Stereoscopic 3D Visual Discomfort Prediction: A Dynamic Accommodation and Vergence Interaction Model

Heeseok Oh, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

Abstract—The human visual system perceives 3D depth following sensing via its binocular optical system, a series of massively parallel processing units, and a feedback system that controls the mechanical dynamics of eye movements and the crystalline lens. The process of accommodation (focusing of the crystalline lens) and binocular vergence is controlled simultaneously and symbiotically via cross-coupled communication between the two critical depth computation modalities. The output responses of these two subsystems, which are induced by oculomotor control, are used in the computation of a clear and stable cyclopean 3D image from the input stimuli. These subsystems operate in smooth synchronicity when one is viewing the natural world; however, conflicting responses can occur when viewing stereoscopic 3D (S3D) content on fixed displays, causing physiological discomfort. If such occurrences could be predicted, then they might also be avoided (by modifying the acquisition process) or ameliorated (by changing the relative scene depth). Toward this end, we have developed a dynamic accommodation and vergence interaction (DAVI) model that successfully predicts visual discomfort on S3D images. The DAVI model is based on the phasic and reflex responses of the fast fusional vergence mechanism. Quantitative models of accommodation and vergence mismatches are used to conduct visual discomfort prediction. Other 3D perceptual elements are included in the proposed method, including sharpness limits imposed by the depth of focus and fusion limits implied by Panum's fusional area. The DAVI predictor is created by training a support vector machine on features derived from the proposed model and on recorded subjective assessment results. The experimental results are shown to produce accurate predictions of experienced visual discomfort.

Index Terms—Stereoscopic 3D, accommodation vergence mismatch, dynamic accommodation and vergence interaction model, visual discomfort prediction.

H. Oh and S. Lee are with the Department of Electrical and Electronics Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: angdre5@yonsei.ac.kr; slee@yonsei.ac.kr).

A. C. Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: bovik@ece.utexas.edu).

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

I. INTRODUCTION

STEREOSCOPIC 3D (S3D) provides the illusion of depth on two-dimensional flat displays causing viewers to experience a 3D effect by the projection of retinal disparities from the left and right views. However, unnatural 3D input stimuli provided to either or both of two of the main depth-sensing modalities (stereopsis and depth-from-focus) can result in conflicting cross-coupled interactions between the oculomotor and crystalline lens control systems, thus causing experienced visual discomfort and asthenopia. Accordingly, research on the causes of, prediction of, and amelioration of symptoms of physiological discomfort arising from viewing S3D content has recently accelerated in the fields of video engineering, ophthalmology, and vision science.

Several sources of visual discomfort when viewing S3D contents have been studied. For example, Richardt *et al.* [1] showed that asymmetric visual artifacts, such as binocular rivalries arising from left-right inconsistencies, shower door effects in non-photorealistic rendering, and random pixel noise can cause visual discomfort. Kooi and Toet [2] demonstrated that crosstalk and vertical disparity can also produce feelings of visual discomfort. Sohn *et al.* [3] considered object size as a factor that could potentially cause visual discomfort. They found that when viewing an object of narrow width against a large relative depth, severe visual discomfort can occur.

In this paper, we focus on the depth sensing from horizontal disparity, because the most significant cause of discomfort felt when viewing S3D are conflicts arising from accommodation and vergence mismatches (AVM) [4]–[7]. Several studies have been devoted to analyzing AVM occurrences in the context of interactions between the vergence and accommodation subsystems [8], [9]. Emoto *et al.* [10] and Okada *et al.* [11] showed that stereoscopic stimuli displayed on flat-panel displays can create inconsistent vergence and accommodation responses, which can give rise to AVM and sensations of visual discomfort. These early studies examined the effects on experienced visual discomfort of changes in vergence relative to changes in accommodation in the absence of retinal disparity, and conversely, changes in accommodation relative to changes in vergence in the absence of retinal blur. Other studies have been directed towards developing accommodation and vergence models based on the cross-coupled relationship between the processes [12]–[15]. However, work on the application-relevant problem of predicting the degree of

visual discomfort that is experienced when viewing S3D, using models of the cross-coupling and feedback mechanisms that drive the dynamics of accommodation and vergence remain scarce. Indeed, prior attempts to model and predict S3D-induced visual discomfort have focused on analyzing excessive and relative depth presentations delivered by S3D content. This is commonly done by analyzing estimated (computed) disparity maps, without considering dynamic interactions between the accommodation and vergence processes or the various 3D visual subsystems involved in implementing them [16]–[20]. For example, Kim and Sohn [19] developed a visual discomfort prediction algorithm that uses a computed depth map, expressed in terms of disparity after calculating relative and absolute depths. Choi *et al.* [18] constructed a metric on horizontal disparity based on a computed depth map, and used it to predict visual discomfort, but without utilizing any models of human visual function. Qi *et al.* [20] developed an objective model without feature regression by estimating the joint entropy and mutual information between two-view images. Previously, we developed a visual discomfort predictor using a model of local image bandwidth derived from principles of physiological optics [21]. The predictor developed there focuses on anomalies of blur when viewing S3D. These are caused by conflicts occurring in a vergence-driven accommodation cross-link process, yet the relevant control subsystems in the midbrain were not functionally modeled. In a complementary direction, we also constructed a neuronal model based on energy consumption when viewing S3D in primary and extrastriate visual cortex [22]. The discomfort prediction model developed there relies on activity measurements expressed as statistics of computed depth and on established functional models of disparity tuning in neurons in visual area MT.

The approach to the problem that we take here is complementary to our previous models [21], [22]; we model the physiological processes of accommodation and vergence using four descriptive functional components, which are embedded in a basic Schor system model (as explained in detail in Sections II and III). These functions quantify scenarios under which AVMs can occur [23]. Specifically, the four components simultaneously describe limits imposed by the depth of focus (DOF) and of stereopsis operating under the limits of Panum's fusional area as quantitative functions of the stimulus: accommodation-to-accommodation (F_{a2a}), vergence-to-accommodation (F_{v2a}), vergence-to-vergence (F_{v2v}) and accommodation-to-vergence (F_{a2v}). The physiological mechanisms that underlie the definitions of those four factors may be summarized as follows:

- **Blur-driven accommodation** F_{a2a}: This function models control of the crystalline lens and ciliary body, i.e., the accommodation response to bring an object to within the DOF in order to achieve a sharp image of it.
- **Fusional vergence** F_{v2v}: This function models the extraocular motor control that moves the two eyes in opposite directions, such that a binocular image pair can be fused into a single cyclopean image without diplopia. This causes the vergence response to fall within Panum's fusional area, thereby allowing binocular fusion to occur.

- **Vergence-driven accommodation** F_{v2a}: This function models accommodation that is driven by the vergence stimulus. Binocular disparity drives vergence movements, but also stimulates the accommodation response.
- **Accommodation-driven vergence** F_{a2v}: This function models the vergence dynamic that is driven by the accommodation stimulus. While the input stimulus of focal distance is used to change the optical power of the crystalline lens, it also stimulates the vergence response.

The two functions F_{a2a} and F_{v2v} may be modeled as mutually independent processes. However, accommodation and vergence are not separate processes, and there is a strong cross-coupling that, in our model, is expressed by the functions F_{v2a} and F_{a2v}. These collaborative processes also affect each other. Clear and comfortable vision is attained by interactions between the processes represented by F_{v2a} and F_{a2v}, which enable depth-consistent retinal focusing and binocular fusion. As we will show, this functional model makes it possible to predict visual discomfort levels that are experienced when viewing S3D content, by expressing the output response as a function of the input depth stimulus. Modeling the dynamic feedback interactions between accommodation and vergence improves our ability to create algorithms that can accurately predict how human viewers will react (physiologically) to S3D images.

We also incorporate factors that account for the DOF and Panum's fusional area, which place limits on accommodation and vergence, respectively. We incorporate these into a model that describes the interactions that occur between the accommodation and vergence processes. The limits on focal distance, wherein only those objects that fall within a specific depth range are perceived with sharp vision [24], [25]. Panum's fusional area is defined with respect to the horopter, as a depth range within which S3D images can be fused into single, clear 'cyclopean' images without diplopia [26].

Using these concepts, we develop a binocular system model that accepts S3D depth information as input. The model describes cooperating processes of accommodation and vergence, and thus we call it the Dynamic Accommodation and Vergence Interaction (DAVI) model (and algorithm). DAVI models the feedback and cross-linked processes of accommodation and vergence, as well as limits on each of these processes.

Under the DAVI model, given an input depth stimulus (computed S3D depth), the measured S3D responses are used to construct a set of response maps. These response maps are analyzed and used to extract features that are predictive of visual discomfort experienced when viewing S3D images. Suitable perceptually relevant limits on these features are imposed using a DOF based model of out-of-focus blur and a model of Panum's fusional area. The features extracted from the response maps, along with subjective S3D image discomfort scores from a large public S3D database dedicated for this purpose are used to train a support vector machine (SVM). The resulting trained SVM achieves accurate visual discomfort prediction on the real S3D image content in this database.
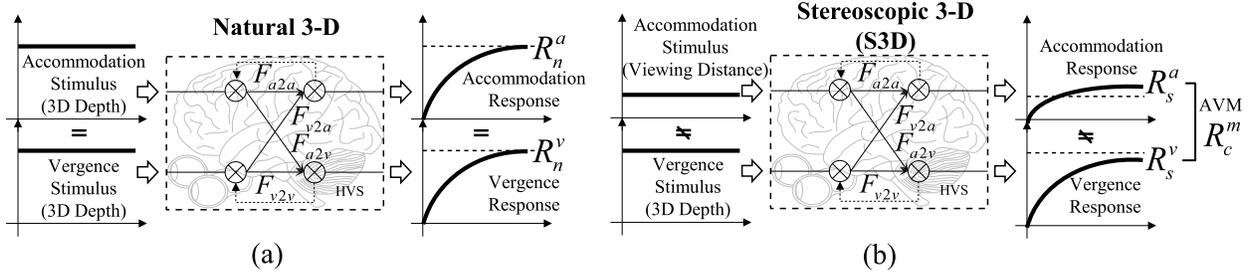
Fig. 1. Overview concept of the DAVI discomfort prediction model: (a) Output responses to natural 3D depth stimuli. (b) Output responses to S3D depth stimuli. When viewing S3D on a flat display the accommodation stimulus is fixed on the screen, while vergence adapts to the projected left-right disparities, thereby creating a potential mismatch between the depth signals communicated by the two modalities.

II. MOTIVATION AND OVERVIEW OF DAVI MODEL

To compute pixel disparities, we used the algorithm of Tanimoto *et al.* [60]. The depth at each pixel of an S3D image is obtained by triangulation of the computed pixel disparities [18], [57]. The resulting depth map is then used as the input stimulus to DAVI. This input stimulus is processed to produce output accommodation and vergence response maps that are computed using the DAVI transfer function model.

A. Model Output Depth Responses

Refer to Fig. 1(a), which depicts the relationship between the accommodation and vergence processes when a real-world 3D environment is being viewed. In the figure, the responses that are used in the DAVI system model are as defined below:

- R_n^a: accommodation output response that would be expected when viewing a naturalistic 3D input depth stimulus
- R_n^v: expected vergence output response to a naturalistic 3D input depth stimulus
- R_s^a: accommodation output response to a displayed S3D input depth stimulus
- R_s^v: vergence output response to a displayed S3D input depth stimulus
- R_c^m: conflict signal between the accommodation and vergence responses to a displayed S3D (AVM) stimulus

The differences between the input depth stimuli and the output responses R_n^a and R_n^v induce feedback processes. The output responses are also affected by the cross-linked processes F_{v2a} and F_{a2v} that respond to the input stimuli. When a normally sighted human views a natural 3D scene, the processes of accommodation and vergence occur in a cross-coupled manner under simultaneous control, as depicted in Fig. 2(a). When viewing such a natural 3D viewing environment, the input depth stimulus is synergistically carried by both output responses R_n^v and R_n^a via feedback from the vergence and accommodation processes.

However, when a human views an S3D picture on a flat display, accommodation remains at the display screen depth even as vergence responds to virtual depths implied by the given compulsory disparities. In other words, the input depth stimuli for accommodation and vergence are different, as depicted in Fig. 1(b), thereby inducing conflicting output responses. The difference between the accommodation and vergence responses, $R_c^m = \left| R_s^a - R_s^v \right|$ (refer to Fig. 1) causes accommodation-vergence mismatches when viewing
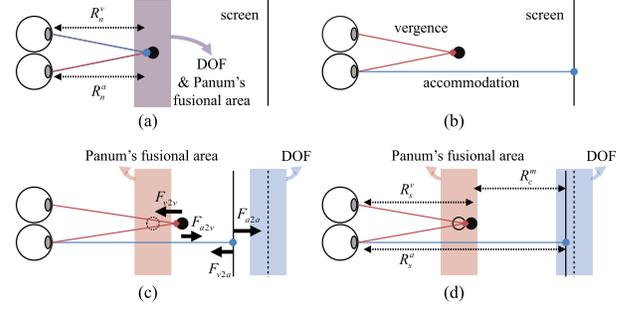

Fig. 2. Illustration of accommodation / vergence mismatch occurrences. (a) The output responses to a natural 3D input depth stimulus are the same. (b) Input depth stimulus when viewing S3D. Unlike vergence, the input accommodation depth stimulus is fixed at a constant viewing distance. (c) Accommodation and vergence control processes collaborate to achieve a steady-state. (d) Model output responses to an S3D picture.

S3D pictures. As shown in Fig. 2(b), when viewing S3D, the accommodation input stimulus is the distance between the viewer and the (assumed planar) screen, so the eyes accommodate to place the image of the fixation point, which will always lie on the screen, into focus. Simultaneously, the binocular disparities induce extraocular movements towards fusing the object within the depth fields.

Fig. 2(c) depicts the dynamic interactions that occur between the accommodation and vergence processes when viewing S3D. The dashed circles represent the vergence input depth stimuli, while the dashed lines indicate the accommodation input depth stimuli. Given the two different stimuli, the accommodation-driven vergence and vergence-driven accommodation functions (F_{a2v} and F_{v2a}) force the vergence response towards the screen (accommodation input stimulus) and the accommodation response towards the object (vergence input stimulus), respectively. Conversely, the fusional vergence function F_{v2v} forces the vergence response towards Panum's fusional area, while the blur-driven accommodation function F_{a2a} forces the accommodation response to lie within the DOF. Conflicts between these four functions can cause disagreements between the accommodation and vergence responses $\left(R_s^a \not\approx R_s^v \right)$.

As a result, although perception of the stereoscopic image may be achieved as a sharp cyclopean image (i.e., the accommodation response lies within the DOF and the vergence response lies within Panum's fusional area), differences between the two different depth stimuli may still cause
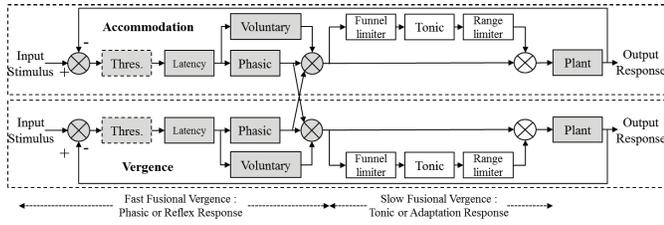
Fig. 3. Schor's dual interaction model: gray-colored components are used to construct the DAVI model of accommodation and vergence.

Fig. 4. Transfer function representation of the dynamic accommodation/vergence model.

a non-zero AVM response R_c^m, as depicted in Fig. 2(d). Such a conflict may be maintained in a stabilized state, but with a high metabolic cost, causing muscular fatigue, and ultimately, visual discomfort. To be able to better predict the degree of experienced visual discomfort, it is therefore desirable to quantitatively model the dynamic feedback interactions occurring during accommodation-vergence conflicts.

B. Dynamic Accommodation and Vergence Interaction Model

The DAVI model is based on Schor's dual interaction model of accommodation and vergence [23], which is depicted in Fig. 3. As indicated at the lower part of the figure, the model consists of two subprocesses, fast fusional vergence and slow fusional vergence [27]–[29]. When an input depth stimulus falls outside of natural perceptual limits, i.e., outside the DOF (for accommodation) or outside Panum's fusional area (for vergence), the fast fusional vergence subsystem is aroused to cross-stabilize the accommodation and vergence responses via feedback.

The fast fusional vergence is associated with an immediate response driven by phasic control. The phasic controller in Fig. 3, which is part of the central nervous system [31], plays the role of accumulating the input depth stimulus to induce an output response. It is strongly affected by the cross-coupling between accommodation and vergence [27]. Conversely, accommodation and vergence are partially a result of voluntary and conscious processes of directed attention that operate largely independent of the cross-link, and are instead controlled by the intention of the viewer to experience a sharp depth image at a specific point. The plant represents the ocular motor movement system (the rectus muscles controlling eye position, and the ciliary muscles controlling the thickness of the crystalline lens), which is driven by signals issued from the phasic controller, which partially determines the input stimulus via the negative feedback loop. When feedback is provided to change the plant, the controller attempts to reduce the error between the desired and actual output responses [31]. The blocks that represent thresholds implied by the DOF and by Panum's fusional area are nonlinear components of the Schor model [31]. When this back and forth process is repeated over a long interval ($\approx$ 30 sec to several hours [27]), neural fatigue accumulates in the tonic controller, which arouses the slow fusional vergence to achieve adaptation at a given depth.

Unlike the fast fusional vergence, slow fusional vergence is only associated with response adaptation. Importantly, this subprocess is no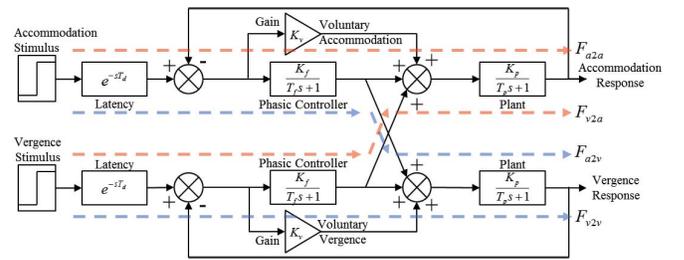t cross-coupled with accommodation and vergence [23], [37]. Therefore, for a given input S3D depth stimulus, interactions between the accommodation and vergence processes mainly arise from the fast fusional vergence responses. Thus DAVI models only factors related to the adjustment of accommodation and vergence for fast fusional vergence, as indicated by the gray boxes in Fig. 3. We will utilize these subsystem models to develop functional features that are predictive of AVM on viewed S3D pictures.

III. ACCOMMODATION AND VERGENCE RESPONSE

Next, we define each component of the DAVI model using a Laplace-domain transfer function representation as summarized in Fig. 3. Four subsystem transfer functions are defined that comprise the overall DAVI model: blur-driven accommodation, vergence-driven accommodation, accommodation-driven vergence, and fusional vergence.

A. Transfer Function of the DAVI Model

Fig. 4 shows a simplified version of the adapted Schor dual interaction model, obtained by extracting the main elements of Fig. 3 related to fast fusional vergence and the plant component. Each element of DAVI is modeled by a sub-transfer function with parameters obtained from clinical experiments on visual signaling. In this model, depth information is first computed in the fast fusional disparity processing center, viz., the phasic controller [27]. The plant transmits signals to the oculomotor system in response to integrated signals issued from the phasic controller. Therefore, the phasic controller and plant are modeled as an integrator as shown in Fig. 4. We use a simple single-pole model of response saturation without divergence and with negative feedback [27], [30]:

$$F(s) = \frac{K}{Ts + 1}. \tag{1}$$

Thus the phasic controller and plant are represented by a first-order system (1), where K is the gain and T is the time delay. As discussed in Section II, the voluntary response is modeled as independently induced and is expressed using the gain without the time delay. Each component in Fig. 4 is modeled using data drawn from subjective tests, including the first-order transfer function in (1) [27], [30]. Parameters used in the model were obtained by regression on data taken from the clinical subjective experiments [23], [27], [30].

Model Parameters	Accommodation	Vergence
• Gains:		
- K_f (Phasic component)	62.5	62.5
- K_v (Voluntary component)	1.5	1.5
- K_p (Plant component)	1.2	1.2
• Time Constants:		
- T_f (Decay time of phasic component)	5 sec	5 sec
- T_p (Decay time of plant component)	0.15 sec	0.25 sec
- T_d (Time delay constant)	0.15 sec	0.30 sec

The parameters of the transfer functions used to model each system component are tabulated in Table I. The time delay parameters T_f, T_p and T_d are taken from the prior studies [23], [31]. However, the gain parameters K_f, K_v, and K_p (substituting K in (1)) depend on the viewer's individuality and environment. Moreover, given that the DAVI model is responsive to depth stimuli when viewing S3D, while the tonic adaptation subsystem is excluded from the basic Schor system model, the normative system parameters require modification. Thus the gain parameter of each component is separately adjusted in order to stabilize the overall system response [32], [33].

In our model, the depth stimulus at any spatial image coordinate is assumed to be constant over time from the moment it is presented to the viewer, hence we take the temporal input depth stimulus to be a simple step function. The depth step response is used to analyze the oculomotor system response to a static S3D stimulus [14], [34]. When the input image is a naturalistic scene that is not viewed on a screen, the accommodation and vergence stimuli will generally remain in agreement, and a viewer with normal vision will be able to comfortably view the input. Therefore, to determine the proper gain parameters, given an input depth stimulus we force the amplitudes of the input and the output to be the same at steady state. For a natural scene, if the amplitude of the input depth stimulus is assumed to be unity, i.e., a unit step function, then the output response is also set to one. Thereby, the gain parameters are determined in order to satisfy the following constraint:

$$S\left(F_{a2a}(s)\right) + S\left(F_{v2a}(s)\right) = S\left(F_{v2v}(s)\right) + S\left(F_{a2v}(s)\right) = 1 \tag{2}$$

where $S(\cdot)$ is a unit step function input depth stimulus, and $F_{a2a}(s)$, $F_{v2a}(s)$, $F_{v2v}(s)$, and $F_{a2v}(s)$ are blur-driven accommodation, vergence-driven accommodation, fusional vergence, and accommodation-driven vergence transfer functions, respectively. Note that the accommodation and vergence sub-responses can be summed to produce total responses [32].

The four transfer functions contained in the overall DAVI model depicted in Fig. 4 can be formularized using the parameters in Table I. It is difficult to quantify the negative feedback response of the closed-loop system, since it is unrealistic to assume that a vergence (accommodation) sub-system will produce a zero response to an accommodation (vergence) system feedback. To resolve this problem, Jiang *et al.* [35] derived a linear model of binocular control expressed as a closed-loop system. In regards to the

functionality of the model presented here, the model of Jiang [35] is relevant to the definition of F_{v2a}. Maxwell *et al.* [36] modeled the first- and second-order relationships between the inputs and outputs of binocular accommodation and vergence systems. Our approach to resolving this dilemma is to approximate the non-linear interactions between stimulus and response as the transfer function of an open-loop system. This was accomplished using the regression tool in Matlab/Simulink, yielding four 4th order transfer functions:

$$F_{a2v}(s) = \frac{100s^2 + 420s + 80}{s^4 + 30.27s^3 + 381.1s^2 + 2357s + 456.4} \tag{3}$$

$$F_{v2v}(s) = \frac{12s^3 + 239.2s^2 + 1914s + 373.4}{s^4 + 30.27s^3 + 381.1s^2 + 2357s + 456.4} \tag{4}$$

$$F_{v2a}(s) = \frac{60s^2 + 412s + 80}{s^4 + 30.27s^3 + 381.1s^2 + 2357s + 456.4} \tag{5}$$

$$F_{a2a}(s) = \frac{7.2s^3 + 197.3s^2 + 1906s + 373.4}{s^4 + 30.27s^3 + 381.1s^2 + 2357s + 456.4}. \tag{6}$$

Here, the input variable is the depth information, i.e., the computed distance from the viewer to an object. The fast fusional vergence responses in Schor's model are thus expressed as a linear combination of the four transfer functions (3), (4), (5), and (6). Similarly, the sum of the responses F_{a2v} and F_{v2v} is the total vergence output response, while the sum of the responses F_{v2a} and F_{a2a} is the total output accommodation response. As explained in Section VI, we validated (3)-(6) by comparing the responses of the DAVI model to real subjective values obtained from clinical experiments.

B. Responses Under the DAVI Model

1) Observation 1 (Natural 3D Responses): The output response to a step input stimulus stabilizes to a steady state after a short transient period. At steady state, the input and output amplitudes are the same, which means that a human viewer will comfortably process the input.

Fig. 5 depicts the output responses of the four functions when a unit step function with a depth amplitude of 1.0 D is applied to the input stimulus. Here, D (diopters) and MA (meter angles) are the reciprocal of the stimulus distance in meters. Fig. 5(a) represents the fusional vergence response $S\left(F_{v2v}(s)\right)$ to a step vergence input stimulus, the accommodation-driven vergence response $S\left(F_{a2v}(s)\right)$ to a step accommodation input stimulus, and the total vergence output response $S\left(F_{v2v}(s)\right) + S\left(F_{a2v}(s)\right)$ for both stimuli. After a short transient period, the output vergence and accommodation responses become identical to each other at depth amplitudes of 1.0 MA and 1.0 D. Note that the cross-coupled feedback process between the accommodation and vergence subsystems is assumed to occur under a natural 3D viewing condition.

2) Observation 2 (Displayed Stereoscopic 3D Responses): Because of the difference between the input depth stimuli for accommodation and vergence, there exists a conflict between the associated output responses. Hence, the steady-state amplitude responses of accommodation and vergence are not equal to the input depth stimulus.
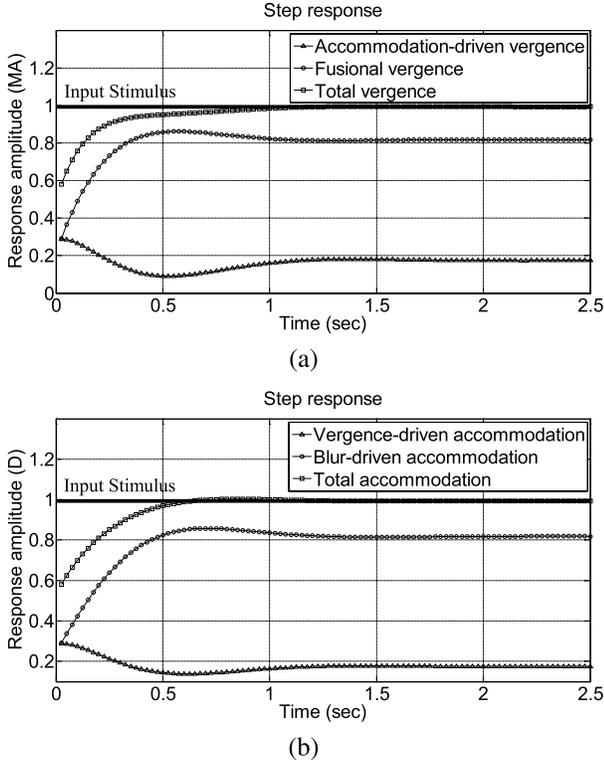
(a)

(b)

Fig. 5. Step response (amplitude of the step input stimulus is unity) of the accommodation and vergence dynamic sub-system models (a) vergence step responses (fusional vergence F_{v2v}, accommodation-driven vergence F_{a2a}, and their sum), (b) accommodation step responses (vergence-driven accommodation F_{v2a}, blur-driven accommodation F_{a2a}, and their sum).

Fig. 5(b) depicts the blur-driven accommodation response $S(F_{a2a}(s))$ to a step accommodation stimulus, the vergence-driven accommodation response $S(F_{v2a}(s))$ to a step vergence stimulus, and the total accommodation response $S(F_{a2a}(s)) + S(F_{v2a}(s))$. In Fig. 6, a temporal unit step function vergence stimulus is also applied, while the accommodation stimulus is held fixed at the viewing distance of 0.58 D (1.7 meter), which is three times the display height, in accordance with the recommended viewing environment [38], [39]. When viewing S3D on a display screen, the eyes adjust the focal length to the screen depth to bring the image displayed on the screen into sharp focus on the retina. Thus, the accommodation depth, or viewing distance, is fixed to be the input stimulus. As mentioned in Section II, the different types of depth stimuli when viewing displayed S3D pictures induce mismatches of accommodation and vergence.

IV. MAP CONSTRUCTION

In this section, we describe how to obtain relevant feature maps that are combined into overall DAVI responses for a given S3D depth stimulus. The DAVI responses are computed at every spatial coordinate, but are weighted using a 3D saliency index. The maps also account for limits implied by the DOF and by Panum's fusional area.

A. DAVI Response Map Construction

The DAVI system input is depth information computed from S3D images, which is then transformed into
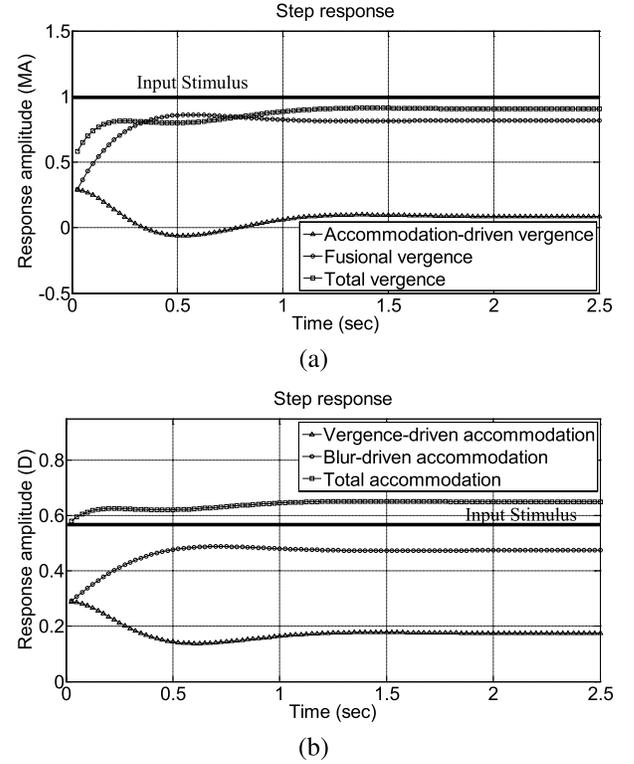

(a)

(b)

Fig. 6. Step response (amplitude of step input stimulus is unity for vergence and viewing distance for accommodation) of the accommodation and vergence dynamic models (a) vergence step responses (fusional vergence F_{v2v}, accommodation-driven vergence F_{a2a}, and their sum), (b) accommodation step responses (vergence-driven accommodation F_{v2a}, blur-driven accommodation F_{a2a}, and their sum).
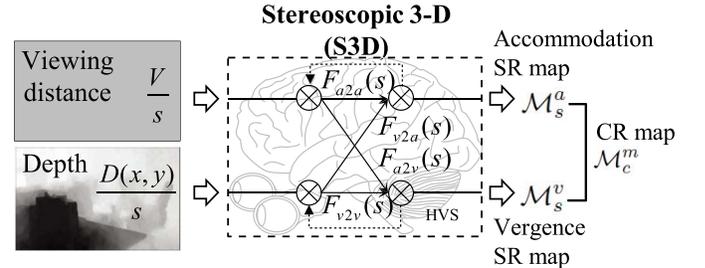

Fig. 7. Response map construction: the SR maps $\mathcal{M}_s^a$ and $\mathcal{M}_s^v$ use the viewing distance and the depth map as the input depth stimulus for accommodation and vergence, respectively. The CR map is obtained from the two SR maps.

a set of 2D feature maps that capture the following information:

- **SR maps**: Of the displayed S3D responses R_s^a and R_s^v
- **CR map**: Of the conflict S3D responses R_c^m (AVM)

1) Stereoscopic Response (SR) Maps: The SR maps represent the output responses induced by processes of accommodation and vergence when viewing the displayed S3D image. Fig. 7 is the flow diagram depicting the construction of the SR maps from an input depth stimulus.

The binocular visual system synchronously controls the accommodation and vergence processes via cross-coupled signaling and feedback [40]. Therefore, when a natural 3D image
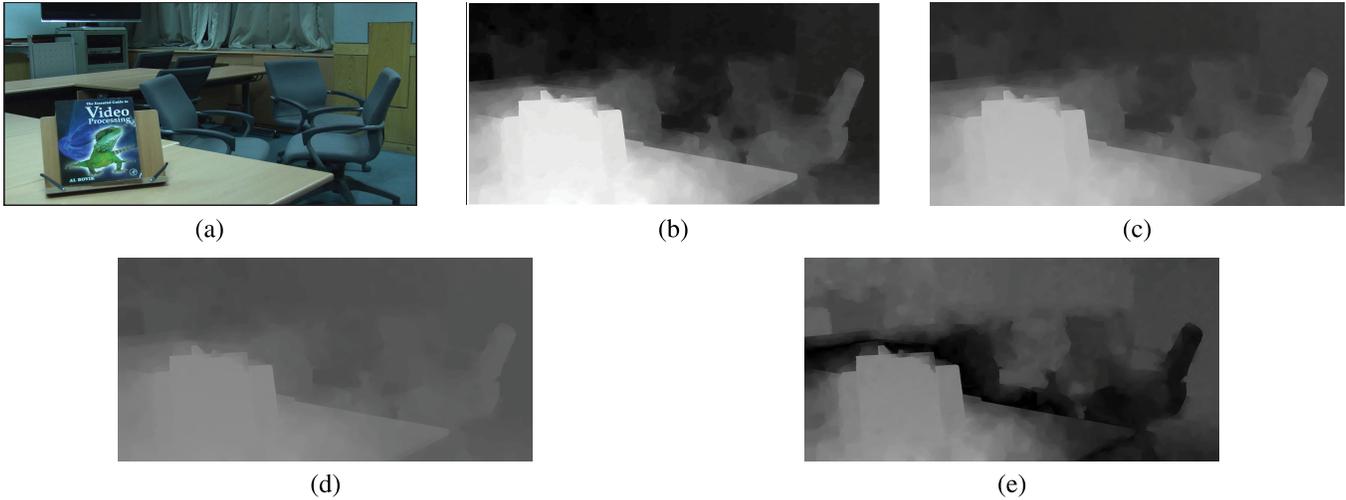
Fig. 8. (a) Original stereoscopic image (left image) from the IEEE-SA database. (b) Computed depth map. (c) Stereoscopic vergence response map $\mathcal{M}_s^v$. (d) Stereoscopic accommodation response map $\mathcal{M}_s^a$. (e) AVM map $\mathcal{M}_c^m$.

is viewed, the accommodation and vergence responses are harmonized with each other, reaching a steady state after a short duration. However, when viewing S3D, the focus is assumed to be fixed on the display screen. Therefore, the input depth stimulus for accommodation is a constant viewing distance V, hence the input accommodation depth stimulus is independent of the depths obtained from stereoscopic disparity. Of course, the input vergence depth stimulus is closely related to the depths implicit in the stereoscopic image, causing the viewer to experience artificial depths.

Since the S3D stimulus is given from the moment of viewing it, it can be assumed that the vergence input depth stimulus at each point is a step function with a constant depth amplitude over time. Hence, the estimated depth map $D_v(x, y)$ can be interpreted as a sequence of 1-D temporal values over 2D space. Thus, the depth map can be expressed as $\frac{D_v(x,y)}{s}$, since the depth value at each pixel is constant once viewing of an S3D image commences. Based on this assumption, we construct two SR maps which consist of S3D responses:

$$\mathcal{M}_s^v(x, y) = \lim_{t \to \infty} \mathcal{L}^{-1}\left\{\frac{V}{s} \cdot F_{a2v}(s) + \frac{D_v(x, y)}{s} \cdot F_{v2v}(s)\right\}(t).$$
(7)

$$\mathcal{M}_s^a(x, y) = \lim_{t \to \infty} \mathcal{L}^{-1}\left\{\frac{V}{s} \cdot F_{a2a}(s) + \frac{D_v(x, y)}{s} \cdot F_{v2a}(s)\right\}(t).$$
(8)

Unlike F_{v2a} and F_{v2v}, the input depth stimulus for $F_{a2v}(s)$ and $F_{a2a}(s)$ is the viewing distance V, which may be expressed as $\frac{V}{s}$ in the Laplace domain.

As a result, the input depth stimulus for accommodation is different from that for vergence, which leads to a difference in the output response obtained from a corresponding natural 3D environment. The SR maps $\mathcal{M}_s^v$ and $\mathcal{M}_s^a$ are computed using (8) and (9) at each pixel (x, y) of the image.

Figs. 8(c) and (d) are the SR maps corresponding to the image in Fig. 8(a). Fig. 8(c) is the SR map $\mathcal{M}_s^v$ for vergence, while Fig. 8(d) is the SR map $\mathcal{M}_s^a$ for accommodation.

Brighter regions are closer to the viewer. Because the input depth stimulus is V, $\mathcal{M}_s^v$ generally has a lower response value than $\mathcal{M}_n^v$ (closer to screen) for negative depth. Additionally, $\mathcal{M}_s^a$ has a lower response value than $\mathcal{M}_n^a$ for negative depth, since vergence-driven accommodation induces negative feedback to $\mathcal{M}_n^a$ owing to the fixed input accommodation depth stimulus V.

2) Conflict Response (CR) Map: Viewing S3D on a flat display screen induces different accommodation and vergence responses relative to viewing a natural 3D environment. It accounts for any disagreement between the two responses, and this quantity is an important discomfort feature. The CR map representing such AVM is defined as the absolute difference between the two SR maps:

$$\mathcal{M}_c^m(x, y) = \left|\mathcal{M}_s^v(x, y) - \mathcal{M}_s^a(x, y)\right|.$$
(9)

The human visual system attempts to force the responses to lie within the DOF and Panum's fusional area, driven by the blur-driven accommodation process and by fusional vergence. However, when viewing S3D, AVM inevitably occurs, which can lead to visual discomfort.

Fig. 8(e) shows an example of the CR map, where brighter regions represent larger conflicts. By contrast, in darker regions, the conflict between the accommodation and vergence responses is less severe or negligible, i.e., the depths are close to the 2D screen with zero disparity.

B. Depth of Focus and Panum's Fusional Area

In the basic Schor model, limits imposed by the DOF and Panum's fusional area are applied prior to the phasic controller. These constraints on the binocular system induce control of the crystalline lens and of the oculomotor system according to the level of input stimuli. These non-linear functions are employed to obtain features relevant to the prediction of visual discomfort, since certain effects are not easily detectable when only using models of the statistics of computed depth. In Schor's model, these thresholds were set as fixed values [23].
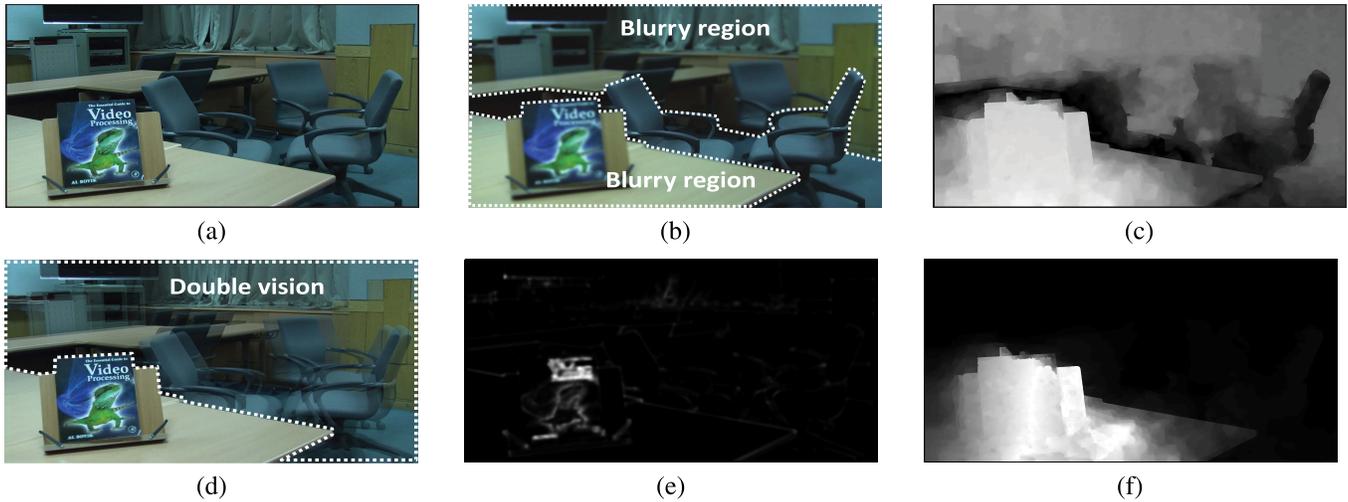
Fig. 9. (a) Original stereoscopic image (left image). (b) Out-of-focus image for an assumed blur. The chair is in the focused region, while the region outside the DOF is perceived as blurry. (c) OF map $\mathcal{M}_d(x, y)$ using (11). Bright regions represent blurry objects. (d) Diplopic image based on Panum's fusional area. The book is at the fixation point, while the region out of Panum's fusional area is perceived with 'double vision'. (e) Detected salient region. (f) PF map $\mathcal{M}_p(x, y)$ using (19). Brighter regions are considered as fused (cyclopean).

However, using fixed values results in hard decisions which may be inaccurate. To overcome such a drawback, we construct 2D feature maps utilizing a continuous model of out-of-focus blur and of Panum's fusional area.

- **OF map**: Of the degree of out-of-focus starting from the focal distance
- **PF map**: Of Panum's fusional area representing how well the 3D object is fused

1) Out-of-Focus (OF) Map: The human visual system attempts to induce a shift of accommodation that agrees with the vergence input depth stimulus as expressed by F_{v2a}. Conversely, the human visual system attempts to perceive a sharp image by situating accommodation within the DOF F_{a2a} in the direction of the screen [41].

The DOF represents a tolerance range around the focal distance that allows a viewer to perceive sharp vision. Additionally, the DOF serves as a threshold on the output accommodation response and can be used to form a "zone of comfort" [42]. When accommodation focuses on an object, the light that passes through the crystalline lens is concentrated at the fovea; thereby the object is sharply perceived, while regions outside the DOF are perceived with blur. Because focus is not perfectly achieved at a given point, a disk called the circle of confusion is defined [24].

Fig. 9(b) depicts a blurry region outside of the DOF when using the depth map shown in Fig. 8(b) with accommodation focused on the screen, unlike the original stereoscopic image in Fig. 9(a) that provides a sharp view throughout. Prior clinical measurements have determined that the zone of comfort occurs over a finite region, e.g., within a range of $\pm0.33D$ based on an assumed fixed viewing distance [5]. However, a continuous model of the circle of confusion presents advantages for analysis. Towards this end, we construct a circle of confusion model on the retina using a thin-lens system based on accommodation, as depicted in Fig. 10. The focal length f is related to the distance from the fixation point and the nodal distance s
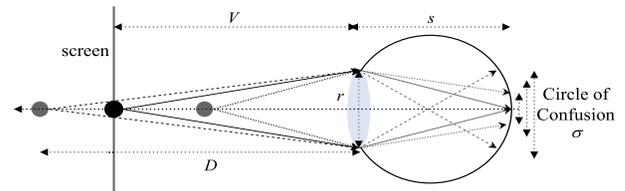

Fig. 10. Circle of confusion geometry in 3D. Generally, the pupil diameter $r \approx 3$ mm and the nodal distance $s \approx 16$ mm [43].

by $\frac{1}{V} + \frac{1}{s} = \frac{1}{f}$. When viewing an S3D picture on a flat screen, the viewing distance V is fixed for accommodation because the fixation point is on the screen. The diameter of the circle of confusion can then be expressed as follows [24], [44]:

$$\sigma = r\left(\frac{s}{V}\right)\left|1 - \frac{V}{D}\right| \tag{10}$$

where D is the depth created by stereopsis when a point on an object has disparity, and where r is the pupil diameter. Since the eye accommodates to the fixation point at viewing distance V, an object on the screen is imaged on the retina in sharp focus. However, when another object has an implied 3D depth away from the screen that has been focused on, i.e., the fixation point, it is perceived as blurry. The diameter of the circle of confusion σ captures occurrences of out-of-focus blur. The OF map $\mathcal{M}_d$ is defined at each pixel (x, y) in terms of the diameter of the circle of confusion σ:

$$\mathcal{M}_d(x, y) = \sigma(x, y). \tag{11}$$

Fig. 9(c) shows the OF map $\mathcal{M}_d$ for the image in Fig. 9(a). Darker regions are coded as closer to the screen. Brighter regions portend an increased likelihood of AVM occurrences.

2) Panum's Fusional (PF) Area Map: Panum's fusional area places a threshold on vergence stimuli that determines whether an object can be fused with single vision. It is defined with respect to the horopter, which is a hypothetical
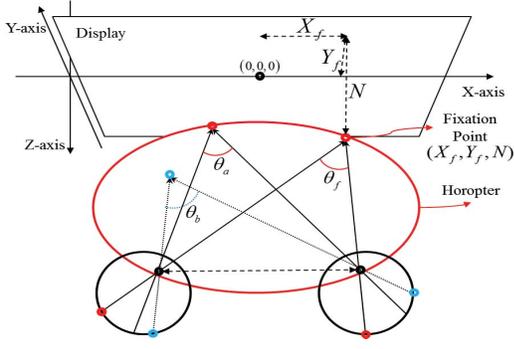
Fig. 11. Definition of horopter and angular disparity in a stereoscopic geometry. The dashed circle is the horopter. The angles θ_f and θ_a on the horopter are the same. Such objects are perceived with single vision. By contrast, the angle θ_b on the horopter causes diplopia.

circle that connects three points (the fixation point on an object and the two nodal points of the eyes) [45]. The red circle in Fig. 11 depicts the Vieth-Muller circle, which is the theoretical horopter. A visual angle on the horopter is equal to θ_f at the fixation point, and a point on an object on this circle is perceived with single vision. Then, the angular disparity is $\Delta\theta = |\theta_f - \theta|$, where θ is the projection angle between the two eyes of an object not lying at the fixation point. An object outside of this circle, such as an object having angle θ_b, will not be fused, and would be perceived with double vision (diplopia), similar to Fig. 9(d).

When a stereoscopic image provides excessive depth, visual discomfort can be caused by diplopia. To establish Panum's fusional area given a fixation point, salient regions should be preferentially processed. Towards this goal, we deploy the 3D saliency detection technique described in Kim *et al.* [57]. The predicted visual fixation points are used to define a reference angular disparity, or horopter [58], [59]. Fig. 9(e) shows salient regions computed from the image in Fig. 9(a).

Diplopia arises continuously with distance from the horopter. A sudden change does not occur from a sense of single vision to one of double vision as vergence goes beyond the proximal distance and the distal distance limitation for angular disparity. Therefore, we employ a continuous model of Panum's fusional area based on the angular disparity $\Delta\theta$ to construct the PF map [46]:

$$\mathcal{M}_p(x, y) = \begin{cases} 1, & (0 \le \Delta\theta \le b) \\ \exp\left(-\dfrac{\Delta\theta(x, y) - b}{c}\right), & (b < \Delta\theta) \end{cases} \quad (12)$$

where b represents the threshold below which the viewer retains single vision, and c is a constant that affects the decay of the weight decrement (here $b = 0$ and $c = 0.62$ as was used [46], [58]). Fig. 9(f) shows the PF map constructed using the image in Fig. 9(a). The viewer is predicted to be able to fuse the stereoscopic image more easily in brighter regions than in darker areas. The value of the PF map peaks at zero disparity where the viewer feels most comfortable.

V. FEATURE EXTRACTION

The DAVI model utilizes features that are predictive of experienced visual discomfort when viewing S3D pictures. These are derived from the histograms of the three types of constructed maps: OF, PF and CR, for each pair of stereoscopic images.

Several authors have studied the effects that the depth category has on the level of visual discomfort felt when viewing S3D [16]–[19]. Generally, depths associated with positive disparities, i.e., behind the display screen (which is assumed to have zero disparity) are associated with higher levels of comfort than depths associated with negative disparities (i.e., in front of the screen). Therefore, separate positive and negative disparity associated mean depths are calculated from each map's histogram:

$$\bar{m}^+(\mathcal{M}(x, y)) = \frac{1}{\bar{n}^+} \sum_{n>0} \mathcal{M}^h(n) \quad (13)$$

$$\bar{m}^-(\mathcal{M}(x, y)) = \frac{1}{\bar{n}^-} \sum_{n<0} \mathcal{M}^h(n) \quad (14)$$

where $\bar{m}^+$ ($\bar{m}^-$) is the mean depth associated with positive (negative) disparities, $\bar{n}^+$ ($\bar{n}^-$) is the total number of pixels associated with positive (negative) disparities, $\mathcal{M}(x, y)$ is the map, and $\mathcal{M}^h$ is the histogram of $\mathcal{M}$.

Relative depth is defined as the ratio between positive and negative depths. When the relative depth ratio is large, feelings of visual discomfort are more likely to occur [4], [6], [47]. By using (13) and (14), the relative depth can be calculated as:

$$\bar{\gamma}(\mathcal{M}(x, y)) = \frac{\bar{m}^+(\mathcal{M}(x, y))}{\bar{m}^-(\mathcal{M}(x, y))} \quad (15)$$

We also deploy a measure of depth dispersion, obtained as

$$\bar{\sigma}(\mathcal{M}(x, y)) = \frac{1}{\max(\mathcal{M}(x, y))} \times \sqrt{\frac{1}{\bar{n}} \sum_n \left(\mathcal{M}^h(n) - \bar{m}(\mathcal{M}(x, y))\right)^2}, \quad (16)$$

where $\bar{n}$ and $\bar{m}(\mathcal{M}(x, y))$ are the total number of pixels and the mean value of $\mathcal{M}(x, y)$, respectively.

Finally, it is known that the most severe impairments of an image or video can have a dominant effect on the overall perceived quality, an observation that is often used to improve objective 2D picture and video quality analyzers [48], [54]. Similarly, we have observed that the likelihood that visual discomfort will occur is increased by the presence of regions containing unusual depth characteristics. Towards quantifying and capturing these spatially localized occurrences, the upper p^{th} percentiles (averages) of the histograms of each map are calculated for both positive and negative depths. We define the upper (mean) p^{th} percentile and the lower $(1 - p)^{th}$ percentile

$$\bar{m}_p^+(\mathcal{M}(x, y)) = \frac{1}{\bar{n}_p^+} \sum_{n>n_p^+} \mathcal{M}^h(n) \quad (17)$$

$$\bar{m}_p^-(\mathcal{M}(x, y)) = \frac{1}{\bar{n}_p^-} \sum_{n<n_p^-} \mathcal{M}^h(n) \quad (18)$$
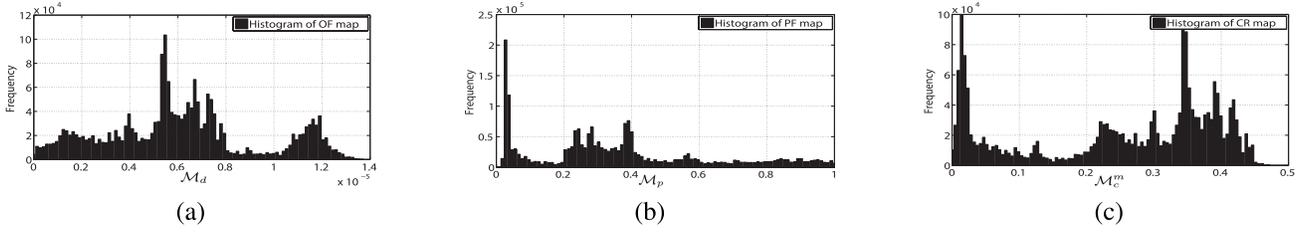
Fig. 12.   Histogram of computed depth feature maps from the original stereoscopic image in Fig. 9 : (a) out-of-focus map $\mathcal{M}_d$, (b) Panum's fusional area map $\mathcal{M}_p$ and (c) conflict AVM response map $\mathcal{M}_c^m$.

where superscripts '+' and '−' indicate that the percentiles are computed over positive (negative) depths. Likewise, $\bar{n}_p^+$ ($\bar{n}_p^-$) is the number of p-percentile pixels at positive (negative) depths. In short, $\bar{n}_p^+ = \bar{n}^+ \cdot p/100$ and $\bar{n}_p^- = \bar{n}^- \cdot p/100$, respectively. In our implementation, we used $p = 5$ [21], [22]. Fig. 12 shows example histograms of each type of map computed from the original stereoscopic image in Fig. 9.

Using (13)-(18), the above-defined four features are extracted from each map and compared with the learned model using a classifier (SVM) along with the mean opinion scores (MOS) obtained from a subjective assessment study. The extracted features can be categorized as descriptive of the implied DOF, of Panum's fusional area (PF) and of DAVI features. The 12 features are summarized below.

- OF features (3 features)
  - ✓ $\bar{m}_p^+(\mathcal{M}_d)$, $\bar{m}_p^-(\mathcal{M}_d)$ and $\bar{\sigma}(\mathcal{M}_d)$: Features extracted from the OF map $\mathcal{M}_d$ capture the degree of perceived sharpness in the stereoscopic image.
- PF features (3 features)
  - ✓ $\bar{m}_p^+(\mathcal{M}_p)$, $\bar{m}_p^-(\mathcal{M}_p)$ and $\bar{\sigma}(\mathcal{M}_p)$: Features extracted from the PF map $\mathcal{M}_p$ are predictive of how easily the stereoscopic image pair can be fused.
- DAVI features (6 features)
  - ✓ $\bar{m}^+(\mathcal{M}_c^m)$, $\bar{m}^-(\mathcal{M}_c^m)$, $\bar{m}_p^+(\mathcal{M}_c^m)$ and $\bar{m}_p^-(\mathcal{M}_c^m)$: Features extracted from the AVM CR map $\mathcal{M}_c^m$ correlate highly with the degree of AVM caused by S3D perception.
  - ✓ $\bar{\gamma}(\mathcal{M}_s^v)$ and $\bar{\gamma}(\mathcal{M}_s^a)$: Features extracted from the accommodation and vergence SR maps $\mathcal{M}_s^a$ and $\mathcal{M}_s^v$ capture the perceived relative depth when viewing S3D.

VI. Performance of the DAVI Model

A. Verification of the DAVI Model

We compared the predictions of the DAVI model to the numerical results of previous clinical experiments on the oculomotor responses of human viewers of S3D content. Fukushima et al. measured the responses of 8 subjects when unbalanced accommodation and vergence stimuli were provided [49]. To analyze oculomotor synergies, they examined the individual dynamic responses to the unbalanced stimuli. They also defined a response differential expressed as the difference between the initial response and the mean of the static steady-state response (response differential = initial response - static response). The initial response is the first
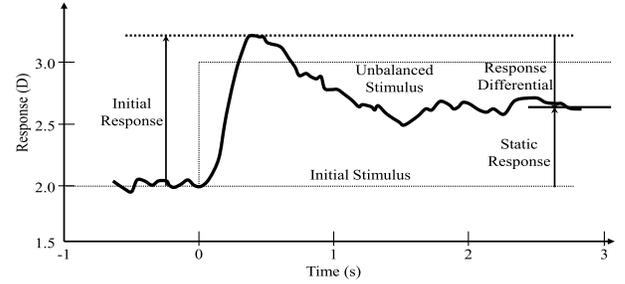

Fig. 13.   Definition of response differential for an unbalanced stimulus [49]. The dashed and bold-dashed lines represent the input stimuli and the maximum response, respectively. The initial response is the maximum local response value after the unbalanced stimulus begins. The static response appears at steady state. The response differential represents the difference between the two responses.

local maximum response value. This indicates the amount of overshoot immediately after the input depth is presented. The static response was defined as the average response from 2.5 to 3 seconds (at steady-state) after the onset of the unbalanced stimulus.

Fig. 13 depicts an example response to a 1.0 D step unbalanced stimuli. When a vergence depth stimulus that is in conflict with the accommodation depth is input, the maximum response value appears after a short period, following which the 3D vision system attempts to stabilize the response via feedback. It is known that the static response at steady state is generally significantly different from the initial input depth stimulus. The response differential is the difference between these responses. In [11] and [49], a practical input depth stimuli was used to study accommodation and vergence responses was 2.0 D (0.5 meter) and 3.0 MA (0.33 meter). Fig. 14 shows the simulated accommodation and vergence output responses under the DAVI model when 2.0 D and 3.0 MA are used as the levels of as the accommodation and vergence input step profiles $\left(\frac{1}{2}S(F_{a2a}) + \frac{1}{3}S(F_{v2a})\right)$ and $\left(\frac{1}{2}S(F_{a2v}) + \frac{1}{3}S(F_{v2v})\right)$, in agreement with the parameters used in the clinical experiments in [49]. The accommodation (vergence) output response of the clinical experiment matches the total response of both the blur-driven accommodation function F_{a2a} and the vergence-driven accommodation function F_{v2a} (fusional vergence function F_{v2v} and accommodation-driven vergence function F_{a2v}) defined earlier.

Fig. 14 shows that the response behavior of the DAVI model is similar to that of the clinical results in Fukushima's research, when the input depth stimuli for accommodation and vergence

TABLE II
RESPONSES MEASURED FROM THE DAVI MODEL AND FROM CLINICAL EXPERIMENTS [49]

|	Accommodation static response (D)	Accommodation response differential (D)	Vergence static response (MA)	Vergence response differential (MA)
Subject1	0.55	0.53	1.06	0.22
Subject2	0.19	0.36	0.79	0.06
Subject3	0.39	0.31	0.94	-0.04
Subject4	0.46	0.18	1.10	0.20
Subject5	0.73	-0.19	0.81	0.28
Subject6	0.87	-0.23	1.12	0.27
Subject7	0.79	-0.05	1.06	0.18
Subject8	0.21	0.01	1.10	0.20
Mean of subjects	0.52 ± 0.26	0.11 ± 0.27	1.00 ± 0.13	0.17 ± 0.11
DAVI model	**0.17**	**0.14**	**0.81**	**0.12**
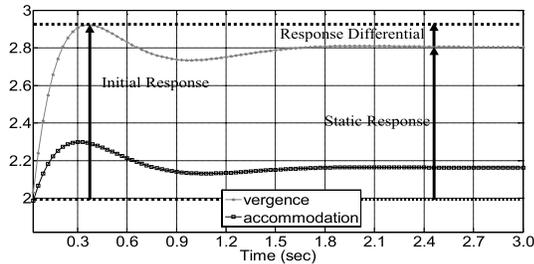
Fig. 14. Matching the responses of the DAVI model with the responses defined in Fukushima's research [49]. The dashed lines represent maximum response and the response at initial state.

are 2.0 D and 3.0 MA (unbalanced stimuli), respectively [49]. In the DAVI simulation, the amplitudes of the accommodation and vergence static responses were 0.17 D and 0.81 MA. The mean accommodation and vergence responses of the subjects were 0.52 ± 0.26 D and 1.00 ± 0.13 MA [49]. In addition, the accommodation and vergence response differential values were 0.14 D and 0.12 MA in the DAVI simulation, while the mean responses in the clinical experiments were 0.11 ± 0.05 D and 0.17 ± 0.11 MA. The individual responses of the subjects and the results of the DAVI simulation are summarized in Table II.

As a way of validating the DAVI model, we compared the static response and the response differential data obtained in the clinical experiments on 8 subjects as reported by Fukushima *et al.* [49], with the results of the DAVI simulation. Overall, the temporal response tendency, response differential and static responses delivered by the DAVI model track the subjective responses measured in the clinical experiment. The numerical model follows a similar trend to that of the subjective data, which reflects the dynamics occurring in the accommodation and vergence mechanism.

B. Visual Discomfort Prediction Results

Just as it is standard procedure to test the performance of image and video quality prediction models against human opinions [50], so also 3D discomfort prediction performance needs to be gauged against the gold standard reference of human subjective judgments of experienced visual discomfort when viewing S3D images. We conducted a subjective study on the degree of visual discomfort experienced by human viewers of S3D pictures using stereo image pairs from the IEEE-SA stereo image database, which consists of 800 S3D

image pairs that were obtained using a PANASONIC AG-3DA1 twin-lens camera [22], [51]. The size of each image is 1920×1080 pixels. The subjective discomfort assessment experiment was conducted in a laboratory environment as prescribed by the current recommendation for conducting subjective studies [38]. Twenty-eight subjects participated in the subjective evaluation whose ages ranged from 22 to 38 years. Each participant was naive in regards to concepts in the 3D image processing field [22]. The number of subjects was about twice the standardized recommendation [38]. Since the IEEE-SA stereo image database contains a large number of S3D image pairs, we divided the subjective test into nine sessions, one training session and eight testing sessions. In the training session, each participant was instructed regarding the methodology of the test and of the general range of discomfort levels by showing them 20 S3D pairs exhibiting the broad range of disparities in the database. In each testing session, 100 randomly shuffled S3D image pairs were evaluated [22]. Each S3D image pair was displayed for 10 seconds. A rest period of 10 minutes was inserted between each pair of consecutive sessions to avoid the accumulation of visual fatigue. Each participant was asked to assign a visual comfort score without supplying any additional cues. The allowable scores ranged from 1 = extremely uncomfortable to 5 = very comfortable. The subjective scores were collected using a tablet PC [56], thereby enhancing the reliability of the subjective test results. Four outliers were discarded in accordance with the recommendation and the collected mean opinion score (MOS) was then normalized [22], [38]. A 46 inch polarized 3D display was used and the viewing distance was set to about 1.7 meters which is three times the display height [38], [39]. The ratio of the luminance of the inactive screen to its peak luminance was less than 0.02. The ratio of the luminance of the screen (when displaying only the black level in a completely dark room) to peak white was about 0.01. The ratio of the luminance of the background behind the picture monitor to the peak luminance of the picture was about 0.15. The luminance levels were measured using a KONICA MINOLTA CA-310. Other sources of room illumination were negligible.

Performance evaluation was done using the linear correlation coefficient (LCC), the Spearman rank order correlation coefficient (SROCC), and the root mean square error (RMSE) relative to the MOS recorded during the subjective assessment task. SVM was used to create a regression model that captures the relationship between the

TABLE III

LCC FOR 2000 TRIALS OF RANDOMLY CHOSEN TRAIN AND TEST SETS

	Mean	Median	Standard Deviation
Yano [16]	0.4030	0.4081	0.0724
Nojiri [17]	0.6938	0.7044	0.0857
Choi [18]	0.6732	0.6808	0.0836
Kim [19]	0.7040	0.7150	0.0755
3DAVM [21]	0.8524	0.8616	0.0454
3DVDP [22]	0.8505	0.8595	0.0457
OF features	**0.8135**	**0.8218**	**0.0484**
PF features	**0.7340**	**0.7369**	**0.0537**
DAVI features	**0.8291**	**0.8353**	**0.0496**
OF+PF+DAVI	**0.8590**	**0.8629**	**0.0452**

TABLE IV

SROCC FOR 2000 TRIALS OF RANDOMLY CHOSEN TRAIN AND TEST SETS

	Mean	Median	Standard Deviation
Yano [16]	0.3361	0.3374	0.0723
Nojiri [17]	0.6063	0.6099	0.0752
Choi [18]	0.5866	0.5957	0.0789
Kim [19]	0.6172	0.6216	0.0703
3DAVM [21]	0.7785	0.7801	0.0406
3DVDP [22]	0.7784	0.7790	0.0417
OF features	**0.7170**	**0.7183**	**0.0465**
PF features	**0.6555**	**0.6568**	**0.0548**
DAVI features	**0.7632**	**0.7669**	**0.0451**
OF+PF+DAVI	**0.7887**	**0.7900**	**0.0405**

TABLE V

RMSE FOR 2000 TRIALS OF RANDOMLY CHOSEN TRAIN AND TEST SETS

	Mean	Median	Standard Deviation
Yano [16]	0.76	0.76	0.06
Nojiri [17]	0.59	0.59	0.07
Choi [18]	0.61	0.61	0.06
Kim [19]	0.58	0.58	0.06
3DAVM [21]	0.42	0.40	0.06
3DVDP [22]	0.42	0.41	0.06
OF features	**0.47**	**0.46**	**0.06**
PF features	**0.56**	**0.56**	**0.06**
DAVI features	**0.44**	**0.43**	**0.06**
OF+PF+DAVI	**0.41**	**0.40**	**0.06**

discomfort-predictive feature vectors and the MOS values. The SVM was implemented using the linear kernel of the libSVM package [52], and the feature vectors were learned by the cross-validation method. The training set (for each train-test iteration) consisted of 640 image pairs (80% of the database) obtained by random selection, while the test set consisted of the remaining 160 image pairs (20% of the database). The learning process was also conducted on previous competitive models for fair comparison.

We compared the DAVI predictor with previous models developed by Yano *et al.* [16], Nojiri *et al.* [17], Choi *et al.* [18], and Kim and Sohn [19], the 3D accommodation-vergence mismatch (3DAVM) predictor [21] and the 3D visual discomfort predictor (3DVDP) [22]. To better understand the contributions of the various elements of the DAVI predictor, we also compared the performance of each subset of OF features, PF features, and the DAVI features in isolation from the other subsets. Tables III, IV and V show the performance of the tested models in terms of LCC, SROCC and RMSE, respectively. Clearly, each subset of proposed features is strongly predictive of visual discomfort, and when combined, exhibit significantly improved performance.

In order to study the degree of dependency of the performance of the model on the proportion between the training and testing sets, we measured the mean values of the LCCs over 2000 trials as a function of the percentage that was the training set as it ranged from 10% to 80% in 5% increments. Fig. 15 shows that the DAVI predictor is not strongly dependent on
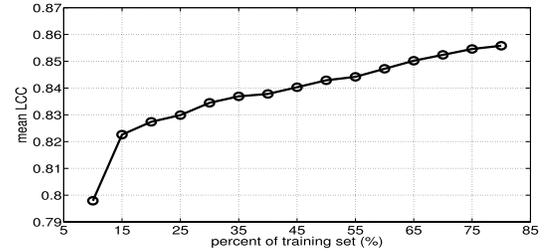

Fig. 15. Mean LCC performance of DAVI predictor as a function of the percentage of the training set.

the percentage of the overall database that is the training set. The performance difference varies less than 4% even as the percentage of entire database that is comprised by the training set is increased from 15% to 80%.

In order to study the degree of statistical significance of the proposed method, we conducted F-tests on the errors between the MOS and the predicted scores. The model residual R was used to test the statistical efficacy of the predictions against the MOS values as:

$$R = \{Q_i - \text{MOS}_i, i = 1, 2, \ldots, N_T\} \qquad (19)$$

where Q_i is the ith predicted score. The F-test was used to test one prediction model against another at the 95% significance level (i.e., at a $p = 0.05$ and critical F-value of 1.2991) and the results are tabulated in Table VI. A symbol "1" ("0") indicates the performance of the model in the row is superior (inferior) to the model in the column, while "−" indicates equivalence. The results show that all of the component features of the DAVI model contribute to the visual discomfort prediction power of the model. Moreover, the DAVI predictor is significantly superior to the methods of Yano *et al.* [16], Nojiri *et al.* [17], Choi [18] *et al.*, and Kim and Sohn [19]. Although the features of the DAVI predictor yield performance that is statistically equivalent to that of the recently developed 3DAVM [21] and 3DVDP models [22] at the 95% significant level, the DAVI method uses features that are different from, and complementary to those used in the other models. Table VII shows the mean LCC, SROCC and RMSE values over 2000 train-test trials by combining the proposed DAVI features including OF and PF with each of the previous models. The performances in every case are significantly improved compared to those of Tables III, IV and V. However, the combined models do not provide predictive performance that exceeds that of OF+PF+DAVI used in isolation. The results strongly indicate that the DAVI model effectively captures key factors predictive of visual discomfort that occurs when viewing S3D that are complementary to the other models.

C. Discussion

The simulation results indicate that the trained DAVI predictor achieves improved performance relative to existing models, which largely depend on measurements on excessive and relative disparities and other statistics of the computed depth, whereas deeper perceptual model features are absent [16]–[19]. While prior perceptual models have been deployed including models of the disparity tuning curves of

TABLE VI

RESULTS OF F-TEST PERFORMED ON THE MODEL RESIDUALS BETWEEN PREDICTED DISCOMFORT SCORES AND MOS VALUES AT THE SIGNIFICANT LEVEL OF 95%

	Yano	Nojiri	Choi	Kim	3DAVM	3DVDP	OF	PF	DAVI	OF+PF+DAVI
Yano	-	0	0	0	0	0	0	0	0	0
Nojiri	1	-	-	-	0	0	0	-	0	0
Choi	1	-	-	-	0	0	0	0	0	0
Kim	1	-	-	-	0	0	0	-	0	0
3DAVM	1	1	1	1	-	-	-	1	-	-
3DVDP	1	1	1	1	-	-	-	1	-	-
OF	1	1	1	1	-	-	-	-	-	-
PF	1	-	1	-	0	0	-	-	-	0
DAVI	1	1	1	1	-	-	-	-	-	-
OF+PF+DAVI	1	1	1	1	-	-	-	1	-	-

TABLE VII

PERFORMANCE OF 2000 TRIALS BY COMBINING FEATURES OF THE PROPOSED (DAVI+OF+PF) WITH PREVIOUS MODELS

	LCC	SROCC	RMSE
Yano [16]	0.8490	0.7780	0.4209
Nojiri [17]	0.8527	0.7785	0.4211
Choi [18]	0.8493	0.7779	0.4207
Kim [19]	0.8538	0.7809	0.4167
3DAVM [21]	0.8589	0.7846	0.4115
3DVDP [22]	0.8531	0.7794	0.4118

neurons in visual area MT and of local visual 3D bandwidth derived using principles of physiological optics [21], [22], the DAVI model complements these other recent models. Unlike other methods, the DAVI features capture potential conflicts between the cross-coupled accommodation and vergence control mechanisms when viewing S3D by quantitatively modeling oculomotor control mechanisms which are highly correlated with recorded human judgments of 3D visual discomfort.

While we have focussed on objective factors that are specifically predictive of visual discomfort experienced when viewing S3D, other factors also affect the overall quality of experience such as perceptual depth quality, image distortions arising from coding, texture gradients, perspective and occlusion. However, these factors need to be deeply analyzed in regards to their contribution to experienced visual discomfort. Looking forward, we plan to extend the DAVI model to be able to analyze and exploit temporal variations of S3D responses, such as adaptation and accumulation of fatigue. Towards this end, quantitative modeling of the slow fusional vergence subsystem in Schor's model would be valuable. We expect that such an approach would be useful for predicting visual discomfort not only on stationary S3D, but also on 3D videos.

VII. CONCLUSION

We explored a novel, physiologically inspired approach to the problem of predicting visual discomfort on S3D viewing. Dynamic physiological responses to viewed S3D pictures must be analyzed by accounting for the interactions that occur between the processes of accommodation and vergence. By suitably characterizing these responses, it is possible to more reliably predict 3D visual discomfort. This is accomplished in the DAVI model, which derives from the dual interaction transfer function model of Schor. The DAVI predictor exemplifies the manner by which physiological responses can be predicted from estimated S3D depth maps.

In the future, we plan to extend this work to encompass the experience of visual discomfort that is felt when viewing S3D

motion pictures (videos). A number of models for predicting video quality have been proposed [53], [55], [63], [64], but only a few have been proposed for predicting the quality of S3D viewing in terms of the degree of experienced visual discomfort [61], [62]. This much deeper problem will benefit by incorporating physiology-based temporal models of neuromuscular responses in the oculomotor system as well as functional models of space-time 3D perception to create an approach that can predict human responses to time-varying 3D content. Towards this goal, we would develop special-purpose features for 3D video to capture the influence of dynamic parameters related to object and camera motion. We also envision creating a database of videos with time-varying disparity statistics and human subject scores on them, similar to the study in [65], which deployed time-varying distortions. We envision that such predictive models would be useful for analyzing discomfort experienced when viewing 3D cinema, and could be useful for generating guidelines for cinematic stereographers.

REFERENCES

[1] C. Richardt, L. Świrski, I. P. Davies, and N. A. Dodgson, "Predicting stereoscopic viewing comfort using a coherence-based computational model," in *Proc. Int. Symp. Comput. Aesthetics Graph. Visualizat., Imag.*, Vancouver, BC, Canada, Aug. 2011, pp. 97–104.

[2] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, nos. 2–3, pp. 99–108, Aug. 2004.

[3] H. Sohn, Y. J. Jung, S.-I. Lee, and Y. M. Ro, "Predicting visual discomfort using object size and disparity information in stereoscopic images," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 28–37, Mar. 2013.

[4] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: A review," *Proc. SPIE*, vol. 6490, p. 64900I, Feb. 2007.

[5] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.*, vol. 8, no. 3, pp. 1–30, Mar. 2008.

[6] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, no. 4, pp. 141–150, Nov. 2004.

[7] 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.

[8] K. Ukai and P. A. Howarth, "Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations," *Displays*, vol. 29, no. 2, pp. 106–116, Mar. 2010.

[9] M. Emoto, T. Niida, and F. Okano, "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," *J. Display Technol.*, vol. 1, no. 2, pp. 328–340, Dec. 2005.

[10] S. Yang and J. E. Sheedy, "Effects of vergence and accommodative responses on viewer's comfort in viewing 3D stimuli," in *Proc. SPIE*, vol. 7863, p. 78630Q, Jan. 2012.

[11] 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, pp. 475–484, Feb. 2005.

[12] V. V. Krishnan and L. Stark, "A heuristic model for the human vergence eye movement system," *IEEE Trans. Biomed. Eng.*, vol. BME-24, no. 1, pp. 44–49, Jan. 1977.

[13] A. S. Eadie, L. S. Gray, P. Carlin, and M. Mon-Williams, "Modelling adaptation effects in vergence and accommodation after exposure to a simulated virtual reality stimulus," *Ophthalmic Physiol. Opt.*, vol. 20, no. 3, pp. 242–251, 2000.

[14] C. J. Erkelens, "A dual visual–local feedback model of the vergence eye movement system," *J. Vis.*, vol. 11, no. 10, pp. 1–14, 2011.

[15] Y.-F. Chen, Y.-Y. Lee, T. Chen, J. L. Semmlow, and T. L. Alvarez, "Behaviors, models, and clinical applications of vergence eye movements," *J. Med. Biol. Eng.*, vol. 30, no. 1, pp. 1–16, 2010.

[16] 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, no. 4, pp. 191–201, Sep. 2002.

[17] 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.

[18] J. Choi, D. Kim, S. Choi, and S. Sohn, "Visual fatigue modeling and analysis for stereoscopic video," *Opt. Eng.*, vol. 51, no. 1, pp. 017206-1–017206-11, Feb. 2012.

[19] 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.

[20] F. Qi, X. Fan, D. Zhao, T. Jiang, and J. Zhang, "A visual comfort assessment metric for stereoscopic images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015.

[21] 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.

[22] 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.

[23] 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.

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

[25] V. De Silva, A. Fernando, S. Worrall, H. K. Arachchi, and A. Kondoz, "Sensitivity analysis of the human visual system for depth cues in stereoscopic 3D displays," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 498–506, Jun. 2011.

[26] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Perception of Space and Motion*, W. Epstein and S. Rogers, Eds. San Diego, CA, USA: Academic, 1995, pp. 69–117.

[27] C. M. Schor, "The relationship between fusional vergence eye movements and fixation disparity," *Vis. Res.*, vol. 19, no. 12, pp. 1359–1367, 1979.

[28] C. M. Schor, "Adaptive regulation of accommodative vergence and vergence accommodation," *Amer. J. Optometry Physiol. Opt.*, vol. 63, no. 8, pp. 587–609, 1986.

[29] C. M. Schor and T. K. Tsuetaki, "Fatigue of accommodation and vergence modifies their mutual interactions," *Invest. Ophthalmolo. Vis. Sci.*, vol. 28, no. 8, pp. 1250–1259, 1987.

[30] D. T. McRuer and D. Graham, "Human pilot dynamics in compensatory system," Tech. Rep. AFFDL-TR-65-15, 1965.

[31] A. S. Eadie and P. J. Carlin, "Evolution of control system models of ocular accommodation, vergence and their interaction," *Med. Biol. Eng. Comput.*, vol. 33, pp. 517–524, Jul. 1995.

[32] J. Semmlow and P. Wetzel, "Dynamic contributions of the components of binocular vergence," *J. Opt. Soc. Amer. A*, vol. 69, no. 5, pp. 639–645, 1979.

[33] J. A. Bullinaria and P. M. Riddell, "Learning and evolution of control systems," *Neural Netw. World*, vol. 10, no. 4, pp. 535–544, 2000.

[34] W. Jaschinski, A. Švede, and S. Jainta, "Relation between fixation disparity and the asymmetry between convergent and divergent disparity step responses," *Vis. Res.*, vol. 48, no. 2, pp. 253–263, Jan. 2008.

[35] B. C. Jiang, "A method for correction of CA/C ratio based on linear model of accommodation and vergence," *Optometry Vis. Sci.*, vol. 71, no. 3, pp. 192–198, 1994.

[36] J. Maxwell, J. Tong, and C. M. Schor, "The first and second order dynamics of accommodative convergence and disparity convergence," *Vis. Res.*, vol. 50, no. 17, pp. 1728–1739, Aug. 2010.

[37] B. C. Jiang, "Accommodative vergence is driven by the phasic component of the accommodative controller," *Vis. Res.*, vol. 36, no. 1, pp. 97–102, Jan. 1996.

[38] "Subjective assessment of stereoscopic television pictures," ITU-R, Tech. Rep. BT.1438.

[39] "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Tech. Rep. BT.500-7.

[40] R. T. Held, E. A. Cooper, and M. S. Banks, "Blur and disparity are complementary cues to depth," *Current Biol.*, vol. 22, pp. 426–431, Mar. 2012.

[41] D. M. Hoffman and M. S. Banks, "Focus information is used to interpret binocular images," *J. Vis.*, vol. 10, no. 5, May 2010.

[42] 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, Jul. 2011.

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

[44] J. Wang, M. Barkowsky, V. Ricordel, and P. Le Callet, "Quantifying how the combination of blur and disparity affects the perceived depth," *Proc. SPIE*, vol. 7865, pp. 78650K-1–78650K-10, Feb. 2011.

[45] I. P. Howard, *Seeing in Depth: Basic Mechanisms*, vol. 1. Toronto, ON, Canada: Porteous Publishing, 2002.

[46] T. Ohshima, H. Yamamoto, and H. Tamura, "Gaze-directed adaptive rendering for interacting with virtual space," in *Proc. IEEE Virtual Reality Annu. Int. Symp. (VRAIS)*, Santa Clara, CA, USA, Mar./Apr. 1996, pp. 103–110 and 267.

[47] J. Li, M. Barkowsky, and P. Le Callet, "The influence of relative disparity and planar motion velocity on visual discomfort of stereoscopic videos," in *Proc. 3rd Int. Workshop Quality Multimedia Exper.*, Sep. 2011, pp. 155–160.

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

[49] T. Fukushima, M. Torii, K. Ukai, J. S. Wolffsohn, and B. Gilmartin, "The relationship between CA/C ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images," *J. Vis.*, vol. 9, no. 13, pp. 1–13, Dec. 2009.

[50] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[51] *IEEE Standards Association Stereoscopic (3D Imaging) Database*. [Online] Available: http://grouper.ieee.org/group/3dhf

[52] C.-C. Chang and C.-J. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm/

[53] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[54] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[55] A. C. Bovik, *The Essential Guide to Video Processing*. San Francisco, CA, USA: Academic, 2009.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," in *Proc. IEL*, Apr. 2008, pp. 1–8.

[61] S.-P. Du, B. Masia, S.-M. Hu, and D. Gutierrez, "A metric of visual comfort for stereoscopic motion," *ACM Trans. Graph.*, vol. 32, no. 6, Nov. 2013.

[62] Y. Bi and J. Zhou, "A visual comfort metric for stereoscopic 3D video based on SMDE approach," in *Proc. Int. Conf. Signal Process., Commun. Comput. (ISPCC)*, Hong Kong, Aug. 2012, pp. 72–77.

[63] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. I-869–I-872.

[64] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.

[65] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

Heeseok Oh received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree. His research interests include 2D/3D image and video processing based on human visual system, machine learning, and computational vision.

Sanghoon Lee (M'05–SM'12) received the B.S. degree from Yonsei University, Seoul, Korea, in 1989, the M.S. degree from the Korea Advanced Institute of Science and Technology, in 1991, and the Ph.D. degree 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 on 3G wireless and multimedia networks. In 2003, he joined the Department of Electrical and Electronics Engineering, Yonsei University, where he is currently a Full Professor. His research interests include image/video quality assessment, medical image processing, cloud computing, sensors and sensor networks, wireless 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 Chair of the IEEE P3333.1 Quality Assessment Working Group (2011-). He currently serves on the IEEE IVMSP Technical Committee (2014-); was a Technical Program Co-Chair of the International Conference on Information Networking (ICOIN) 2014, and 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 a 2015 Yonsei Academic Award from Yonsei University, a 2012 Special Service Award from the IEEE Broadcast Technology Society, and a 2013 Special Service Award from the IEEE Signal Processing Society.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) holds the Cockrell Family Endowed Regents Chair in Engineering at The University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering. He is a Faculty Member with the Department of Electrical and Computer Engineering and the Institute for Neuroscience. He has authored over 750 technical articles in these areas and holds several U.S. patents. His publications have been cited more than 45 000 times in the literature, his current H-index is about 75, and he is listed as a Highly-Cited Researcher by Thompson Reuters. His several books include the companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009). His research interests include image and video processing, digital television and digital cinema, computational vision, and visual perception.

Dr. Bovik received Televisions highest honor, an individual Primetime Emmy Award for Outstanding Achievement in Engineering Development from the Academy of Television Arts and Sciences (The Television Academy) in 2015, for his work on the development of video quality prediction models which have become standard tools in broadcast and postproduction houses throughout the television industry. He has also received a number of major awards from the IEEE Signal Processing Society, including: the Society Award (2013); the Technical Achievement Award (2005); the best paper award (2009); the Signal Processing Magazine Best Paper Award (2013); the Education Award (2007); the Meritorious Service Award (1998), and (coauthor) the Young Author Best Paper Award (2013). 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 is also a recipient of the Joe J. King Professional Engineering Achievement Award (2015) and the Hocott Award for Distinguished Engineering Research (2008) from the Cockrell School of Engineering 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 Pattern Recognition Society. He is a fellow of the Optical Society of America and the Society of Photo-Optical and Instrumentation Engineers, and is a member of the Television Academy and the National Academy of Television Arts and Sciences. He co-founded and was the longest-serving Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996-2002), created and served as the first General Chair of the IEEE International Conference on Image Processing, held in Austin, TX, in 1994, along with numerous other professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996-1998; Editorial Board, THE PROCEEDINGS OF THE IEEE, 1998-2004; and Series Editor of *Image, Video, and Multimedia Processing* (Morgan and Claypool Publishing Company, 2003-present). His was also the General Chair of the 2014 Texas Wireless Symposium, held in Austin, in 2014.

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