the brain where they are interpreted by perceptual and cognitive processes [35]. However, in general, while visual discomfort often accumulates due to the immersion of the viewer, his/her reactions may also slow in proportion to the accumulated visual discomfort. This may result in less accurate and delayed responses. However, we hypothesize that the availability of multimodal cues may reduce unfamiliarity relative to the subject's assessment task, both in terms of perception and cognition. Generally, when recording judgments using MICSQ, subjects should be better able to concentrate on their 3D task(s) with less distraction from the scoring process.

MICSQ allows for a wider dynamic range of perception, while leading the subjects to concentrate on the assessment task longer, with faster responses. Figs. 11(a) and (b) show the improved, higher dynamics in the perceptual responses reported via MICSQ relative to those using SSCQE. In particular, we can observe that the influences of the multimodal cues on human judgements are very significant by comparing the results between MICSQ and SSCQE-slider. In addition, the scores obtained via SSCQE-slider show a slightly wider range than those obtained via SSCQE-mouse. Therefore, the effect on human judgements varies according to the assessment tool.

In order to quantify how much the subjective score varies as a function of the subject's intention, we also measured the dispersion of the subjective scores by each subjective methodology.

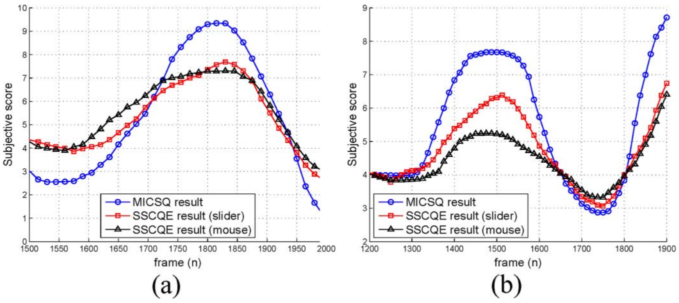


Fig. 11. Subjective results on (a) 1500–2000th frames of "SEQ. IEEE", (b) 1200–1900th frames of "SEQ. EPFL".

Let $\mathbf{V}_k$ be the magnitude of the score variation over one second surrounding the scene change frames:

$$\begin{aligned}\mathbf{V}_k &= \text{mean}(\triangle\mathbf{S}_k/\triangle t) && (9)\\ &= \text{mean}\{|\mathbf{S}_k(n_k) - \mathbf{S}_k(n_k + r_k \cdot 1\ \text{sec})|/1\ \text{sec}\} && (10)\end{aligned}$$

where $\mathbf{S}_k$, n_k and r_k are the subjective score, the index of a scene change frame and the frame rate of the test sequence k respectively.[6] Table V shows the std and $\mathbf{V}_k$ for the six test sequences. Apparently, MICSQ delivers a much more dynamic representation of human response as compared to SSCQE.

3) Performance Comparison: In order to determine what the relationship might be between the ACR of each composed sequence and the corresponding assessment results obtained via MICSQ and SSCQE, 16 subjects conducted an ACR on "SEQ. IEEE" and "SEQ. EPFL". Instead of extracting scores over the entire 10 second sequence duration, only the scores over the last 8 seconds of each sequence were sampled, presuming a similar response time to adjust the slider [3]. The correlations were calculated by using the mean value of 16 sampled MICSQ and

[5] These sequences have a frame rate of 30 fps and a length of 30 seconds. The subjective assessment was conducted by 16 subjects as shown in Appendix A

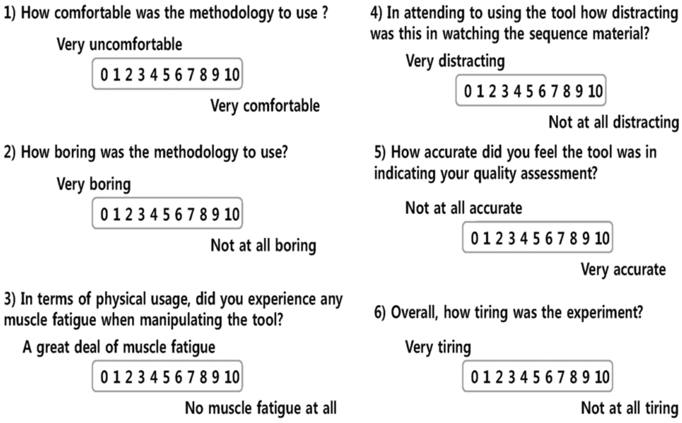
[6] $r_1 = 30$ and $n_1 = 1, \ldots, 8$ for "SEQ. IEEE" and $r_2 = 25$ and $n_2 = 1, \ldots, 17$ for "SEQ. EPFL".

TABLE VI
SCORE OF THE ACR OF NINE COMPOSED SEQUENCES "SEQ. IEEE" AND THE CORRELATIONS BETWEEN THE ACR AND THEIR CORRESPONDING MICSQ AND SSCQE SCORES

	$\#1$	$\#2$	$\#3$	$\#4$	$\#5$	$\#6$	$\#7$	$\#8$	$\#9$	**Average**
Mean ACR score	6.875	7.5	6.813	5.625	5.938	5.125	5.625	4.75	5.5625	**5.979**
Corr. with MICSQ	0.114	0.336	0.374	0.456	0.335	0.321	0.213	0.308	0.255	**0.301**
Corr. with SSCQE	0.137	0.305	0.374	0.455	0.308	0.296	0.188	0.342	0.186	**0.288**

TABLE VII
SCORES OF THE ACR OF EIGHTEEN COMPOSED SEQUENCES "SEQ. EPFL" AND THE CORRELATIONS BETWEEN THE ACR AND THEIR CORRESPONDING MICSQ AND SSCQE SCORES

	$\#10\alpha$	$\#11\gamma$	$\#12\beta$	$\#13\beta$	$\#10\gamma$	$\#14\alpha$	$\#15\gamma$	$\#13\alpha$	$\#11\alpha$	
Mean ACR score	7.938	6.063	6.125	7.813	7.563	5.625	6.063	5.938	5.813	
Corr. with MICSQ	0.232	0.237	0.303	0.298	0.354	0.386	0.286	0.341	0.382	
Corr. with SSCQE	0.201	0.174	0.254	0.263	0.354	0.101	0.209	0.334	0.274	
	$\#10\beta$	$\#12\gamma$	$\#15\alpha$	$\#14\beta$	$\#13\gamma$	$\#11\beta$	$\#12\alpha$	$\#15\beta$	$\#14\gamma$	**Average**
Mean ACR score	4.875	6.188	8.25	5.375	7.25	6	7.75	5.688	5.5	**6.434**
Corr. with MICSQ	0.311	0.362	0.403	0.452	0.355	0.287	0.385	0.207	0.363	**0.330**
Corr. with SSCQE	0.268	0.285	0.294	0.199	0.304	0.244	0.364	0.159	0.328	**0.256**


1) How comfortable was the methodology to use ?
Very uncomfortable
0 1 2 3 4 5 6 7 8 9 10
Very comfortable

2) How boring was the methodology to use?
Very boring
0 1 2 3 4 5 6 7 8 9 10
Not at all boring

3) In terms of physical usage, did you experience any muscle fatigue when manipulating the tool?
A great deal of muscle fatigue
0 1 2 3 4 5 6 7 8 9 10
No muscle fatigue at all

4) In attending to using the tool how distracting was this in watching the sequence material?
Very distracting
0 1 2 3 4 5 6 7 8 9 10
Not at all distracting

5) How accurate did you feel the tool was in indicating your quality assessment?
Not at all accurate
0 1 2 3 4 5 6 7 8 9 10
Very accurate

6) Overall, how tiring was the experiment?
Very tiring
0 1 2 3 4 5 6 7 8 9 10
Not at all tiring

Fig. 12. The questionnaires used in MICSQ and SSCQE.

TABLE VIII
MEAN RESPONSES TO MICSQ AND SSCQE QUESTIONNAIRE

		Q.1	Q.2	Q.3
MICSQ	**Mean**	**9.625**	**8.813**	**9.688**
	Standard deviation	0.619	0.981	0.479
SSCQE	**Mean**	**4.938**	**7.750**	**6.813**
	Standard deviation	2.568	1.483	1.642
		Q.4	Q.5	Q.6
MICSQ	**Mean**	**8.125**	**9.063**	**9.188**
	Standard deviation	0.885	0.998	0.911
SSCQE	**Mean**	**6.563**	**5.250**	**8.813**
	Standard deviation	1.365	1.770	0.981

SSCQE scores (8 seconds sampled at 2 Hz) from each sequence against the ACR scores from 16 subjects.

All of the correlations between the ACR and the MICSQ and SSCQE scores are low, as shown in Tables VI and VII. This tendency of low correlation between a single assessment of a short sequence and a continuous assessment of the same sequence was also found in [3]. The correlations are somewhat higher on the fourth sequence of "SEQ. IEEE", since $f_s(n)$ and $f_t(n)$ are static and the scores from both methodologies are almost flat. Subsequently, the MICSQ scores in '$\#14\alpha$' and '$\#14\beta$' contain much higher dynamic energy than the SSCQE scores. Hence, the correlations between ACR and MICSQ are markedly higher than between ACR and SSCQE. Although the difference between the average correlations of MICSQ and SSCQE is not significant, the correlation between the ACR and corresponding MICSQ results are higher than between SSCQE and the ACR. This difference is more notable on "SEQ. EPFL" than on "SEQ. IEEE" because "SEQ. EPFL" contains higher dynamics than "SEQ. IEEE", as mentioned in Section III-E.2.

After finishing their assessment tasks, each subject was asked to complete a six item questionnaire in which discomfort and distraction experienced when using MICSQ and SSCQE were probed. Answers were collected using an 11-element Likert scale, as shown in Fig. 12. The mean responses were higher for MICSQ than SSCQE for all questions, as shown in Table VIII. Importantly, subjects using MICSQ found the experience to be comfortable and less distracting to the assessment task as compared to SSCQE, as indicated by the responses to Q.1 and Q.4.

IV. STUDY ON EMPIRICAL 3D DISTORTIONS

The high reliability of MICSQ suggests that we can perform more accurate analyses on empirical 3D distortions. For example, for 3D computer graphics applications, it is possible to control the shooting environment (focal length, CCD width, position of cameras, etc.) with a high degree of freedom as functions of parameters of the assessment (viewing) environment (viewing distance, display size, resolution, etc.). Thus, we constructed several 3D computer graphics sequences to use with the viewing and shooting environment parameters depicted in Fig. 13.

If the score is lower than the initial score (5), the subject experiences noticeable visual discomfort, but generally feels satisfactory visual comfort when a score is over 8. Thus, we set a score of 8 as a reasonable threshold for comfortable 3D QoE based on the results of the experiment described in Section IV.A. Using the experimental setup depicted, we conducted three experiments: A) dynamic response of the CVZ as a function of the duration of exposure to uncomfortable stimuli; B) the effects of motion in depth on visual comfort; and C) relating the 3D depth of field to the sensation of naturalness. Table IX outlines these

TABLE IX
OUTLINE OF THE THREE EXPERIMENTS

3D QoE factors	Variable parameters	New discoveries
Visual comfort	Exposure duration of uncomfortable stimuli	CVZ shrinks with increasing time spent viewing an uncomfortable stimuli
	Speed of motion in depth	Visual comfort is affected differently by the speed of motion in depth as a function of forward versus backward direction
Naturalness	Artificial DOF	Effect of artificial blur on naturalness is significant when the diameter of the aperture is similar to the diameter of the cornea

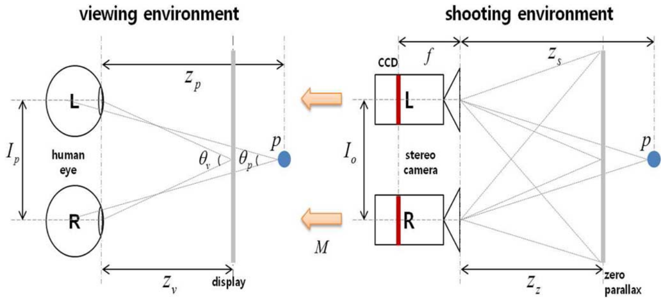


Fig. 13. Viewing and shooting environment parameters when making the computer graphic 3D sequences. **Notation**: I_p—inter-pupillary distance (IPD). I_o—inter-ocular distance (IOD). z_v—viewing distance. z_p—the distance between the two eyes and the object P. z_s—the distance between the camera and the object P. z_z—the distance between the camera and the zero-parallax position. θ_p—the angle between the two eyes and the object P. θ_v—the angle between the two eyes and the projection point to the display of the object P. M—the magnification factor. d_c—the length of the width of the CCD sensor. f—focal length of the camera.

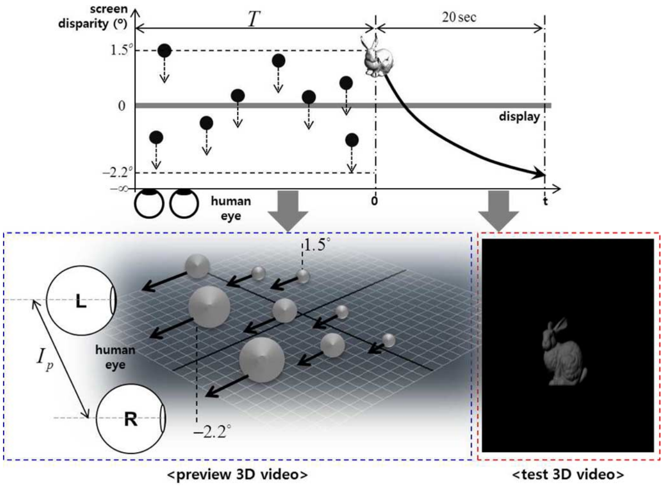


Fig. 14. Temporal evolution of the experiment.

experiments, the effects of variable parameters on two 3D QoE factors, and a summary of new discoveries. Detailed descriptions of the viewing environment and the subjects are given in the Appendix.

A. Comfortable Viewing Zone (CVZ)

1) Objective: The stereoscopic fusion process combines the retinal images from the two eyes into a unified percept. This occurs within a short distance from the horopter (Panum's area). It is difficult to fuse objects outside of Panum's area, and attempting to do so may cause visual discomfort. In general, as rule-of-thumb, for clear and single binocular vision, objects need to lie within the Panum's area while the disparity needs to be smaller than 1°. However, this CVZ may decrease with an increasing duration of exposure to an uncomfortable state [2], [36]. Here, we measure the degree to which the CVZ shrinks as a function of exposure duration when viewing dynamic 3D video.

2) Experimental Set-Up: As shown in Fig. 14, the subjects watched a preview 3D video of duration T ($T = 0$, 300 and 600) before viewing the test sequence. We denote the three experiments as 'EXP I', 'EXP II' and 'EXP III' respectively. In the preview 3D video, multiple 'meteors' moved continuously through a range of screen disparities of 1.5° to $-2.2°$, giving the viewer a sense of flying through space. The initial speed of each meteor was set at 60 cm/s and then increased in increments of 10 cm/s until the speed reached 150 cm/s. When the speed of the multiple meteors was 150 cm/s, it was held constant until the end of the presentation. Subjects viewing this sequence may be expected to experience visual discomfort after approximately 1 second due to the rapid motion in depth, as verified in Fig. 6.

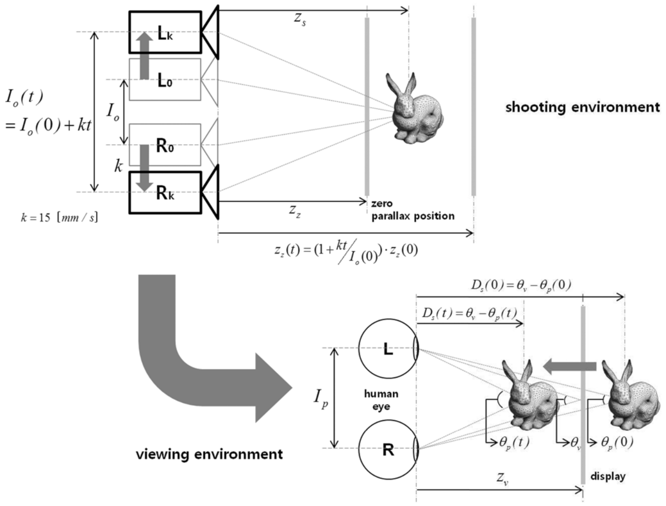


Fig. 15. An object (rabbit) located at the center of the frame is captured by the left and right cameras while linearly increasing the value of the IOD ($I_o(t)$). The rabbit is shown as having the screen disparity ($D_s(t) = \theta_v - \theta_p(t)$) at a viewing distance of z_v.

After watching the 3D preview video over time T, the subjects immediately watched the test sequences while performing the subjective visual comfort assessment task over a period of 20 seconds.

Fig. 15 depicts the shooting environment used to generate the test sequences. A gray rabbit was placed in the center and captured by the stereo camera, while the value of the inter-ocular distance (IOD) was increased according to $I_o(t) = I_o(0) + kt$ over a period of 20 seconds at increments of $k = 15$ mm/s. Then, the distance between the camera and the zero parallax

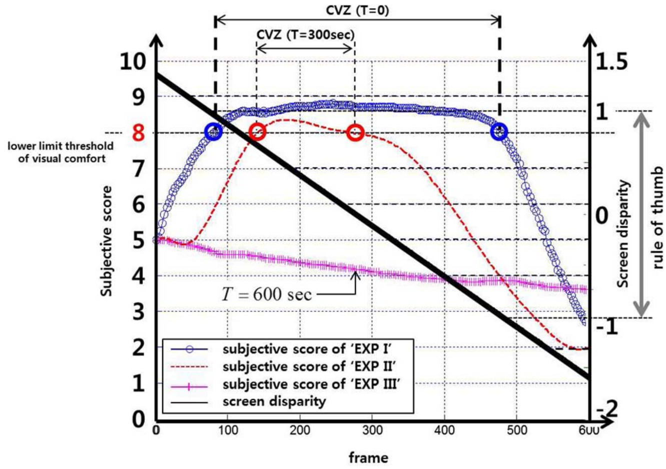


Fig. 16. Subjective visual comfort assessment results. After viewing the preview 3D video ($T = 0, T = 300$ and $T = 600$ secs.), the CVZ was obtained in the range of ($-0.985^\circ \sim 0.789^\circ$, $0.006^\circ \sim 0.730^\circ$ and none, respectively).

position (3D focal point) is varied as $z_z(t) = \{1 + kt/I_o(0)\} \cdot z_z(0)$. It is important that the screen disparity can be decreased without changing the size of the rabbit:

$$\begin{aligned} D_s(t) &= \theta_v - \theta_p(t) \\ &= \theta_v - 2\tan^{-1}\left[\left\{I - fM\left(\frac{1}{z_s} - \frac{1}{z_z(t)}\right) I_o(t)\right\} \Big/ 2z_v\right] \end{aligned} \tag{11}$$

where $M = d_c/w_d$ is the magnification factor. Unlike typical experiments for studying the effect of disparity on visual comfort [11], we can observe the effect of disparity more precisely because the object size remains unchanged and therefore any variation of disparity is thus isolated.

3) Results and Discussion: The CVZ must be determined differently depending on the time spent viewing the preview 3D video, as shown in Fig. 16. In 'EXP I', the CVZ falls in the range $-0.985^\circ < D_s < 0.789^\circ$, which is very similar to the rule-of-thumb. However, in 'EXP II', the CVZ occupies the range $0.006^\circ < D_s < 0.730^\circ$ which is narrower than in 'EXP I'. In addition, the overall subjective score is also lower. However, when the preview 3D video was played for 600 secs., the scores obtained in 'EXP III' monotonically decreased from the beginning of the test sequence. Thus, no CVZ could be found because the score failed to rise above the threshold of eight, which we applied as the guideline for comfortable 3D QoE.

Using MICSQ, the CVZ was accurately determined to be $-0.985^\circ < D_s < 0.789^\circ$, which differs from the rule-of-thumb $-1^\circ < D_s < 1^\circ$. Moreover, as expected, most of the subjects experienced an accumulation of visual discomfort. Table X summarizes the CVZ found in this manner as compared to the conventional rule-of-thumb.

B. Speed and Direction of Depth Motion

1) Objective: The effect of depth motion on visual comfort varies with the direction and speed of motion [11]. Object motion toward the viewer (looming) is more sensitive than motion away from the viewer [8], [38]. Nevertheless, a study of depth motion as a function of direction in depth on visual comfort when viewing 3D content remains a topic worthy of inquiry. The goal of our next experiment using MICSQ was to study the way a subject's feelings of visual discomfort vary with the speed and direction of motion in depth. Specifically, we sought to find a threshold for speed in each direction at which subjects began to feel visual discomfort.

TABLE X
EXPERIMENTAL RESULTS OF THE CVZ

	near $(^\circ)$	far $(^\circ)$
Rule-of-thumb	-1°	1°
'EXP I' $(T = 0)$	-0.985°	0.789°
'EXP II' $(T = 300$ secs.)	0.006°	0.730°
'EXP III' $(T = 600$ secs.)	-	-

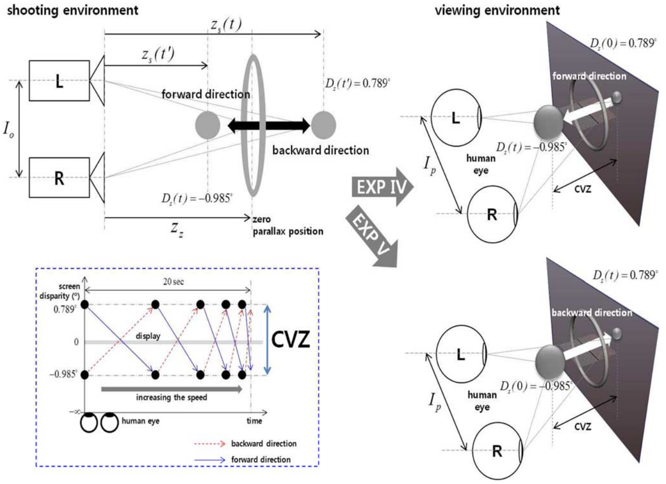


Fig. 17. Detailed content of test sequences. A sphere positioned in the center of the frame moves toward ('EXP IV') and away from ('EXP V') the viewer. The speed of the sphere increased from 15 cm/s to 235 cm/s within the CVZ ($-0.985^\circ < D_s(t) < 0.789^\circ$).

2) Experimental Set-Up: We constructed two 3D computer graphic sequences where a gray sphere is presented in the center of the frame which moves either towards or away from the viewer, as shown in Fig. 17. In order to help the subject accurately perceive the change of depth, a gray ring is positioned at zero disparity surrounding the sphere, all against a black background. Using these two sequences, two experiments 'EXP IV' and 'EXP V' were conducted.

In order to not cause visual discomfort from disparity alone, the object is restricted to movement within the CVZ obtained for 'EXP I' in Table X. First, for the direction towards the viewer ('EXP IV'), the sphere moves from $D_s(0) = 0.789^\circ$ to $D_s(t) = -0.985^\circ$. For the reverse motion ('EXP V'), the sphere moves from $D_s(0) = -0.985^\circ$ to $D_s(t) = 0.789^\circ$. In both experiments, after the sphere reached the end position, it began moving again from the start position. The initial speed was set at 15 cm/s and then increased in increments of 10 cm/s each time the sphere returned to the start position during an overall presentation length of 20 seconds. However, unlike the previous experiment (Fig. 15), the IOD was fixed at $I_o(t) = I_o$. Hence the size of the sphere is gradually increased and decreased in 'EXP IV' and 'EXP V', respectively.

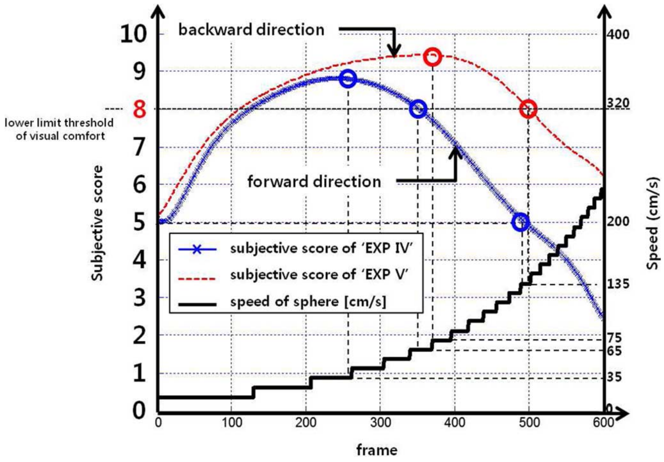


Fig. 18. Subjective visual comfort assessment results. The visual discomfort began at the speed of 65 cm/s (135 cm/s) in the forward (backward) direction.

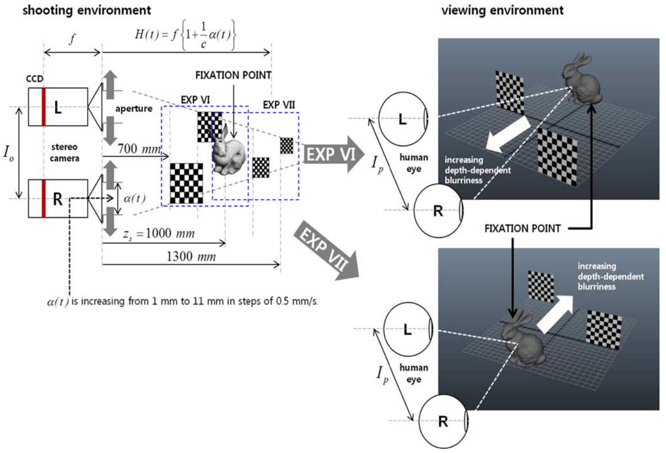


Fig. 19. Content of the test sequences. The rabbit was positioned at the center of the frame and two checkered 'occluders' were presented in front of ('EXP VI') and behind ('EXP VII') the rabbit. The aperture diameter of the camera was increased from 1 mm to 11 mm in steps of 0.5 mm/s.

3) Results and Discussion: For both experiments, the initial score began at 5, and rose to around 9. In 'EXP IV', the subjective score decreased as the sphere in the video reached 35 cm/s as shown in Fig. 18. The scores fell below 8 (our threshold for comfortable 3D QoE), as the speed exceeded 65 cm/s. When the speed exceeded 135 cm/s, the scores fell below the initial score of 5 and most subjects experienced significant visual discomfort. In 'EXP V', the subjective scores began decreasing at a speed of 75 cm/s, falling to around 8 at 135 cm/s, which is twice the speed in the direction towards the viewer. These results strongly suggest that human viewers are more sensitive to looming motion and begin to experience visual discomfort from looming at a lower speed than from motion in the opposite direction. Furthermore, we found that even if the object is displayed within the CVZ, visual discomfort may be induced if the speed in depth is excessive.

C. Relating DOF to Naturalness

1) Objective: The DOF is the amount of retinal defocus whereby accommodation is accomplished such that a scene appears acceptably sharp in the 3D vicinity of the point of gaze [1]. Generally, the value of the DOF is $\pm 0.2D$ around the fixation point [8]. In natural viewing, the human eye perceives objects that are located in front of and behind the focal point as blurred. Thus, if all objects outside the DOF are displayed sharply, it is plausible that human viewers may experience an unnatural sense or even annoyance from the excess of sharpness. Thus, automatic generation of artificial blur has been widely discussed for 3D applications [1]. In order to study the relationship between perceived naturalness and the synthetic DOF, we measured subjects' feelings of naturalness as a function of the aperture diameter of the stereo camera.

2) Experimental Set-Up: A gray rabbit was placed in the center with two neighboring checkered 'occluders' imaged by the stereo camera while the aperture diameter was increased. The occluders were placed in front of (behind) the rabbit and presented in the test sequence for 'EXP VI' ('EXP VII') as shown in Fig. 19. The subjects were instructed to fix their eyes on the rabbit during the experiments. When the value of the aperture was very small, the minute amount of blurring in front of and behind the rabbit gives rise to unnatural viewing. When the value of the diameter was made very large, too much blur also caused unnatural viewing. Unlike prior experiments where a blurring filter was used [39], we implemented blurring more naturally and accurately by controlling the degree of blur in proportion to the distance from the rabbit by increasing the aperture diameter as follows.

Denoting the value of the aperture diameter at time t as $\alpha(t)$, the *f-number* of the aperture of the camera is $N(t) = f/\alpha(t)$. Then, the hyperfocal distance of the camera is

$$H(t) = \frac{f^2}{N(t)c} + f = f\left\{1 + \frac{1}{c}\alpha(t)\right\} \tag{12}$$

where c is the radius of the circle of confusion and f is the focal length.[7] The near and far distances of the DOF in the shooting environment are expressed in terms of z_s and f, respectively:

$$D_N(t) = \frac{f z_s}{c(z_s - f)/\alpha(t) + f} \text{ [mm]} \tag{13}$$

$$D_F(t) = \frac{f z_s}{c(f - z_s)/\alpha(t) + f} \text{ [mm]}. \tag{14}$$

Thus, $D_N(t)$ increases and $D_F(t)$ decreases with increases in $\alpha(t)$ ($\because z_s \gg f$). In both experiments, $\alpha(t)$ was increased from 1 mm to 11 mm in increments of 0.5 mm/s over an interval of 20 seconds.

3) Results and Discussion: Fig. 20 depicts the result of the subjective naturalness assessment study as a function of the variation in the aperture diameter. In 'EXP VI', the subjective score started rising from five at the 64th frame at the value of $\alpha = 2$ mm. The score continuously increased until the maximum value of $\alpha = 7.7$ mm, then decreased rapidly to three. Similarly, as shown in Fig. 20, the subjective score in 'EXP VII' began rising at the value $\alpha = 2.4$ mm, and increasing to around seven until the value $\alpha = 8.05$ mm was reached. Then, the score

[7] In this experiment, we set $c = 0.029$ mm and $f = 35$ mm.

TABLE XI
NUMBER OF SUBJECTS AND THE EQUIPMENT FOR THE EXPERIMENTS AND ASSESSMENTS IN EACH SECTION

		Section II-B.2	Section III	Section IV	Equipment
Single-user assessment	male	-	36	20	46-inch polarized stereoscopic display
	female	-	15	15	Tablet (Samsung Galaxy-Tab)
Multi-user assessment	male	23	3	3	3D projector (BenQ W710ST)
	female	17	2	2	Tablets (Samsung Galaxy-Tab)
Number of subjects		40	56	40	-

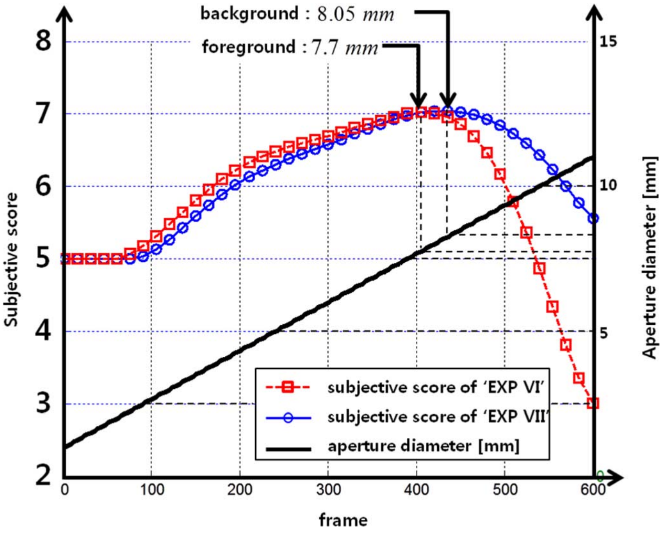


Fig. 20. Subjective 3D QoE (naturalness) assessment result with the value of the aperture diameter.

decreased to 5.5 at a relatively slower rate than in 'EXP VI'. Observe that the obtained value of the aperture diameter when the naturalness was maximum in 'EXP II' was a little larger than in 'EXP VI'. However, when the blur was very significant, the reduced degree of naturalness was very noticeable in 'EXP VI' as compared to 'EXP VII'. This suggests that human viewers are more sensitive to variations in naturalness when foreground objects are present in front of the fixation point, at least as compared to the blurring of background objects.

Although individuals may experience changes in naturalness due to different blurs, our results suggest that a proper degree of blur brings more naturalness to 3D content. In addition, the aperture diameter values of 7.7 mm and 8.05 mm, where the subjective scores were maximized in each experiment, are similar to the diameter of the cornea of the human eye (around 7.8 mm) [40]. Therefore, the effect of artificial blur on naturalness is significant when the diameter of the aperture is similar to the diameter of the cornea. However, when the degree of blur becomes excessive, human viewers begin to feel a sense of unnatural viewing.

V. CONCLUSION

Human subjects experience 3D visualization very deeply. Moreover, they feel visual fatigue due to the accumulation of visual discomfort, complicating the assessment of 3D QoE. Here we proposed a new methodology, named MICSQ, for subjective 3D QoE assessment experiments. Unlike conventional methods, MICSQ utilizes external stimuli such as vibration, flickering and sound to improve human concentration during 3D QoE evaluation. We conducted a number of relevant experiments to verify the utility of the new MICSQ methodology. We also contributed new findings on visual comfort as it relates to disparity and motion, and found an interesting relationship between naturalness and DOF. As advanced techniques for 3D signal processing and demand for 3D content continues to expand, we envision that comprehensive 3D QoE protocols such as MICSQ will prove increasingly valuable. The software of MICSQ can be downloaded from http://insight.yonsei.ac.kr.

APPENDIX
EXPERIMENTAL SET-UP

A. Viewing Environment

To perform single-user subjective 3D QoE assessment experiments using MICSQ, a forty-six inch polarized stereoscopic display with resolution 1920 × 1080 and display height $h_d = 0.6$ m was used. The viewing distance was set to $z_v = 3 \times h_d = 1.8$ m and $H = 1$ m. A Samsung Galaxy-Tab (model: SEC-SHW M180W, $h_t = 0.12$ m) was used as the assessment tool. It was fixed on the table at a distance of $z_t \approx 0.4$ m and $\theta_t \approx \pi/4$ by setting $\theta_e = \pi/4$ and $l = 0.7$ m. The range of subjective scores was set to 0–10 with an initial score of 5, with equally spaced marks [bad]-[poor]-[fair]-[good]-[excellent], following ITU-R Rec. 500–13 [23]. For multi-user subjective assessment, the same viewing environment was used except for an increase in the screen size ($h_d = 1.5$ m) and the viewing distance ($z_v = 4.5$ m) and a 3D projector (BenQ W710ST: $\mathrm{frame\ rate} = 119$ Hz) in Section II.B.2.

B. Subjects

Fifty-six subjects (39 male and 17 female) participated in the assessment study. Fifteen of the subjects are involved in 3D research while the others were naive. The ages of the subjects ranged from 24 to 31 with an average of 27. All subjects were tested and found to have good or corrected visual acuity of greater than 1.25 (the Landolt C-test) and good stereoscopic acuity of less than 60 arc (the RANDOT stereo test). The viewers conducted subjective 3D QoE assessment one hour a day for a month. If the rating score of a subject was found to be much different from the group results, the subject was regarded as an outlier following the rejection procedure in [23]. The number of subjects and the equipment used in the assessment described in each of the preceding sections of this paper are summarized in Table XI.

REFERENCES

[1] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, p. 030201, 2009.

[2] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual comfort," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 335–346, Jun. 2011.

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

[4] S. J. Daly, R. T. Held, and D. M. Hoffman, "Perceptual issues in stereoscopic signal processing," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 347–361, Jun. 2011.

[5] K. Lee, A. K. Moorthy, S. Lee, and A. C. Bovik, "3D Visual activity assessment based on natural scene statistics," *IEEE Trans. Image Process.*, to be published.

[6] F. Zilly, J. Kluger, and P. Kauff, "Production rules for stereo acquisition," *Proc. IEEE*, vol. 99, no. 4, pp. 590–606, Apr. 2011.

[7] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays*, vol. 23, pp. 191–201, 2002.

[8] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, pp. 141–150, 2004.

[9] T. Naganuma, I. Nose, K. Inoue, A. Takemoto, N. Katsuyama, and M. Taira, "Information processing of geometrical features of a surface based on binocular disparity cues: An fMRI study," *Neurosci. Res.*, vol. 51, no. 2, pp. 147–155, 2005.

[10] X. Cao, A. C. Bovik, Y. Wang, and Q. Dai, "Converting 2D video to 3D: An efficient path to a 3D experience," *IEEE Multimedia*, vol. 18, no. 4, pp. 12–17, Apr. 2011.

[11] F. Speranza *et al.*, "Effect of disparity and motion on visual comfort of stereoscopic images," in *Proc. Stereoscopic Displays and Virtual Reality Syst. XIII*, 2006, vol. 6055, pp. 60550B-1–60550B-9.

[12] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vision*, vol. 11, no. 8, pp. 1–29, 2011.

[13] I. Ohzawa, G. C. Deangelis, and R. D. Freeman, "Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors," *Science*, vol. 249, pp. 1037–1041, 1990.

[14] W. IJsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and the feeling of presence in 3DTV," *Displays*, vol. 18, pp. 207–214, 1998.

[15] S. Pastoor, "Human factors of 3DTV: An overview of current research at Heinrich-Hertz-Institut Berlin," in *Proc. IEE Colloq. Stereoscopic Television*, 1992, pp. 11/1–11/4.

[16] *Standard for the Quality Assessment of Three Dimensional (3D) Displays, 3D Contents and 3D Devices Based on Human Factors*, IEEE P3333.1, 2012 [Online]. Available: http://grouper.ieee.org/groups/3dhf

[17] *Image Safety. Reducing the Incidence of Undesirable Biomedical Effects Caused by Visual Image Sequences*, International Standard Organization: IWA3: 2005, 2005.

[18] *Subjective Video Quality Assessment Methods for Multimedia Applications*, ITU-T Recommendation P.910, ITU, Geneva, Switzerland, 2008.

[19] G. Saygili, C. G. Gurler, and A. M. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 593–601, Jun. 2011.

[20] L. Zhang, Q. Peng, Q. H. Wang, and X. Wu, "Stereoscopic perceptual video coding based on just-noticeable-distortion profile," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 572–581, Jun. 2011.

[21] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," *Electron. Imag., 3D Image Process. Applicat.*, 2010.

[22] S. Yang, T. Schlieski, B. Selmins, S. Cooper, R. Doherty, P. Corriveau, and J. Sheedy, "Stereoscopic viewing and reported perceived immersion and symptoms," *Optom. Vis. Sci.*, vol. 89, no. 7, pp. 1068–1080, Jul. 2012.

[23] ITU-R, Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R, Tech. Rep. BT.500-13, 2012.

[24] M. Steffin, Visual-Haptic Interfaces: Modification of Motor and Cognitive Performance [Online]. Available: http://emedicine.medscape.com/article/1136674-overview

[25] D. Jia, A. Bhatti, S. Nahavandi, and B. Horan, "Human performance measures for interactive haptic-audio-visual interfaces," *IEEE Trans. Haptics*, vol. 6, no. 1, pp. 46–57, Jan.–Mar. 2013.

[26] M. Massimino, "Sensory substitution for force feedback in space teleoperation," Ph.D. dissertation, MIT, Dept. Mech. Eng., Cambridge, MA, USA, 1992.

[27] S. Ricciardi *et al.*, "Dependability issues in visual-haptic interfaces," *J. Visual Lang. Comput.*, vol. 21, no. 1, pp. 33–40, Feb. 2010.

[28] T. Barbieri and L. Sbattella, N. Tokareva and B. Kotsik, Eds., UNESCO Institute for Information Technologies in Education, Moscow, Russia, "Providing Enhanced Content Accessibility to Images for Visually Impaired Users," *Specialized Training Course ICTs in Education for People With Special Needs*, pp. 234–235, 2006.

[29] K. Hale, K. Stanney, and L. Malone, "Enhancing virtual environment spatial awareness training and transfer through tactile and vestibular cues," *Ergonomics*, vol. 52, no. 2, pp. 187–203, 2009.

[30] R. J. Stone, "Applications of virtual environments: An overview," in *Handbook of Virtual Environments*. Mahwah, NJ, USA: Lawrence Erlbaum, 2002, pp. 827–856.

[31] L. Kim, M. Han, S. Shin, and S. Park, "Haptic dial system for multimodal prototyping," in *Proc. 18th Int. Conf. Artificial Reality and Telexistence (ICAT 2008)*, 2008.

[32] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," in *ISOIIEC JTClISC29/WGlI MPEG 20081 MI5377*, Apr. 2008.

[33] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[34] ETRI & GIST 3DV Sequences [Online]. Available: FTP://203.253.128.142

[35] C. D. Wickens*, Engineering Psychology and Human Performance*, 2nd ed. New York, NY, USA: Harper Collins, 1992.

[36] G. C. S. Woo, "The effect of exposure time on the foveal size of Panum's area," *Vis. Res.*, vol. 14, no. 7, pp. 473–480, 1974.

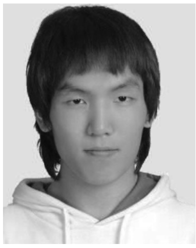
[37] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "On the usefulness of attention for object recognition," in *Proc. 8th Eur. Conf. Computer Vision*, 2004.

[38] T. A. Munch, R. Azeredo da Silveira, S. Siegert, T. J. Viney, G. B. Awatramani, and B. Roska, "Approach sensitivity in the retina processed by a multifunctional neural circuit," *Nature Neurosci.*, vol. 12, no. 10, pp. 1308–1316, 2009.

[39] W. Blohm, I. P. Beldie, K. Schenke, K. Fazel, and S. Pastoor, "Stereoscopic image representation with synthetic depth of field," *J. Soc. Inf. Displ.*, vol. 5, no. 3, pp. 307–313, 1997.

[40] H. Liou and N. Brennan, "Anatomically accurate, finite model eye for optical modeling," *J. Opt. Soc. Amer. A*, vol. 14, no. 8, pp. 1684–1695, 1997.

Taewan Kim received his B.S. degree and the M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea in 2008 and 2010, respectively. He is currently working toward the Ph.D. from 2010. His research interests include quality assessment of 2D and 3D image and video, 3D video coding, cross-layer optimization and wireless multimedia communications. He has participated in the IEEE standard working group for 3D quality assessment (IEEE P3333.1). He was accepted in the Samsung Humantech Thesis Prize in 2013.



Jiwoo Kang received his B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea in 2011. He is currently working toward the joint M.S. and Ph.D. from 2011. His research interests include medical image processing, multimedia processing and GPU programming.

Sanghoon Lee (M'05–SM'12) received the B.S. in E.E. from Yonsei University in 1989 and the M.S. in E.E. from Korea Advanced Institute of Science and Technology (KAIST) in 1991. From 1991 to 1996, he worked for Korea Telecom. He received his Ph.D. in E.E. from the University of Texas at Austin in 2000. From 1999 to 2002, he worked for Lucent Technologies on 3G wireless and multimedia networks. In March 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, Seoul, Korea, where he is a full professor. He has been an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2010–) and an Editor of the *Journal of Communications and Networks* (JCN) (2009–), and the Chair of the IEEE P3333.1 Quality Assessment Working Group (2011–). He served as the General Chair of the 2013 IEEE IVMSP workshop and a guest editor of IEEE TRANSACTIONS ON IMAGE PROCESSING 2013. He has received a 2012 special service award from IEEE Broadcast Technology Society and 2013 special service award from IEEE Signal Processing Society. His research interests include image/video quality assessments, medical image processing, cloud computing, wireless multimedia communications and wireless networks.

Alan C. Bovik is the Curry/Cullen Trust Endowed Chair Professor at The University of Texas at Austin, where he is Director of the Laboratory for Image and Video Engineering (LIVE). He is a faculty member in the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published more than 650 technical articles in these areas and holds two U.S. patents. His several books include the recent companion volumes The Essential Guides to Image and Video Processing (Academic Press, 2009).

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award (2009); the Education Award (2007); the Technical Achievement Award (2005), and the Meritorious Service Award (1998). He also was named recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, received the SPIE Technology Achievement Award for 2012, and was the IS&T/SPIE Imaging Scientist of the Year for 2011. He received the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008), the IEEE Third Millennium Medal (2000) and two journal paper awards from the international Pattern Recognition Society (1988 and 1993). He is a Fellow of the IEEE, a Fellow of the Optical Society of America (OSA), a Fellow of the Society of Photo-Optical and Instrumentation Engineers (SPIE), and a Fellow of the American Institute of Medical and Biomedical Engineering (AIMBE). He has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996–1998; co-founder and Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996–2002; Editorial Board, THE PROCEEDINGS OF THE IEEE, 1998–2004; Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003–present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, Texas, in November, 1994.

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions.