Enhancement of Visual Comfort and Sense of Presence on Stereoscopic 3D Images

Heeseok Oh, Jongyoo Kim, Jinwoo Kim, Taewan Kim, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*— Conventional stereoscopic 3D (S3D) displays do not provide accommodation depth cues of the 3D image or video contents being viewed. The sense of content depths is thus limited to cues supplied by motion parallax (for 3D video), stereoscopic vergence cues created by presenting left and right views to the respective eyes, and other contextual and perspective depth cues. The absence of accommodation cues can induce two kinds of accommodation vergence mismatches (AVM) at the fixation and peripheral points, which can result in severe visual discomfort. With the aim of alleviating discomfort arising from AVM, we propose a new visual comfort enhancement approach for processing S3D visual signals to deliver a more comfortable 3D viewing experience at the display. This is accomplished via an optimization process whereby a predictive indicator of visual discomfort is minimized, while still aiming to maintain the viewer's sense of 3D presence by performing a suitable parallax shift, and by directed blurring of the signal. Our processing framework is defined on 3D visual coordinates that reflect the nonuniform resolution of retinal sensors and that uses a measure of 3D saliency strength. An appropriate level of blur that corresponds to the degree of parallax shift is found, making it possible to produce synthetic accommodation cues implemented using a perceptively relevant filter. By this method, AVM, the primary contributor to the discomfort felt when viewing S3D images, is reduced. We show via a series of subjective experiments that the proposed approach improves visual comfort while preserving the sense of 3D presence.**

***Index Terms*— 3D TV, enhancement, foveation, human factor, optimization, stereoscopic, visual discomfort, visual presence.**

I. Introduction

STEREOSCOPIC 3D (S3D) visual services continue to gain increasing acceptance as a way of providing viewers with an enhanced, immersive experience [1]–[6]. Because of disparities between the left and right eyes, viewers are able to perceive 3D-depth by adapting to and controlling the accommodation and vergence (AV) processes that are responsive to the distribution of depths in a scene. However, immersive 3D multimedia experiences may be accompanied by feelings of physical or visual discomfort or fatigue arising from a variety of factors [2], [4], [6]. Among these, accommodation-vergence mismatches (AVM) are regarded as a primary and difficult to address cause of visual discomfort [9], [10].

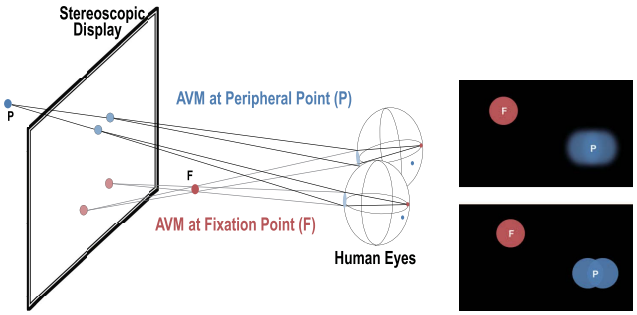

Fig. 1. (a) Viewing geometry of a stereoscopic display, where F and P show fixation and peripheral points, respectively. (b) Stereo view with de-focus blur at P (upper) and without de-focus blur (lower).

When a human with normal binocular vision observes a naturalistic real-world 3D environment, the AV processes, which are the two important oculomotor mechanisms involved in 3D perception, collaboratively drive the vision system towards creating a comfortable sense of clear 3D presence [11]. Vergence cues are easily realized by separating an S3D image pair on the stereoscopic display for capture by the respective eyes by, commonly, alternating polarized left-right views. However, the degree of accommodation needed to achieve a sharp image with an implied focusing depth distribution may not be consistent with the depths implied by vergence on the S3D content. Two distinctive types of AVM may arise:

1) AVM at a fixation point (F): When the eyes converge on a point F, creating a parallax with implied depth, the focus or accommodation of the eyes remains adjusted to the screen depth (Fig. 1 (a)). This discrepancy may induce an AVM at F that falls outside of a comfortable viewing zone (CVZ) [4], [5], [37]. This suggests that visual discomfort could be improved by applying a parallax shift, thereby forcing F into the CVZ [12], [13].
2) AVM at a peripheral point (P): When viewing natural 3D, efforts to merge the two views may be inhibited by de-focus blur at a peripheral point P, since any difference in accommodation between points F and P will cause a natural de-focus blur at P, as shown in Fig. 1 (b) (upper). However, when viewing S3D, a sharp image will be formed at point P (Fig. 1 (b) (lower)) because the focal distance is nearly

Manuscript received August 17, 2016; revised February 24, 2017 and April 11, 2017; accepted April 17, 2017. Date of publication May 9, 2017; date of current version June 7, 2017. This work was supported by the National Research Foundation of Korea through Korea Government under Grant 2016R1A2B2014525. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aljosa Smolic. *(Corresponding author: Sanghoon Lee.)*

H. Oh is with the Center for IT, Wireless Network Laboratory, Yonsei University, Seoul 120-749, South Korea.

J. Kim, J. Kim, T. Kim, and S. Lee are with the Electrical and Electronics Engineering Department, Yonsei University, Seoul 120-749, South Korea (e-mail: slee@yonsei.ac.kr).

A. C. Bovik is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 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/TIP.2017.2702383

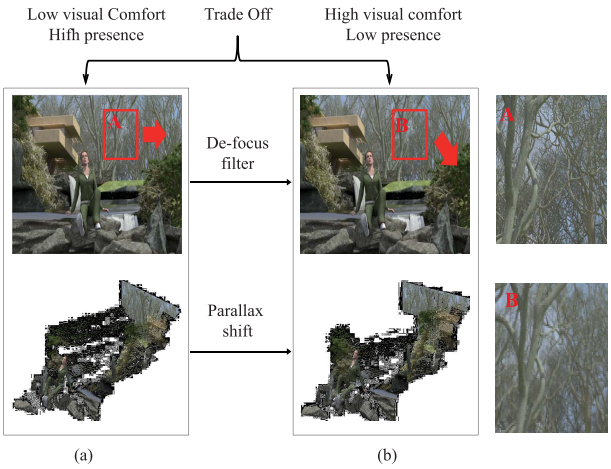

Fig. 2. Trade off between visual comfort and sense of presence. (a) Original image and disparity. (b) Visualization of modified image and disparity by means of de-focus filtering and parallax shifting.

constant over the stereoscopic display. This causes an increased metabolic demand towards achieving acceptable binocular fusion on the sharp data, again causing visual discomfort or fatigue. This suggests that artificial de-focus blur methods can be used to further alleviate AVM in peripheral regions [5], [14]–[16].

However, as shown Fig. 2, there is a trade-off between visual comfort and sense of presence. In Fig. 2 (a), an original image and its disparity are visualized. By modifying the image by means of de-focus filtering and parallax shifting, some reduction of the 3D sense of presence may be expected, but with the benefit of some amount of enhanced visual comfort, as illustrated in Fig. 2 (b). Our aim here is to develop a model of image processing for visual discomfort reduction while maintaining the 3D sense of presence as much as possible.

A key experiential difference between 2D and 3D content is that viewers can feel an increased sense of presence when viewing characters or objects on S3D. Although this sense of presence may greatly affect the 3D viewing experience, methods for modeling and predicting it remain relatively unexplored space [17], [18]. While image processing tools exist that can be used to reduce discomfort felt when viewing S3D content, such as parallax shifting and de-focus filtering, these processes may also reduce the perception of depth, color or texture, leading to degradation of the sense of presence. In order to better balance these effects, we develop a scheme to predict the sense of visual presence as functions of parallax shifting and de-focus filtering.

We quantitatively describe the attribute of visual presence using measurements of the *visual entropies* of color, texture and disparity.

The concept of *visual entropy* has been previously employed to quantify the visual quality of foveated video streams [19], [20], where visual resolution varies according to the fixations of the eyes. The non-uniform variation of visual resolution induced by foveation also affects the sense of presence, and can be predicted using the notion of saliency. Towards capturing this aspect, we deploy a "3D saliency strength" model to quantify the degree of potential visual attraction to a region based on measures of regional visual information [21], [23].

Hence, the sensation of 3D presence is predicted by combining 3D statistical image features descriptive of available visual information and saliency.

To better understand the relationship between experienced visual comfort and presence felt when viewing S3D content, it is beneficial to consider the causes of AVM as well as the statistics of 3D scenes [24]–[28]. Recently, we developed a novel visual comfort prediction framework we dubbed the "3D-AVM predictor" to quantify the degree of visual comfort felt when viewing S3D images, by modeling the neural responses to sensory stimuli that are input to the AV processes [1]. This model uses several features that are predictive of anomalies of the vergence-accommodation/vergence (VA/V) ratio, the absence of de-focus blur, the absence of differential blur and anomalies of the accommodation-vergence/accommodation (AV/A) ratio. We previously showed that the 3D-AVM predictions correlate highly with experienced visual comfort [1].

The main contribution that we make here is the creation of a new visual comfort enhancement method that modifies a given pair of S3D images by processes of parallax shifting and de-focus filtering. As we will show, S3D images processed in this way can be stereoscopically viewed with greater visual comfort, and with little or no loss of the sense of 3D presence as recommended in ITU-R BT.2021 [7]. There exists a trade-off between experienced visual comfort and the sense of presence when the S3D images being viewed are blurred. Using this observation, we define an optimization problem to minimize visual discomfort as predicted by the 3D-AVM predictor, while maintaining the predicted sense of 3D presence by mediating the processes of parallax shifting and de-focus blurring. To achieve the best tradeoff between discomfort and sense of presence, the de-focus blurring filter is designed in accordance with the physiological optics of ocular accommodation. Finally, the physical and cognitive interactions between responses to visual, haptic and aural cues tend to mutually increase the efficacy of each modality [27], [28]. Hence, to demonstrate the performance of the proposed method, we employ a recently developed image assessment methodology, called multimodal interactive continuous scoring of quality (MICSQ) [6], on images drawn from a large public stereo database.

II. Motivation and Overview

Picture quality, depth quality and visual discomfort are the 3 principal factors affecting 3D subjective viewing experience [7]. Our system approach addresses these three factors as depicted in Fig. 3, which shows an overall schematic composed of the following two modules.

1) *3D Content Analysis*: Discomfort predictions delivered by the 3D-AVM predictor, and features predictive of the 3D sense of presence on an S3D image are computed using measurements of texture, color and disparity. The 3D sense of presence is made on 3D visual coordinates, and it incorporates descriptors of foveation, binocular fusion and 3D saliency strength.

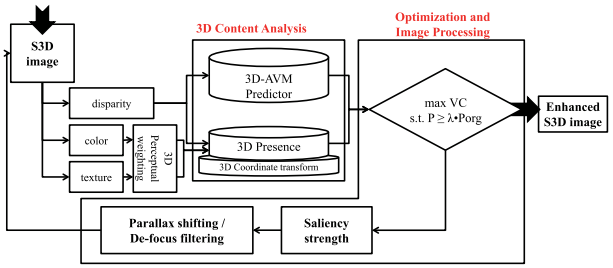

Fig. 3. Proposed visual comfort enhancement of S3D images using parallax shifting and de-focus filtering based on discomfort predictions delivered by the 3D-AVM predictor and features predictive of the 3D sense of presence.

2) *Optimization and Image Processing*: Parallax shifting and de-focus filtering procedures are performed to decrease the level of experienced visual discomfort implied by local 3D-AVM predictor scores, while simultaneously minimizing the loss of sense of presence.

A. 3D Content Analysis

To achieve clear binocular vision in the brain, the interaction between neural processes of accommodation and vergence occurs via inter-operation of two cross-links: AV and VA. However, when people view S3D images on a stereoscopic display, these interactions may be interrupted due to alternations of the cross-link gains, which are regarded as the main cause of visual discomfort experienced when viewing S3D images [1], [15]. In addition, the absence of accommodation depth cues can induce anomalous demands on the fusion process, resulting in visual discomfort. Therefore, in order to quantify the degree of visual discomfort experienced by a viewer of an S3D image, we deploy the successful 3D-AVM predictor to extract four types of features descriptive of anomalies of the VA/V ratio, the AV/A ratio, of de-focus blur and of differential blur [1].

In general, the sense of presence is improved if luminance and color texture is naturally distributed and when the disparity range of a scene is widely distributed. Therefore, measurements of color, texture, and disparity are computed and used to predictive sense of presence. The available visual resolution varies with viewing distance and retinal eccentricity relative to gaze direction. This affects the capture of visual information and the sense of presence. We deploy a foveation model of 3D saliency strength to place limits on the available visual resolution as imposed by foveation and binocular fusion, as explained in section IV. The nonuniform resolution image thus obtained is then mapped onto a uniform domain over curvilinear coordinates. The resulting texture, color and disparity maps are used to define visual entropies that are predictive of the feeling of 3D presence. We also deploy a model of 2D and 3D visual sensitivity, or the degree of perceptibility, based on the predicted visual resolution under an assumed viewing geometry, similar to [3]. Our overall model of the sense of presence is expressed in terms of viewing distance, viewing angle and display type.

B. Optimization and Image Processing

We define and solve an optimization problem aimed at increasing experienced visual comfort, expressed in terms of physiological responses implied by 3D-AVM predictor scores, while minimizing any loss of a sense of 3D presence:

$$\begin{aligned} &\min\ \mathcal{A}\left[\mathbf{F_B}\left(\mathbf{F_S}(o_L, o_R,\ \delta)\right)\right] \\ &\text{subject to } \mathcal{P}\left[\mathbf{F_B}\left(\mathbf{F_S}(o_L, o_R)\right)\right] \geq \lambda \cdot \mathcal{P}(o_L, o_R) \end{aligned} \tag{1}$$

where $\mathcal{A}[\cdot]$ and $\mathcal{P}[\cdot]$ are predictive measurements that correlate with visual discomfort and the sense of 3D presence, respectively. $\mathbf{F_S}(o_L, o_R,\ \delta)$ and $\mathbf{F_B}(o_L, o_R)$ denote the processes of parallax shift and de-focus blur that are applied to the left (o_L) and right (o_R) images of the stereopair to be enhanced, respectively. δ is the amount of parallax shift, and λ ($0 \leq \lambda \leq 1$) is a target threshold on the measurement of 3D presence. The goal is to optimize by balancing "discomfort" vs "presence" via appropriate processes of parallax shift and blur. In order to account for likely regions of visual gaze attraction during the enhancement process, the degrees of the parallax shift and of blur are determined as a function, in part, of the computed 3D saliency strength. In particular, the de-focus filtering used in our model is based on a model of physiological optics (detailed in the following section), so that the simulated blur results in the generation of images having a naturalistic appearance.

III. De-Focus Filtering Using a Point Spread Function

In previous methods of 3D discomfort reduction by the introduction of blur, the degree of blur has been determined absent of any model of physiological optics [14]–[16]. However, when a viewer focuses at a fixation point, the eyes automatically accommodate via changes in the corneal curvature, axial movement of the crystalline lens, changes in axial length, etc [29], [30]. However, among these alterations, changes in the optical power of the crystalline lens are the most significant and relevant [1].

Once the shape of the crystalline lens is determined by the accommodation process, the optics of the eye can be modeled by the thin-lens equation [31]:

$$\frac{1}{s_f} + \frac{1}{d_f} = \frac{1}{f} \tag{2}$$

where f is the focal length, s_f is the posterior nodal distance and d_f is the distance from the nodal point to the fixation point, as shown in Fig. 4. When fixating at a point F, a peripheral point P will not come into perfect focus, resulting in retinal blur. In this case, the blur strength can be characterized by the circle of confusion (CoC), based on the thin-lens model:

$$\beta = r \cdot s_f \left| \frac{1}{d_f} - \frac{1}{d_p} \right|. \tag{3}$$

where β is the diameter of the de-focused point on the retina, d_p is the distance of the peripheral point and r is the pupil diameter, as shown in Fig. 4.

When the eye focuses at a display pixel of width Δ located at distance d_p from the viewer's eye as shown in Fig. 5, its image projects onto the retina with a width $\Delta \cdot s_f / V$ [1], where V is viewing distance. However, if the eye focuses at point F, the pixel width appears wider to a degree determined by the CoC [32].

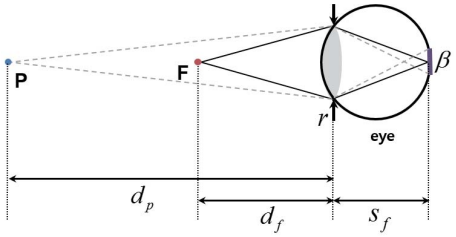

Fig. 4. Assumed geometry of the eye, a fixation point and a peripheral point. A blur circle of diameter of β is created caused by de-focus blur at a peripheral point P when the eye is fixated at point F.

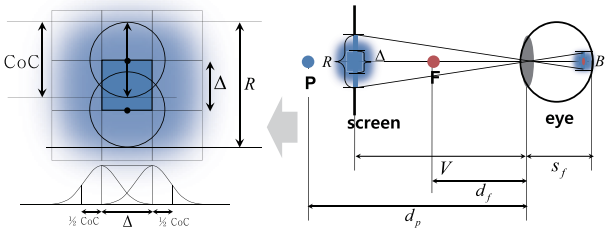

Fig. 5. Circle of confusion (CoC) on the stereoscopic screen (right) and the resulting point spread function (PSF) model of a blurred pixel (left).

To apply the de-focus blur on the retina to the S3D image on the screen, calculate the corresponding blurred pixel width on the screen using the triangles in Fig. 5:

$$R = r \cdot V \left| \frac{1}{d_f} - \frac{1}{d_p} \right| + \Delta \tag{4}$$

The width of the blurred pixels on the screen, (4) is defined by the CoC diameter and the pixel width as depicted in Fig. 5. Since the CoC surrounds the pixel, the pixel is blurred to a degree determined by the CoC diameter, which is a function of the depth distance between the fixated and peripheral points. The point spread function (PSF) describes the response of an imaging system to a point source. The PSF is often approximated by a two-dimensional Gaussian function, where the standard deviation determines the degree of de-focus blur which is proportional to the CoC diameter [32]. Thus, model the PSF of the blurred pixel $p(u, v)$ as:

$$h(u, v) = A_p^{-1} \cdot \exp\left(-\frac{u^2 + v^2}{2\sigma_p^2} \right), \tag{5}$$

$$A_p = \int_u \int_v \exp\left(-\frac{u^2 + v^2}{2\sigma_p^2} \right) du dv \tag{6}$$

where σ_p is the standard deviation of the Gaussian function at pixel p.

To consider the strength of the proposed de-focus filter, foveation and saliency strength kernels are utilized in the PSF. The cut-off frequency of the Gaussian filter is defined by the full-width at half maximum (FWHM) [33]. Thus, σ_p is calculated by finding the diameter of the Gaussian at half-peak, and equating it to R in (4):

$$\sigma_p = \frac{1}{\sqrt{2 \ln 2}} \left(r \cdot V \left| \frac{1}{d_f} - \frac{1}{d_p} \right| + \Delta \right). \tag{7}$$

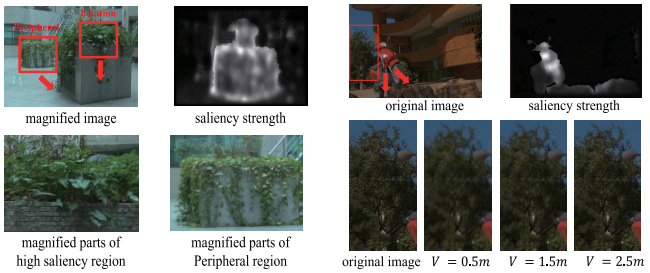

Fig. 6. Example results of de-focus blurring of "ISL5$_{50}$" (left) and the ninth (right) S3D images from the IEEE-SA and IVY databases, respectively. (a) Modified areas according to saliency strength. (b) Modified areas according to viewing distance ($V = 0.5m, V = 1.5m, V = 2.5m$). Output images are obtained using the filter (5).

Fig. 6 depicts two examples showing the results of our method on the "ISL5$_{50}$" image from the IEEE-SA database and on the ninth image from the IVY database [34]. The most salient region was observed on the foreground objects in both images. The other regions were blurred to a degree determined by the spatial distance from the salient object(s) and the viewing distance using the filter in (5).

IV. 3D Content Analysis: Visual Discomfort and 3D Presence

A. 3D-Accommodation Vergence Mismatches

We briefly describe the model employed for S3D visual discomfort prediction, the 3D-AVM predictor (for details, consult [1]). In the nervous system, the sensory and motor processes are integrated to allow a human to take in sensory information and to utilize it to induce motor actions [35], [36]. When a large retinal disparity occurs that exceeds Panum's fusion area, diplopia occurs and vergence eye movements are triggered to fuse the stimulus. Since the sensory stimulus of accommodation is not properly provided to the motor processes when viewing S3D, conflicts between the motor responses can cause sensations of visual discomfort and fatigue [37].

The 3D-AVM predictor uses four kinds of anomalous binocular features to predict visual discomfort, as shown in Fig. 7. AV/A ratio anomalies are related to vergence activity, while anomalies of the VA/V ratio, and absences of de-focus and differential blurs, are related to accommodation activity [1], [39], [40]. The combination of these four features is used to learn a visual discomfort prediction model using a support vector regressor (SVR) trained on human responses to S3D images associated with varying degrees of visual discomfort, as shown in Fig. 7.

B. 3D Sense of Presence

1) 3D Perceptual Weighting by 3D Coordinate Transform: To express the non-uniform visual resolution of S3D images, a method is deployed which incorporates a 3D curvilinear transform weighted by foveation, fusion and saliency factors.

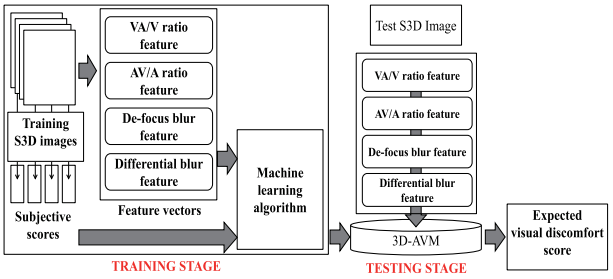

Fig. 7. Procedure for computing 3D-AVM predictor scores on S3D images.

This transformation may be regarded as a model of non-uniform 3D visual resolution mapped onto a uniform 3D domain (see details in [2]).

Suppose there exist coordinate transforms $\mathbb{Y} = (y_1, y_2, y_3)^T$ and $\mathbb{Y}' = (y_1', y_2', y_3')^T$ for $\mathbb{X} = (x_1, x_2, x_3)^T$. If a one-to-one correspondence exists among $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Y}'$, where y_1, y_2, y_3, y_1', y_2', and y_3' are continuous and uniquely invertible, then $\mathbb{Y}$ and $\mathbb{Y}'$ are called 3D curvilinear coordinates.

Let $\mathbb{X}$ denote Cartesian coordinates while $\mathbb{Y}$ and $\mathbb{Y}'$ are 3D curvilinear coordinates computed by processes of foveation filtering $o_f(\mathbb{Y})$ and fusion filtering $o_{f^2}(\mathbb{Y}')$, respectively, as shown in Fig. 8. The original 3D space between the observer and the stereoscopic display $o(\mathbb{X})$ is subjected to a locally band-limiting operation, $o(\mathbb{X}) \rightarrow o_f(\mathbb{Y}) \rightarrow o_{f^2}(\mathbb{Y}')$. The fundamental perceptual factors of 3D saliency, foveation, and fusion at depth are then accounted for as follows.

a) 3D saliency strength: When viewing 2D and 3D content, visual resolution varies with fixation in 3D space, influencing the sense of presence. We developed a successful model of 3D saliency strength that incorporates luminance, contrast, visual discomfort and depth attributes, which is explained in detail in [21] and [22]. This 3D saliency strength model is used here.

b) Foveation coordinate transform: The foveation coordinate transform is used to map a blurred version of an image onto a uniform 2D coordinate system [45]. Given a fixation region, non-uniform gaussian foveation filtering is performed, causing peripheral regions to be blurred, as shown in Fig. 8 (a). The local spatial cutoff frequency in ω_{fov} is defined such that any higher frequency component is less visible or invisible. By setting the maximum possible contrast sensitivity to 1.0, the foveation weight ω_{fov} for both eyes is obtained as

$$\omega_{fov}(V,\ \mathbb{X}) = \min\left\{\frac{e_2 \ln(1/CT_0)}{\alpha(e+e_2)}, \frac{\pi V W}{360}\right\} \tag{8}$$

where $CT_0 = 1/64$ is a minimum contrast threshold, $e_2 = 2.3$ is a half-resolution eccentricity constant, W is the width of the display, and $\alpha = 0.106$ is a spatial frequency decay constant, all determined empirically in [44]. Thus, the relationship between $o(\mathbb{X}) \rightarrow o_f(\mathbb{Y})$ is given by $o_f(\mathbb{Y}) = \omega_{fov}(V, o(\mathbb{X}))$ as exemplified by the image in Fig. 8 (b).

c) Fusion coordinate transform: To model nonuniform resolution in terms of the depth of field based on Panum's

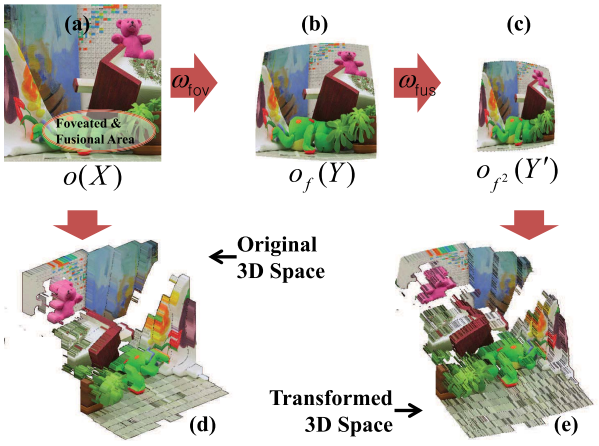

Fig. 8. 3D coordinate transform implementing foveation and fusion. (a) Original image. (b) Foveated coordinate transformation in 2D. (c) Fused image following 2D coordinate transform. (d) Original 3D space. (e) Fused 3D percept following 3D coordinate transform.

fusional area, a fusion weight is defined [46]:

$$\omega_{fus}(\Delta\phi) = \begin{cases} 1, & 0 \le \Delta\phi \le \delta \\ \exp\left(-\frac{\Delta\phi - \delta}{\epsilon}\right), & \delta < \Delta\phi \end{cases} \tag{9}$$

where $\Delta\phi = \phi - \phi_0$, ϕ and ϕ_0 are the angle of vergence for the fixation and another neighboring region, respectively, where δ is a threshold width on the fusional area ($\delta = 0$ **in [21]**) and $\epsilon \simeq 6.2°$ is a fixed coefficient that has been estimated in physiological experiments [2].

Thus, the relationship between $o_f(\mathbb{Y}) \rightarrow o_{f^2}(\mathbb{Y}')$ is given by $o_{f^2}(\mathbb{Y}') = \omega = \omega_{fus}(\omega_{fov}(\mathbb{X}, V))$ in Fig. 8 (c). Consequently, the volume of an S3D image as visualized in 8 (d) is then shrunk, as depicted in 8 (e), by applying the two coordinate transforms.

2) 3D Visual Presence Based on Geometry: We now describe how we account for the affects of the viewing conditions on the predicted sense of presence. As shown in [3], the experienced visual resolution varies with the viewing distance and the viewing angle, further affecting the sense of presence, as follows.

a) 2D visual sensitivity: In [3], we defined 2D visual sensitivity to be the resolving power of the human eye expressed as a function of viewing geometry. Intuitively, when a viewer is further from the display, 2D visual sensitivity decreases. To model this, a model of the spatial contrast sensitivity function (CSF) is used [47]:

$$W_{2D}(f_s) = c_0 \cdot (c_1 + c_2 \cdot f_s) \exp[-(c_2 \cdot f_s)^{c_3}], \tag{10}$$

where spatial frequency f_s has units of cycles/degree, $c_0 = 2.6$, $c_1 = 0.0192$, $c_2 = 0.114$, and $c_3 = 1.1$ [48]. By employing the CSF, the degree of resolving power of the human eye can be estimated. In [3], we analyzed spatial frequency f_s in terms of viewing distance using the concept of display visual resolution (DVR). The DVR *(pixels/degree)* at a given viewing distance V *(m)* is defined by

$$r = \frac{N_x V}{l_x} \cdot \tan\left(\frac{\pi}{180}\right). \tag{11}$$

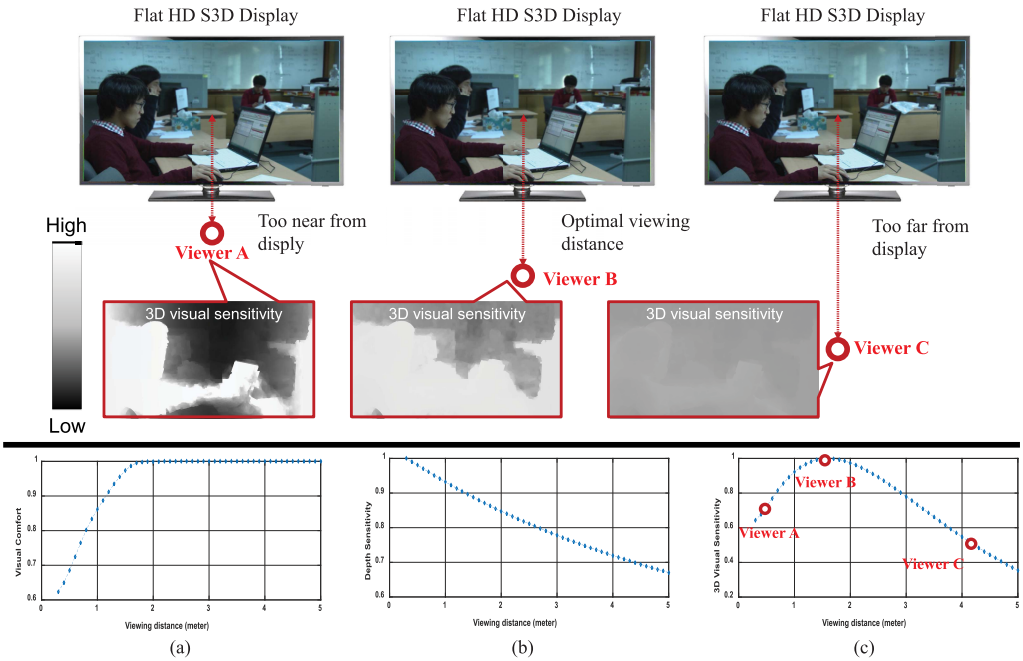

Fig. 9. Variation of 3D sensitivity parameters as function of viewing distance : (a) visual comfort W_{3D}^{D}, (b) depth sensitivity W_{3D}^{S}, (c) 3D visual sensitivity W_{3D}.

where $\frac{N_x}{l_x}$ is the number of pixels per meter. The Nyquist frequency f_s is obtained by halving the maximum sampling rate of the DVR, $f_s = \frac{r}{2}$ *(pixels/degree)*.

b) 3D visual sensitivity: The 3D visual sensitivity is modeled in terms of depth which depends on the viewing position and angular disparity [3]. In our model, to estimate the 3D visual sensitivity (W_{3D}) we use measurements of depth sensitivity, visual discomfort and 2D visual sensitivity. The depth sensitivity (W_{3D}^{S}) decreases as the perceptual depth $D_d = \frac{\alpha_d V^2}{l_e}$ increases, where α_d is angular disparity, V is viewing distance and l_e is the interpupil distance [3]. Thus, the depth sensitivity is expressed as a function of perceptual depth as:

$$W_{3D}^{S} = \frac{1}{D_d + 1}. \tag{12}$$

In order to define 3D visual sensitivity, it is necessary to incorporate angular disparity, which is highly correlated with visual comfort (W_{3D}^{D}). We model the normalized visual comfort as [21]:

$$W_{3D}^{D} = \begin{cases} 1, & \text{if } |\alpha_d| < 1 \\ \dfrac{1}{|\alpha_d|}, & \text{otherwise,} \end{cases} \tag{13}$$

and, as in [3], we define the 3D visual sensitivity W_{3D} as the product

$$W_{3D} = W_{2D} \cdot W_{3D}^{S} \cdot W_{3D}^{D}. \tag{14}$$

In Fig. 9, the upper images illustrate the idea of 3D visual sensitivity as a weight map that varies with viewing distance on a 45 *inch* screen (width: 1.016 *m*, height: 0.5715 *m*) with 1920 × 1080 resolution display. Viewer B is at the optimal viewing distance while viewers A and C are positioned too near and too far from the display, respectively. The 3D visual sensitivity varies with viewing distance, where brighter regions represent more sensitive regions. Viewer A is too near to the display, hence most regions are associated with very low 3D visual sensitivity, except for few objects. When the viewer C is too far from the display, the 3D visual sensitivities also take low values across the image. However, at an optimal viewing distance (viewer B), all regions are viewed with a high 3D visual sensitivity

The graphs at the bottom of in Fig. 9 plot the 3D visual sensitivity against viewing distance.As shown in Fig. 9 (a), the visual comfort weight W_{3D}^{D} is constant; however, this value sharply decreases when the angular disparity falls outside of the aforementioned ZOC (13). Thus, the levels of 3D comfort experienced by viewers B and C appear higher than those of viewer A. In Fig. 9 (b), the depth sensitivity W_{3D}^{S} decreases when the viewing distance increases (12), where the depth sensitivity of viewer A is higher than those of viewers B and C. As a result, W_{3D} in Fig. 9 (c), is maximized at a certain viewing distance, and gradually decreases beyond that viewing distance.

3) Visual Entropy: Here, we use the concepts of visual entropy of texture, color and disparity defined on 3D curvilinear coordinates [23]. In order to capture image texture information, the luminance of a binocular image (e.g. left image) is used to construct a texture map, $\mathbf{M}_T = o_L(u, v)$ where $o_L(u, v)$ are the luminance values at spatial coordinates (u, v) on the original left S3D image. Next, to obtain a color map, low pass and median filtering are applied on each image, thus simplifying and improving the regional clarity of the color components. The color map for each color component Cb and Cr in YCbCr space is denoted $\mathbf{M}_C = \mathbf{F_M}\{\mathbf{F_L}(o_L(u, v)|\mathrm{c})\}$, where $\mathbf{F_M}$ and $\mathbf{F_L}$ denote the median filtering and low pass

TABLE I
FEATURE (TEXTURE, COLOR AND DISPARITY) MAPS OF 3D PRESENCE AND VISUAL ACTIVITY

	Texture	Color	Disparity
Map	$\mathbf{M}_T = o_L(u,v)$	$\mathbf{M}_C = \mathbf{F}_\mathbf{M}\{\mathbf{F}_\mathbf{L}(o_L(u,v)\|_\mathrm{C})\}$	$\mathbf{M}_D = d(u,v)$

filter operations, respectively, and $o_L(\cdot)|_\mathrm{C}$ represents the chromatic components (Cb and Cr) of the original left S3D image. Finally, a disparity map $\mathbf{M}_D = d(u,v)$ is also obtained, where $d(u,v)$ are the angular disparity values at (u,v) computed using the depth estimation reference software [51]. All of the feature maps are summarized in Table I.

In order to apply texture and color maps under a statistical model, the maps are transformed into the wavelet domain. We use the steerable wavelet decomposition over three scales and three orientations [52]. Once each map is transformed, both maps are weighted in the wavelet domain. The visual entropy of each feature map is then calculated using the weighted responses. However, since the sense of 3D presence is strongly affected by perceived depth, disparity entropy is measured over the original curvilinear coordinates by applying the magnitude of angular disparity as a weight on each coordinate.

a) Texture and color: Let $\mathbf{p}_{k,i}$ be the empirical probability mass function (PMF) of the i^{th} subband of factor k over the wavelet domain after applying the visual weights. The factor k represents one of the components, either texture (T) or color (C). Let $\mathbf{B}$ be the set of bins of the histogram of wavelet coefficients and $p_{k,i}(j)$ be the empirical PMF of the j^{th} bin. Then, $\mathbf{p}_{k,i}$ is written as

$$\mathbf{p}_{k,i} = \left\{ p_{k,i}(j) \,|\forall j \in \mathbf{B} \right\}. \tag{15}$$

If $\mathbf{W}_i(\cdot)$ is the wavelet coefficient matrix of the i^{th} subband and $\bar{B}$ is the interval between bins, then the wavelet coefficients $\mathbf{W}_i(\cdot)$ belonging to the j^{th} bin fall in the range $j - \bar{B}/2 \le \mathbf{W}_i(\mathbf{M}_k)(m,n) \le j + \bar{B}/2$. Then, $p_{k,i}(j)$ becomes

$$p_{k,i}(j) = \frac{\sum_{n=1}^{i_n}\sum_{m=1}^{i_m} \Lambda_j(m,n)}{\sum_{n'=1}^{i_n}\sum_{m'=1}^{i_m} \omega^2_{(i_m,i_n)}(m',n')}, \tag{16}$$

where

$$\Lambda_j(m,n) = \begin{cases} \omega^2_{(i_m,i_n)}(m,n), \\ \quad \text{if } j - \bar{B}/2 \le \mathbf{W}_i(\mathbf{M}_k)(m,n) \le j + \bar{B}/2 \\ 0, \quad \text{otherwise} \end{cases} \tag{17}$$

and i_m and i_n are the width and height of the i^{th} subband, respectively, $\omega_{(i_m,i_n)}$ is the visual weight ω of the i^{th} subband and $\omega_{(i_m,i_n)}(m,n)$ is the $(m,n)^{th}$ element of $\omega_{(i_m,i_n)}$. Conceptually, $\omega_{(i_m,i_n)}(m,n)$ $(\omega^2_{(i_m,i_n)}(m,n))$ is the ratio of the length of a unit length element of a 1D-line (2D-area) in Cartesian coordinates to its projection onto curvilinear retinal coordinates at (m,n). If $(m,n)^{th}$ lies in a high-saliency strength region, then $\omega_{(i_m,i_n)}(m,n)$ is relatively higher than in a low-saliency strength region. Therefore, high-saliency strength regions will tend to dominate the computed image statistics. Given these parameters, we define the visual entropy of the i^{th} subband of factor k over 3D visual coordinates as [23].

$$H_{k,i} = -\sum_j W_{2D}(j) \cdot p_{k,i}(j) \log_2(p_{k,i}(j)). \tag{18}$$

b) Disparity: Let $\mathbf{p}_D$ be the empirical PMF of the disparity map $\mathbf{M}_D$:

$$\mathbf{p}_D = \left\{ p_D(j) \,|\forall j \in \mathbf{B} \right\}. \tag{19}$$

Letting $\overline{B}$ denote the interval between bins, the values of the disparity map belonging to the j^{th} bin fall in the range $j - \overline{B}/2 \le \omega(\mathbf{M}_D(u,v)) \le j + \overline{B}/2$. Then, $p_D(j) = \frac{N(j)}{\sum_j N(j)}$, where $N(j)$ is the number of histogram elements of $\omega(\mathbf{M}_D(u,v))$ belonging to the j^{th} bin.

$$H_D = -\sum_j W_{3D}(j) \cdot p_D(j) \log_2(p_D(j)). \tag{20}$$

In the above function, W_{3D} is 3D visual sensitivity. Finally, the measure of the sense of 3D presence $\mathcal{P}$ for an S3D image is

$$\begin{aligned} \mathcal{P} &= \sum_k \tau_k H_k, \quad k \in \{\text{T, C and D}\} \\ &= \sum_k \tau_k \left(\sum_i H_{k,i} \right) + \tau_D H_D, \quad k \in \{\text{T and C}\}. \end{aligned} \tag{21}$$

where τ_k is the parameter used to adjust the relative importances of the three factors that compose the measure of the sense of 3D presence. The value of τ_k is obtained by fitting the value of $\mathcal{P}$ to subjective assessment results on large public databases of S3D images, as described in a later section.

V. IMAGE ENHANCEMENT FOR IMPROVING COMFORT AND SENSE OF PRESENCE

To modify an S3D image so that it can be more comfortably viewed while evoking a more powerful, natural sense of 3D presence, we define a mathematical optimization problem to drive processes of parallax shift and blur. A presumed fixation point needs to be estimated before the processes of depth control and de-focus blur can be applied. In the absence of actual measurement of a viewer's fixation, we instead define the notion of *salient depth* by the value of angular disparity at each point that yields the highest saliency strength value:

$$d_s = \underset{d(u,v)}{\arg\max}\ \mathcal{S}(o_L(u,v)) \tag{22}$$

where $\mathcal{S}(o_L(u,v))$ is the saliency strength values at spatial coordinates (u,v) on the original left S3D image [21]. In other words, the 3D point having the highest saliency strength is taken as the 3D fixation. If the salient depth falls outside the CVZ ($d_s \notin$ CVZ),[1] then the original S3D image can be modified by solving the optimization problem.

[1]The CVZ is set to $-0.985° \le d(u,v) \le 0.789°$ as obtained in [6].

Algorithm 1 Algorithm for Optimizing the Comfort Level of a S3D Image

1: Given d_s, $\mathcal{A}(o_L, o_R)$, $\mathcal{P}(o_L, o_R)$ and CVZ
2: **if** ($d_s \notin$ CVZ) **then**
3: $\delta = 0$
4: **while** $\delta \leq \delta_{\max}$ **do**
5: Obtain o'_L and o'_R using $\mathbf{F_S}(o_L, o_R,\ \delta)$
6: Measure $\mathcal{S}\left(o'_L(u,v)\right)$ corresponding to d'_s of o'_L
7: Obtain o''_L and o''_R using $\mathbf{F_B}(o'_L, o'_R,\ d'_s)$
8: Measure $\mathcal{A}\ (o''_L, o''_R)$ and $\mathcal{P}\ (o''_L, o''_R)$ after measuring $\mathcal{S}\left(o''_L(u,v)\right)$
9: $\delta = \delta + \Delta\delta$
10: **end while**
11: From $\delta = 0$ to $\delta_{\max}$, compare the values of $\mathcal{A}\ (o''_L, o''_R)$
12: Find (o^*_L, o^*_R) with the lowest value of $\mathcal{A}\ (o^*_L, o^*_R)$ under the constraint $\mathcal{P}\ (o^*_L, o^*_R) \geq \lambda \cdot \mathcal{P}(o_L, o_R)$
13: **else**
14: break
15: **end if**

A. Optimization Problem

The optimization problem seeks to minimize the degree of experienced visual discomfort as predicted by the 3D-AVM model while limiting any loss of the sense of 3D presence:

$$\begin{aligned} &\min\ \mathcal{A}\left[\mathbf{F_B}\left(\mathbf{F_S}(o_L, o_R,\ \delta),\ d_s\right)\right] \\ &\text{subject to } \mathcal{P}\left[\mathbf{F_B}\left(\mathbf{F_S}(o_L, o_R,\ d_s)\right)\right] \geq \lambda \cdot \mathcal{P}(o_L, o_R) \end{aligned} \quad (23)$$

where $\mathbf{F_S}(o_L, o_R,\ \delta)$ and $\mathbf{F_B}(o_L, o_R,\ d_s)$ denote the parallax shift and de-focus blur filtering processes applied to the original left and right images, respectively, δ is the amount of parallax shift and λ ($0 \leq \lambda \leq 1$) is the presence threshold.[2]

In Algorithm 1,[3] the pixel disparity is shifted in increments of $\Delta\delta$, but does not exceed Δ_{max}, which depends on the viewing geometry. The saliency strength is also measured. This procedure is repeated until the degree of depth at the fixation point falls within the ZOC. Blur filtering is then performed to obtain an optimal image, as in Fig. 10. We assume a 45 inch display (width: 1.016 m, height: 0.575 m) and viewing distance $V = 1.5$ m, which satisfies the well-known recommendation for an HD display (three times height) [53]. This setting is reasonable to optimize the enhancement level, since the 3D visual sensitivity is indeed maximized near a viewing distance of three times height, as shown in Fig. 9 (c).

Figs. 10 (a) and 10 (b) show examples of the blur and disparity increments, respectively. The average disparity of the processed S3D image gradually decreases at each iteration, where the depths in the high saliency region are reduced by the optimization process in stages 8 and 9 of Algorithm 1. The decrease in disparity is accompanied by a decrease of de-focus blur, as shown in Fig. 10 (a), which corresponds to (7). Fig. 10 (c) visualizes cyclopean images as a function of iteration, showing the evolution of the cyclopean image, the disparity map, and the defocus blur. Figs. 11 (a) and 11 (b) show the evolution of the visual discomfort and 3D sense of presence scores, respectively, as functions of iteration number. In the seventh iteration, our model shows the optimal computed conditions for reducing at a maintained level of 3D presence. Clearly, the predicted degree of visual discomfort is significantly reduced via the processes of parallax shifting and de-focus filtering. The estimated 3D sense of presence score however, is only a little changed. These results demonstrate that the proposed algorithm does indeed largely maintain the predicted sense of 3D visual presence while reducing predicted visual discomfort.

[2]We set $\lambda = 0.8$, which stands for the acceptable sense of presence [5], [13].

[3]We set $\Delta\delta = 0.05°$, and $\delta_{max} = 2°$ in Algorithm 1.

VI. Experiments and Results

We conducted a relatively large-scale subjective study on experienced discomfort and sense of presence felt when viewing S3D images in order to better understand the performance of the proposed method. A total of 920 high-resolution (1920×1080) S3D images were used: 800 S3D images from the IEEE-SA database [54], 100 stereo S3D images from the EPFL database [55], and 20 S3D images from the IVY database [34]. Unlike the first two databases, 20 S3D images from the IVY database were generated by a computer graphics tool with infinite depth of field and also used to test the de-focus blurring schemes against high-precision ground truth.

A. Environment of Subjective Tests

The subjective assessment experiments were conducted in a dark room, compliant with ITU recommendations for the subjective evaluation of picture quality [53]. A 46-inch polarized stereoscopic display with a resolution of 1920×1080 and a display height of 0.6 m was used. The average viewing distance was set to about 1.8 m in accordance with ITU-R Rec. 500-13 [53]. In addition, to familiarize the subjects with the MICSQ methodology [6] and the process of 3D image evaluation, ten sample S3D images were shown to them to allow them to practice scoring before recording actual subjective assessments. The testing sessions were separated by one week in order for the subjects to rest their eyes.

Thirty-seven subjects (27 male and 10 female) participated in the subjective studies. Five of the subjects had been previously involved in 3D research, while the others were inexperienced. The age of the subjects ranged from 21 to 33 years with an average age of 25 years. All of the subjects were tested and found to have good or corrected visual acuity greater than 1.25 (the Landolt C-test) and good stereoscopic acuity of less than 60 arc (the RANDOT stereo test). In the subjective evaluations, on S3D dataset of $920 \times 4 = 3680$ enhanced image pairs was collected. Among these, 800 S3D pairs were obtained from the IEEE-SA database, 100 from the EPFL database, and 20 from the IVY database. These stereopairs were all processed by the four discussed enhancement algorithms, including the model proposed here. We then conducted a series of human subjective tests targeting the experiences of 3D visual discomfort and sense of presence

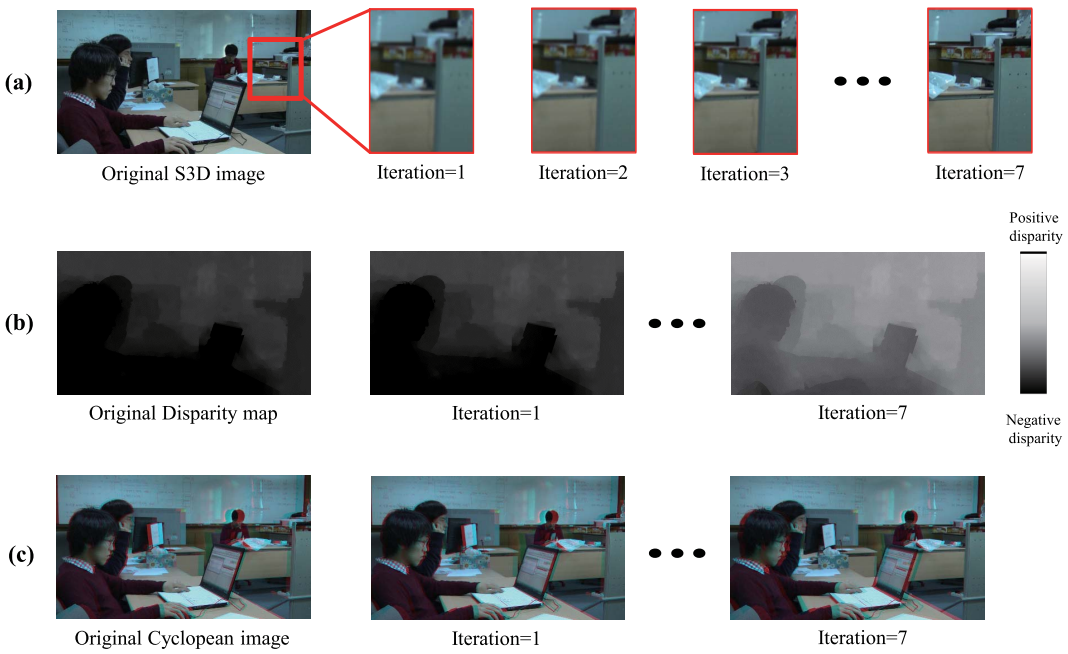

Fig. 10. Optimization visualized as a function of iteration: (a) degree of de-focus filtering on peripheral regions, (b) associated disparity map, (c) associated cyclopean image.

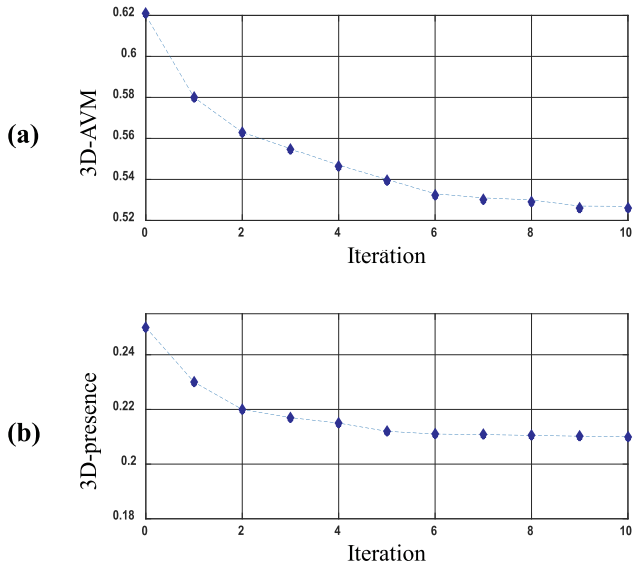

Fig. 11. Plots of (a) visual discomfort scores and (b) 3D sense of presence scores as functions of iteration number.

on the original and processed (comfort enhanced) images. We divided the experiment into 38 separate sessions, including one training session and 37 testing sessions. In the training session, instructions were given regarding the methodology of the test and the general range of both comfort and sense of presence levels that each subject would experience by showing them 20 representative S3D image pairs without labels or comment. The S3D images in the training session broadly spanned the range of parameters in the database. In each testing session, the subjects evaluated no more than

TABLE II
CORRELATIONS BETWEEN 3D PRESENCE FEATURE AND SUBJECTIVE SCORES

	IEEE-SA database		EPFL database	
	Mean	Standard deviation	Mean	Standard deviation
LCC	**0.7812**	0.0385	**0.8799**	0.0436
SROCC	**0.8065**	0.0412	**0.8431**	0.0487
RMSE	**0.2589**	0.0617	**0.2487**	0.0576
OR	**0.0272**	0.0346	**0.0247**	0.0319

100 S3D image pairs (80 S3D image pairs in the last session). The stimuli were randomly shuffled, but no stimuli (enhanced by different methods) was displayed consecutively. The subjects were allowed 10 seconds to evaluate of each S3D pair, thus each testing session required about 17 minutes. A rest period of 10 minutes was allocated between every two consecutive testing sessions. Each subject participated in only four test sessions within one day [1], [20], [57].

B. Performance of 3D Presence Prediction

Each subject was asked to assess the degree of sense of presence experienced after viewing each S3D image for 10 seconds using a Likert-like scale [3]: 5 = very realistic (very strong feeling of presence), 4 = realistic (strong feeling of presence), 3 = mildly realistic (normal feeling of presence), 2 = unrealistic (weak feeling of presence), 1 = very unrealistic (no feeling of presence).

After rejecting outliers following the procedure in [53], the mean opinion scores (MOS) were averaged. We then analyzed the results using two commonly used metrics - the Pearson linear correlation coefficient (LCC) and the Spearman rank-order correlation coefficient (SROCC) [58].

TABLE III

MEAN AND STANDARD DEVIATION OF ENHANCED DISCOMFORT AND PRESENCE MOS VALUES OF THE PROPOSED, JUNG AND LEROY METHODS FOR IMAGES FROM THE IEEE-SA, EPFL, AND IVY DATABASES

Database	Discomfort										Presence									
	Original		Sohn et al.		Leroy et al.		Jung et al.		**Proposed method**		Original		Sohn et al.		Leroy et al.		Jung et al.		**Proposed method**	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
IEEE-SA	3.102	0.296	3.611	0.335	3.135	0.484	3.392	0.396	4.282	0.206	3.861	0.315	3.425	0.295	3.756	0.326	3.355	0.338	3.699	0.298
EPFL	3.056	0.335	3.613	0.371	3.303	0.411	3.584	0.328	4.256	0.224	3.624	0.294	3.248	0.321	3.543	0.375	3.192	0.362	3.501	0.314
IVY	2.877	0.365	3.835	0.312	3.056	0.329	3.503	0.301	4.172	0.274	3.588	0.362	3.224	0.298	3.514	0.296	3.237	0.351	3.522	0.375
Average	**3.012**	**0.332**	**3.686**	**0.339**	**3.165**	**0.408**	**3.493**	**0.235**	**4.237**	**0.216**	**3.691**	**0.324**	**3.299**	**0.305**	**3.604**	**0.332**	**3.261**	**0.350**	**3.574**	**0.329**

The mean values and standard deviations of the correlation scores are tabulated in Table II, which quantitatively demonstrates that the 3D presence prediction scores correlate highly with the MOS on both of the databases. More specifically, the mean LCC against the IEEE-SA and EPFL databases was found to be 0.7812 and 0.8799, respectively, the mean SROCC was 0.8065 and 0.8431, respectively, the root mean squared error (RMSE) was 0.2589 and 0.2487, respectively, and the outlier ratio (OR) was 0.0272 and 0.0247, respectively. These results show that the visual entropy-based 3D presence feature performs commendably as a predictor of the experienced sense of presence when viewing S3D content.

C. Subjective Assessment Results

To gather the MOSs on visual discomfort, each subject was asked to assess the degree of visual discomfort experienced after viewing each S3D image for 10 seconds using a Likert-like scale [1]: 5 = very comfortable, 4 = comfortable, 3 = mildly comfortable, 2 = uncomfortable, 1 = very uncomfortable.

We measured and compared the performance of the proposed method with that of several prior 3D visual comfort enhancement methods proposed by Sohn *et al.* [13], Leroy *et al.* [15], and Jung *et al.* [16]. Sohn *et al.* proposed global and local disparity remapping techniques to ameliorate visual discomfort while maintaining the sense of naturalness. Leroy *et al.* proposed blurring regions of high spatial frequency and large screen disparity. The cut-off frequency for a given discomfort level was deduced from relationships between visual discomfort, disparity and spatial frequency. A recent paper by Jung *et al.* proposed a synthetic DOF blur technique that uses a space-varying Gaussian low-pass kernel based on an estimated visual importance map.

1) MOS Comparisons: Table III tabulates the mean and standard deviations of the subjective scores of visual discomfort and sense of presence on the original images and the enhanced images, respectively. After processing the S3D images using our comfort enhancement model algorithms, the subjective scores were increased by about 1.2 MOS units, which represents a significant improvement in resulting experienced visual comfort as compared to the other two approaches, as shown in Table III. In addition, the standard deviation of the subjective scores on the results of the proposed method was the smallest amongst the compared methods. Generally, our proposed model provided outstanding relative performance.

Several subjects reported that the amount of blur resulting from Jung's method was insufficient to produce comfortable images. One of the drawbacks of Jung's method is that the imposed blurring is not based on physiological or optical factors; rather, all of the coefficients are empirically set for a specific experimental environment [16]. Moreover, because the blur strength was expressed only as a function of the relative disparity and not of the viewing distance or polarity of disparity, the blur strength cannot be optimized for other viewing environments. Conversely, in our method, the amount of blur is determined as a function of the viewing distance and geometrical relationships based on the principles of physiological optics.

Furthermore, observe that the reported sense of presence on the images that were optimized by our model enhancement algorithm was generally not significantly different than on the original images. It is worth noting that the rendered computer graphic images in the IVY database were generally rated as having a relatively lower sense of 3D presence. We might surmise that naturalistic real-world images deliver a stronger sense of 3D presence than do virtual images.

2) Reliability Comparisons: The reliability of the subject results was evaluated using a commonly used measure, known as Cronhach's α, typically deployed as an estimate of reliability in psychometric tests [16], [60]:

$$C_\alpha = \frac{Q}{Q-1}\left|1 - \frac{\sum_{q=1}^{Q}\sigma_{Y_q}^2}{\sigma_X^2}\right|, \tag{24}$$

where Q is the number of test images , σ_X^2 is the variance of the visual comfort scores of each subject, and $\sigma_{Y_q}^2$ is the variance of scores on the q^{th} test image. In general, the value of Cronbach's α increases as a function of the intercorrelations among the test items and is regarded as a method of estimating the reliability of test scores. It is commonly accepted that the internal consistency of a set of test data is good when the value of Cronbach's α is higher than 0.8 [61]. In our case, Cronbach's α was found to be 0.8853.[4]

3) Statistical Significance Comparisons: To investigate the statistical significance of the relative performances of the prior and proposed methods, we performed standard F-tests for discomfort and presence on the MOS values of the original and processed images. The F-tests were conducted on the ratios of variance of both the original and enhanced images' subjective scores for each method, relative to every other one at a 95% significance level [58]. A value of '1' indicates that the statistical performance of the method in the row is superior

[4]Note that a Cronbach's α of 0.7 is often considered acceptable [61].

TABLE IV

RESULTS OF THE F-TEST ABOUT DISCOMFORT AMONG ORIGINAL IMAGE, THE SOHN, LEROY, JUNG AND PROPOSED METHODS

	Original	Sohn *et al.*	Leroy *et al.*	Jung *et al.*	**Proposed**
Original	-	0	-	0	0
Sohn *et al.*	1	-	1	1	0
Leroy *et al.*	-	0	-	0	0
Jung *et al.*	1	0	1	-	0
Proposed	1	1	1	1	-

TABLE V

RESULTS OF THE F-TEST ABOUT PRESENCE AMONG ORIGINAL IMAGE, THE SOHN, LEROY, JUNG AND PROPOSED METHODS

	Original	Sohn *et al.*	Leroy *et al.*	Jung *et al.*	**Proposed**
Original	-	1	1	1	-
Sohn *et al.*	0	-	0	-	0
Leroy *et al.*	0	1	-	1	-
Jung *et al.*	0	-	0	-	0
Proposed	-	1	-	1	-

TABLE VI

NUMBER OF SUBJECTS AND STIMULI IN THE EXPERIMENTS

	Male	Female	Number of Stimuli		
			IEEE-SA	EPFL	IVY
Section IV	19 (4)	6 (1)	800	100	—
Section V	27 (4)	10 (1)	800	100	20
Total Number of Subjects & Stimuli	37 (5)		920		

to that of the method in the column and a symbol value of '0' indicates that the method in the row is statistically inferior to that of the method in the column. A symbol of '-' indicates that the statistical performance of the method in the row is equivalent to that of the method in the column. The results indicate that our proposed model algorithm significantly reduced visual discomfort, as shown in Table IV. With respect to enhancing the sense of presence, the performance obtained by our model was almost equivalent to that of the method of Leroy *et al* and of the original, as shown in Table V.

VII. CONCLUSION

Unlike viewing of 2D images, human beings perceive dynamic levels of visual discomfort when viewing realistic immersive S3D images. Therefore, it is desirable that 3D content should be fine-tuned, either during capture (ideally) or by post-processing (as here) to deliver a minimum level of visual discomfort, but a maximum amount of 3D visual information. Balancing these goals is the challenge addressed in this paper. We envision that the 3D comfort/presence model and algorithm introduced here may play a role in the development of more advanced techniques for 3D picture processing, thereby better meeting increases in demand for high-quality 3D content. In future, the methods described here could be used for perceptually optimized [62] content creation.

APPENDIX: EXPERIMENTAL SET-UP

A. Viewing Environment

The subjective assessment experiments were conducted in a dark room, compliant with ITU recommendations for the subjective evaluation of picture quality [53]. A 46-inch polarized stereoscopic display with a resolution of 1920×1080 and a display height of 0.6 m was used. The average viewing distance was set to about 1.8 m in accordance with ITU-R Rec. 500-13 [53]. In addition, to familiarize the subjects with the MICSQ methodology [6] and the process of 3D image evaluation, ten sample S3D images were shown to each of them to allow them to practice scoring, before recording actual subjective assessments. The testing sessions were separated by one week in order for the subjects to rest their eyes.

B. Subjects

Thirty-seven subjects (27 male and 10 female) participated in the subjective studies. Five of the subjects had been previously involved in 3D research, while the others were inexperienced. The age of the subjects ranged from 21 to 33 years with an average age of 25 years. All of the subjects were tested and found to have good or corrected visual acuity greater than 1.25 (the Landolt C-test) and good stereoscopic acuity of less than 60 arc (the RANDOT stereo test). The number of subjects and the stimuli used in the assessments described in each section are summarized in Table VI.

REFERENCES

[1] J. Park, S. Lee, and A. C. Bovik, "3D visual discomfort prediction: Vergence, foveation, and the physiological optics of accommodation," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 415–427, Jun. 2014.

[2] K. Lee, A. K. Moorthy, S. Lee, and A. C. Bovik, "3D visual activity assessment based on natural scene statistics," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 450–465, Jan. 2014.

[3] H. Oh and S. Lee, "Visual presence: Viewing geometry visual information of UHD S3D entertainment," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3358–3371, Jul. 2016.

[4] 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.

[5] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. CImag. Sci. Technol.*, vol. 53, no. 3, pp. 30201-1–30201-14, May 2009.

[6] T. Kim, J. Kang, S. Lee, and A. C. Bovik, "Multimodal interactive continuous scoring of subjective 3D video quality of experience," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 387–402, Feb. 2014.

[7] J. Li, M. Barkowsky, and P. L. Callet, "Subjective assessment methodology for preference of experience in 3DTV," in *Proc. IEEE 11th IVMSP Workshop*, Jun. 2013, pp. 1–4.

[8] P. Hanhart and T. Ebrahimi, "Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience," in *IS&T/SPIE Electronic Imaging*.International Society for Optics and Photonics, Mar. 2014, p. 90110D.

[9] J. Park, H. Oh, S. Lee, and A. C. Bovik, "3D visual discomfort predictor: Analysis of disparity and neural activity statistics," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1101–1114, Mar. 2015.

[10] T. Kim, S. Lee, and A. C. Bovik, "Transfer function model of physiological mechanisms underlying temporal visual discomfort experienced when viewing stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4335–4347, Nov. 2015.

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

[12] 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.

[13] H. Sohn, Y. J. Jung, S.-I. Lee, F. Speranza, and Y. M. Ro, "Visual comfort amelioration technique for stereoscopic images: Disparity remapping to mitigate global and local discomfort causes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 745–758, May 2014.

[14] Y. Okada, K. Ukai, J. S. Wolffsohn, B. Gilmartin, A. Iijima, and T. Bando, "Target spatial frequency determines the response to conflicting defocus- and convergence-driven accommodative stimuli," *Vis. Res.*, vol. 46, no. 4, pp. 475–484, Feb. 2006.

[15] L. Leroy, P. Fuchs, and G. Moreau, "Visual fatigue reduction for immersive stereoscopic displays by disparity, content, and focus-point adapted blur," *IEEE Trans. Ind. Electron.*, vol. 59, no. 10, pp. 3998–4004, Oct. 2012.

[16] Y. J. Jung, H. Sohn, S.-I. Lee, F. Speranza, and Y. M. Ro, "Visual importance- and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1408–1421, Aug. 2013.

[17] W. IJsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and the feeling of presence in 3DTV," *Displays*, vol. 18, no. 4, pp. 207–214, May 1998.

[18] K. Hopf, "An autostereoscopic display providing comfortable viewing conditions and a high degree of telepresence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 359–365, Apr. 2000.

[19] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[20] J. Park, H. Lee, S. Lee, and A. C. Bovik, "Optimal channel adaptation of scalable video over a multicarrier-based multicell environment," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1062–1071, Oct. 2009.

[21] H. Kim, S. Lee, and A. C. Bovik, "Saliency prediction on stereoscopic videos," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1476–1490, Apr. 2014.

[22] H. Kim and S. Lee, "Transition of visual attention assessment in stereoscopic images with evaluation of subjective visual quality and discomfort," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2198–2209, Dec. 2015.

[23] K. Lee and S. Lee, "A new framework for measuring 2D and 3D visual information in terms of entropy," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 26, no. 11, pp. 2015–2027, Nov. 2016.

[24] Y. Liu, A. C. Bovik, and L. K. Cormack, "Disparity statistics in natural scenes," *J. Vis.*, vol. 8, no. 11, pp. 1–19, Aug. 2008.

[25] Y. Liu, L. K. Cormack, and A. C. Bovik, "Statistical modeling of 3-D natural scenes with application to Bayesian stereopsis," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2515–2530, Sep. 2011.

[26] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Color and depth priors in natural images," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2259–2274, Jun. 2013.

[27] K. S. Hale, K. M. 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, Feb. 2009.

[28] 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, 1st Quart., 2013.

[29] D. A. Atchison, "Accommodation and presbyopia," *Ophthalmic Physiol. Opt.*, vol. 15, no. 4, pp. 255–272, Jul. 1995.

[30] H. Schneider, A. Bacskulin, and R. Guthoff, "History of accommodation research," in *Current Aspects of Human Accommodation*, R. Guthoff and K. Ludwig, Eds. Heidelberg, Germany: Kaden Verlag, 2001, pp. 11–23.

[31] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 4, pp. 523–531, Jul. 1987.

[32] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognit.*, vol. 44, no. 9, pp. 1852–1858, 2011.

[33] A. C. Bovik, *The Essential Guide to Image Processing*. New York, NY, USA: Academic, 2009.

[34] *IVY Lab Stereoscopic Image Database for Synthetic DOF Blur*,. accessed on Dec. 23, 2015. [Online]. Available: http://ivylab.kaist.ac.kr/demo/3D/dofblur.htm

[35] S. J. Huston and V. Jayaraman, "Studying sensorimotor integration in insects," *Current Opinion Neurobiol.*, vol. 21, no. 4, pp. 527–534, Aug. 2011.

[36] R. Shadmehr, M. A. Smith, and J. W. Krakauer, "Error correction, sensory prediction, and adaptation in motor control," *Annu. Rev. Neurosci.*, vol. 33, pp. 89–108, Jul. 2010.

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

[38] Y. Nojiri, H. Yamanoue, A. Hanazato, and F. Okano, "Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV," *Proc. SPIE*, vol. 5006, pp. 195–205, May 2003.

[39] K. Lee and S. Lee, "3D perception based quality pooling: Stereopsis, binocular rivalry, and binocular suppression," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 3, pp. 533–545, Apr. 2015.

[40] J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, "Video quality pooling adaptive to perceptual distortion severity," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 610–620, Feb. 2013.

[41] D. Kim and K. Sohn, "Visual fatigue prediction for stereoscopic image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 2, pp. 231–236, Feb. 2011.

[42] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, May 2000.

[43] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[44] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," in *Photonics West'98 Electronic Imaging*. International Society for Optics and Photonics, Jul. 1988, pp. 294–305.

[45] U. Jang, H. Lee, and S. Lee, "Optimal carrier loading control for the enhancement of visual quality over OFDMA cellular networks," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1181–1196, Oct. 2008.

[46] H. Oh, H. Lee, I. Lee, and S. Lee, "Cooperative content and radio resource allocation for visual information maximization in a digital signage scenario," *Digit. Signal Process.*, vol. 45, pp. 24–35, Oct. 2015.

[47] J. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 525–536, Jul. 1974.

[48] S. J. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," in *Photonics West'98 Electronic Imaging*. International Society for Optics and Photonics, Jul. 1998, pp. 180–191.

[49] I. Lee, J. Park, S. Kim, T. Oh, and S. Lee, "Device-aware visual quality adaptation for wireless N-screen multicast systems," *IEICE Trans. Commun.*, vol. E96-B, no. 12, pp. 3181–3189, Dec. 2013.

[50] H. Kim, J. Kim, T. Oh, and S. Lee, "Blind sharpness prediction for ultrahigh-definition video based on human visual resolution," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 27, no. 5, pp. 951–964, May 2017.

[51] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mor, "Reference softwares for depth estimation and view synthesis," *ISO/IEC JTC1/SC29/WG11 MPEG*, vol. 20081, p. M15377, 2008.

[52] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. Int. Conf. Image Process. (ICIP)*, Washington, DC, USA, Oct. 1995, pp. 444–447.

[53] Assembly, ITU Radiocommunication, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Tech. Rep. BT.500-13, 2012.

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

[55] L. Goldman, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, Feb. 2010, p. 75260S.

[56] A. Benoit, P. L. Callet, and P. Campisi, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, pp. 1–13, Dec. 2009.

[57] H. Oh, S. Lee, and A. C. Bovik, "Stereoscopic 3D visual discomfort prediction: A dynamic accommodation and vergence interaction model," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 615–629, Feb. 2016.

[58] 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.

[59] (2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment*. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtvphaseI

[60] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrik*, vol. 16, no. 3, pp. 297–333, Sep. 1951.

[61] P. Kline, *The Handbook of Psychological Testing*, 2nd ed. London, U.K.: Routledge, 1999.
[62] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 857–872, Jun. 2008.

Heeseok Oh received the B.S., M.S., and Ph.D. degrees in electrical and electronics engineering from Yonsei University, Seoul, South Korea, in 2010, 2012, and 2017, respectively, where he is currently with Samsung Electronics Company Ltd., Suwon, South Korea. His current research interests include 2D/3D image and video processing based on human visual system, machine learning, and computational vision.

Jongyoo Kim received the B.S. degree and the M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree. His current research interests include 2D/3D image and video processing based on human visual system, quality assessment of 2D/3D image and video, 3D computer vision and deep learning.

Jinwoo Kim received the B.S. degree in electrical and electronic from Hongik University, Seoul, South Korea, in 2016. He is currently pursuing the M.S. degree with the Multidimensional Insight Laboratory, Yonsei University. His research interests are in the area of quality assessment, computer vision, and machine learning.

Taewan Kim received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2008, 2010, and 2015, respectively. He is currently with the Video Technology Laboratory of SK Telecom, Seoul. His current research interests include the 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 received the Samsung Humantech Thesis Prize in 2013.

Sanghoon Lee (M'05–SM'12) received the B.S. degree from Yonsei University, Seoul, South Korea, in 1989, the M.S. degree from the Korea Advanced Institute of Science and Technology in 1991, and the Ph.D. from The University of Texas at Austin in 2000, all in electrical engineering. From 1991 to 1996, he was with Korea Telecom. From 1999 to 2002, he was with Lucent Technologies, where he was involved in 3G wireless and multimedia networks. In 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, where he is a Full Professor. His current research interests include image/video quality assessment, computer vision, graphics, cloud computing and multimedia communications, and wireless networks. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2010–2014). He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (2014–) and the *Journal of Electronic Imaging* (2015–) and the Chair of the IEEE P3333.1 Quality Assessment Working Group (2011–). He currently serves as a member in the Technical Committee of the IEEE Multimedia Signal Processing (2016–) and the IEEE IVMSP Technical Committee (2014–), and the APSIPA IVM TC Vice Chair (2016–). Previously, he was the Technical Program Co-Chair at the International Conference on Information Networking 2014, and at the Global 3D Forum 2012 and 2013, and was the General Chair of the 2013 IEEE IVMSP Workshop. He also served as a special issue Guest Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING in 2013, and as an Editor of the *Journal of Communications and Networks* (2009–2015). He received the 2015 Yonsei Academic Award from Yonsei University, the 2012 Special Service Award from the IEEE Broadcast Technology Society, and the 2013 Special Service Award from the IEEE Signal Processing Society.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) is the Curry/Cullen Trust Endowed Chair Professor with The University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering. He is a Faculty Member of the Department of Electrical and Computer Engineering and the Center for Perceptual Systems with the Institute for Neuroscience. He has authored or co-authored over 650 technical articles in his research 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). His current research interests include image and video processing, computational vision, and visual perception. He is a fellow of the Optical Society of America, the Society of Photo-Optical and Instrumentation Engineers, and the American Institute of Medical and Biomedical Engineering. He 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 has been involved in numerous professional society activities, including: the Board of Governors, the IEEE Signal Processing Society, 1996–1998; the Co-Founder and an Editor-in-Chief, the IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996–2002; the Editorial Board, The Proceedings of the IEEE, 1998–2004; the Series Editor of *Image, Video, and Multimedia Processing* (Morgan and Claypool Publishing Company, 2003–present); and the Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, TX, USA, in 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions.