

3D Visual Discomfort Predictor: Analysis of Horizontal Disparity and Neural Activity Statistics

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

Abstract—Being able to predict the degree of visual discomfort that is felt when viewing stereoscopic 3D (S3D) images is an important goal toward ameliorating causative factors, such as excessive horizontal disparity, misalignments or mismatches between the left and right views of stereo pairs, or conflicts between different depth cues. Ideally, such a model should account for such factors as capture and viewing geometries, the distribution of disparities, and the responses of visual neurons. When viewing modern 3D displays, visual discomfort is caused primarily by changes in binocular vergence while accommodation is held fixed at the viewing distance to a flat 3D screen. This results in unnatural mismatches between ocular fixations and ocular focus that does not occur in normal direct 3D viewing. This accommodation vergence conflict can cause adverse effects, such as headaches, fatigue, eye strain, and reduced visual ability. Binocular vision is ultimately realized by means of neural mechanisms that subserve the sensorimotor control of eye movements. Realizing that the neuronal responses are directly implicated in both the control and experience of 3D perception, we have developed a model-based neuronal and statistical framework called the 3D visual discomfort predictor (3D-VDP) that automatically predicts the level of visual discomfort that is experienced when viewing S3D images. 3D-VDP extracts two types of features: 1) coarse features derived from the statistics of binocular disparities and 2) fine features derived by estimating the neural activity associated with the processing of horizontal disparities. In particular, we deploy a model of horizontal disparity processing in the extrastriate middle temporal region of occipital lobe. We compare the performance of 3D-VDP with other recent discomfort prediction algorithms with respect to correlation against recorded subjective visual discomfort scores, and show that 3D-VDP is statistically superior to the other methods.

Index Terms—Visual discomfort assessment, middle temporal neural activity, accommodation vergence conflict, stereoscopic 3D viewing, S3D, vergence.

Manuscript received December 29, 2013; revised May 5, 2014; accepted November 12, 2014. Date of publication December 18, 2014; date of current version February 11, 2015. This work was supported by the Ministry of Science, ICT and Future Planning, Korea, through the Information and Communication Technology Research and Development Program in 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sergio Goma.

J. Park, H. Oh, and S. Lee are with the Multidimensional Insight Laboratory, Department of Electrical and Electronics Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: dewofdawn@yonsei.ac.kr; 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, University of Texas at Austin, Austin, TX 78712 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>.

Digital Object Identifier 10.1109/TIP.2014.2383327

I. INTRODUCTION

STEREOSCOPIC 3D (S3D) multimedia services provide a more immersive quality of experience (QoE) by enabling depth perception. S3D perception brings a richer experience to viewers that is uniquely different from a 2D visual experience: a feeling of on-site presence in a 3D scene. However, unwanted side effects in the form of different types of visual discomfort can occur while one is participating in the stereoscopic experience. The possible sources of visual discomfort have been extensively studied with respect to safety and health issues, such as asthenopia (eyestrain), a feeling of pressure in the eyes, nausea, a reduced visual sensitivity, a reduced ability to accommodate and/or converge the two eyes, headaches and neck pain [1]–[3].

Several factors that can cause visual discomfort when viewing S3D have been identified. In [9], for example, the authors studied the issue of visual discomfort caused by misalignment of viewed S3D image pairs in regards to vertical and torsional disparities. They showed that these regressed factors are tightly correlated with experienced visual discomfort when they occur. In [10], the authors demonstrated that keystone artifacts captured by toed-in binocular capture systems also correlate with visual discomfort. The authors of [11] developed a visual comfort improvement technique based on the horizontal disparity range and on window violations in S3D content. They mentioned that window violations may cause severe discomfort. However, this type of distortion can generally be prevented during capture by aligning the main objects in the frame without window violation. Flawed presentations of horizontal disparity, such as excessively large or otherwise unnatural disparities, can also lead to severe visual discomfort [7], [8]. In [12], various other factors that could cause visual discomfort were reviewed, including optical distortions and motion parallax.

In the absence of geometrical distortions and window violations, factors related to horizontal disparity are the dominant factors that cause visual discomfort. Accordingly, here we focus on the horizontal disparity and on analyzing its neural activity statistics related to the perception of horizontal disparities. Visual discomfort caused by viewing 3D images typically results from a perceptual discordance of the depth signals perceived on a flat stereoscopic display. For example, under natural viewing conditions, the accommodation and vergence processes are connected with each other. Varying the

vergence via eye movement induces proportional changes in accommodation, and vice versa. However, when viewing a stereo image on a flat stereoscopic display, discrepancies may occur between the degree of accommodation required to achieve a sharp image for a given amount of vergence, which causes perceptual confusion and conflicts in the visual control system [4], [6].

Horizontal disparity is a fundamental depth cue that modifies the visual perception of the immediate 3D environment by inducing vergence movements, which are deeply related to visual discomfort [13]. The mechanical oculomotor movements that cause vergence are driven by cortical signaling from the brain, hence a good model of the appropriate neural responses to viewed S3D stimuli expressed in terms of horizontal disparity could be a very useful tool for predicting the degree of discomfort that is felt. We approach the problem under the assumption that no reference data describing the stereo image is available *a priori*. This type of assessment is a difficult problem, since the goal is to understand and predict the experience of viewing an image over a 3D visual space without an established reference for comparison. The problem is similar in this regard to recent blind image quality models for 2D and 3D images [14], [15], [20]–[22] that extract features from a training set of a database.

Numerous studies have studied the question of visual discomfort arising from horizontal disparity anomalies that are experienced when viewing stereo images. The authors of [23] and [24] report experimental studies on the effect of excessive horizontal disparity on visual comfort. Diplopia (double vision) begins when horizontal disparity exceeds Panum's fusional area, thereby causing visual discomfort [25]. The authors of [26]–[28] argue that the accommodation-vergence (AV) conflict is the primary cause of visual discomfort. In [26] and [27], a “comfort zone” of comfortable 3D viewing is defined that is limited by extremes of horizontal disparity within which clear single binocular vision can be achieved [4]. Several studies suggest a value of about $\pm 1^\circ$ (degree of visual angle) as a comfort limit, based on empirical measurements [12], [26]. In [16]–[20], the authors argue that the entire scene being viewed should be positioned in depth behind the viewing screen for a more comfortable viewing experience, implying that negative disparities induce more discomfort than do positive disparities, at least relative to the context provided by the fixed depth reference of the screen boundaries [29]. In addition, visual discomfort can also be caused by optical or geometrical misalignments between the left and right binocular images [30]–[32].

More recent efforts have been directed towards extracting measures of visual discomfort from the statistics of horizontal disparities. Yano *et al.* [26] computes the ratio of sums of horizontal disparities near the screen and those far from the screen. The horizontal disparities near and far are determined by defining the comfort zone to be 60 arcmin. The degree of actual experienced visual discomfort was recorded by human subjects viewing S3D movie clips along with measured waveforms of each viewer's accommodation response. The results on 6 subjects indicated that the computed horizontal disparity ratio closely relates to experienced visual discomfort

when viewing S3D. Nojiri *et al.* [20] compute a variety of discomfort factors from parameters of the distribution of experienced horizontal disparity, such as the minimum and maximum values, range, dispersion, absolute average, and average. They carried out a subjective study of experienced visual discomfort and sense of 3D presence on 20 subjects. The results indicate that the range of the horizontal disparity distribution has a high correlation with visual discomfort (~ 0.80). Choi *et al.* [21] distinguish three kinds of features: spatial, temporal, and differential components. The 3D spatial components derive from spatial depth complexity and depth position, calculated based on the variance and absolute mean of the disparity map, as a way of capturing both AV conflicts and excessive horizontal disparity. They find a high correlation (~ 0.77) between a model regressed on their computed features and the results of a subjective test involving 20 subjects. Kim *et al.* [22] proposed several metrics that predict 3D visual discomfort, including the experienced horizontal disparity range and maximum angular disparity, assuming a comfort zone of 60 arcmin. They found the range of maximum experienced angular disparity to have the highest correlation (~ 0.87) with the outcomes of the subjective test, among the features tested.

The use of statistical features such as these generally stems from the observation that larger horizontal disparities are more likely to cause severe visual discomfort. Horizontal disparity magnitude can provide a good predictor of 3D visual discomfort, yet a more elaborate statistical formulation of horizontal disparity should produce even better models of stereoscopic visual discomfort. Further, visual discomfort arises from other factors than the amplitude of horizontal disparity, and other 3D statistical features might also be relevant to visual discomfort, thereby deepening the available quantitative description of visual discomfort. This is the approach we take, using models of neural responses to derive more specific aspects of horizontal disparities.

We have developed a visual discomfort model and algorithm dubbed the 3D Visual Discomfort Predictor (3D-VDP), which extracts two types of statistical features. The first type is a “coarse” feature extracted from a horizontal disparity map. It is defined in terms of known causative factors of visual discomfort that have been identified by psychophysical studies of binocular vision. This follows the same basic philosophy as the statistical features used in previous models [16]–[22]. The other feature is a “fine” feature that is derived from a neural coding model used in computational neuroscience. The underlying assumption is that, since visual discomfort is mainly caused by changing the vergence eye movements while accommodation is fixed on a screen (resulting in AV conflict), stereo images requiring a similar degree of vergence would induce a similar level of visual discomfort. Thus, the fine features are defined in terms of estimated neural activity levels in the middle temporal (MT) region of the brain, which plays an important role in encoding horizontal disparity for vergence eye movements [34], [35]. In Section II, we take a broad view of the neural pathway along which horizontal disparity perception occurs and from which vergence eye movements are directed. Section III details the coarse/fine

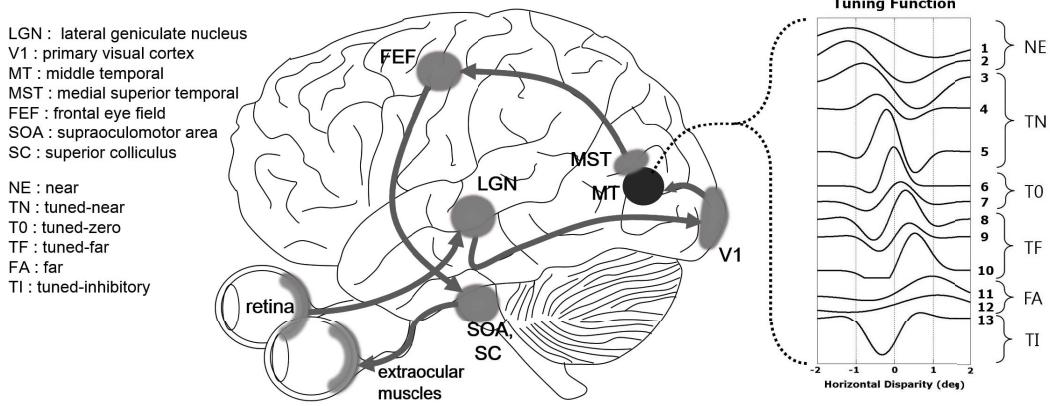


Fig. 1. Horizontal disparity and vergence control in the brain. Left: The neural pathways between horizontal disparity processing in cortical areas V1 and MT/MST and control of vergence eye movements by the extraocular (rectus) muscles [34], [36], [52], [53]. Right: 13 types of measured horizontal disparity tuning profiles exhibited by MT neurons [35]. See Section II and Section III-B for details.

statistical feature based model of visual discomfort that is used in 3D-VDP. The coarse and fine features are combined using a regression analysis, and visual discomfort is predicted using the regressed quality model.

II. NEURAL PROCESSING CONTROLLING VERGENCE EYE MOVEMENT

The main goal of vergence eye movement is to minimize the horizontal disparity of a fixated target object to near zero in order to simultaneously project the target onto the fovea of each eye. As shown in Fig. 1, eye movements are controlled via a feedback system between vision and optomotor control. While there are large cortical areas involved in 3D perception and numerous interconnections among them [36], we shall focus our attention on those areas along the neural pathway that are essential for accomplishing vergence eye movements.

When an image is projected onto the retina in the form of light, it is transformed into an electrical signal via transduction by the photoreceptors. The outputs of the photoreceptors are transmitted to the retinal ganglion cells via an intrinsic local neural network, the responses of which form the first receptive field (RF) of the visual system. This processed visual information is then relayed via the lateral geniculate nucleus (LGN) to primary visual cortex (area V1) [38]. The information from the two eyes is segregated until the LGN, and first combined in V1 [39]. Certain neurons in V1 are activated by stimuli from both eyes, and encode phase differences in horizontal disparity between the signals from the two eyes [40]. Broadly speaking, the separate neural pathways diverge from V1, termed the ventral and dorsal streams, both having a complete retinotopic mapping available. The ventral stream largely follows the path $V1 \rightarrow V2 \rightarrow V4 \rightarrow$ temporal lobe and is sometimes called the “*What Pathway*”, as processing is largely implicated with shape recognition and object representation [42]. The dorsal stream follows the path $V1 \rightarrow V2 \rightarrow MT \rightarrow$ parietal lobe and is sometimes called the “*Where Pathway*” as it is associated with motion computations, object locations and trajectories, and control of the eyes and arms. The secondary visual area, V2, is located next to V1 and is a gateway to the higher visual areas. The two streams also play distinct roles

in binocular depth perception. The neurons along the ventral stream create perceptual representations of 3D object shapes and the sense of 3D arrangements in space [43]. The neurons along the dorsal stream are predominantly involved in computations of low-level motion and horizontal disparity primitives, such as optical flow [44]. The dorsal stream encodes the sense of spatial arrangement and provides data used in the guidance of vergence eye movements [33], [34], [41].

Visual area MT is a key processing stage along the dorsal stream that plays important roles in motion perception, eye movements, and the computation and processing of binocular disparity. The visual responses of area MT neurons are tuned to attributes of the stimuli, such as retinal position, direction of motion, speed of motion, stimulus size, and binocular disparity [36], [46]. Early studies of binocular disparity processing focused on V1 since it is the first visual processing stage that encodes stereopsis, and therefore horizontal disparity tuning of MT is derivative of that in V1. However, recent studies indicate that MT plays a major role in subsequent horizontal disparity processing and horizontal disparity selectivity in this area is considerably stronger than in other cortical areas, such as V1 or V4, although neurons in V4 produce strong responses to relative disparities, as might be useful in the computation of 3D depths [35], [36], [41], [78]. The horizontal disparity tuning curves of MT neurons can be accurately described using the family of Gabor functions [35]. Although V1 neurons also have horizontal disparity tuning functions that are also well-modeled by Gabor functions, MT neurons exhibit a broader horizontal disparity tuning range than V1 neurons at comparable eccentricities [76]. Importantly, MT neurons directly feed medial superior temporal (MST) neurons [48], whose collective activity carries substantial information regarding the initiation of vergence eye movements [49]. Therefore, it is likely that the responses of MT neurons play a key role in the perception of depth as it relates to the guidance of vergence eye movements [41].

As such, our visual discomfort model includes neural features that describe activity in area MT. We make use of data reported in [35], which provides parametric fits to horizontal disparity tuning curves using Gabor functions for 13 typical

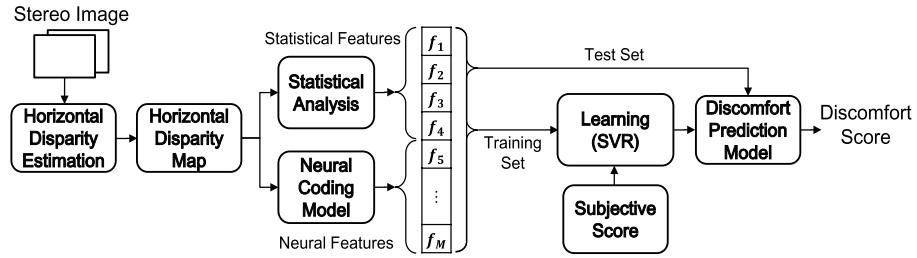


Fig. 2. Overall processing flow of the neural and statistical feature based 3D Visual Discomfort Predictor (3D-VDP). Statistical and neural features are extracted from the estimated horizontal disparity map of a stereo image pair. A support vector regressor (SVR) is trained on the extracted features and the subjective discomfort scores to construct a discomfort prediction model.

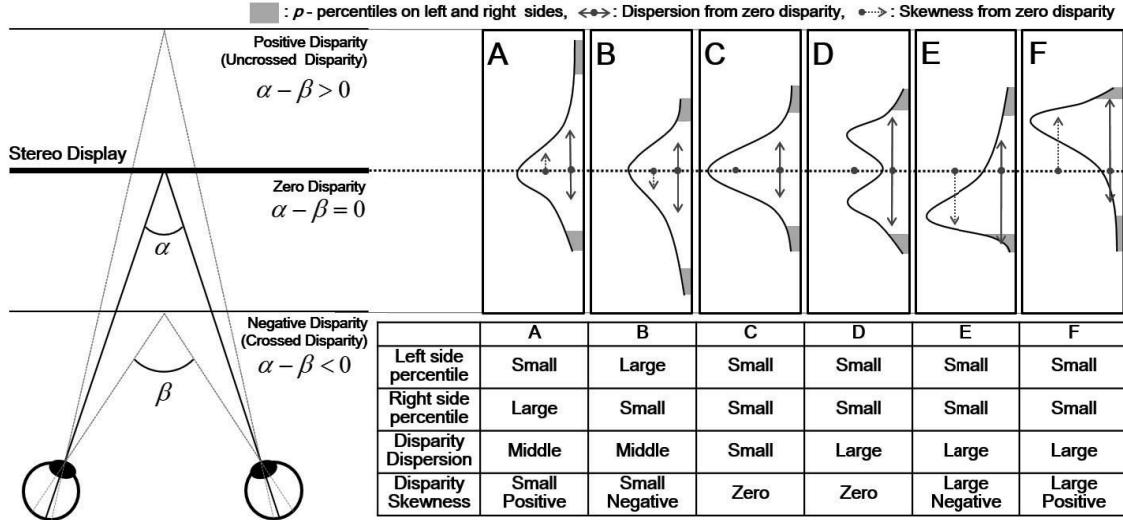


Fig. 3. Definition of horizontal disparity relations and examples of idealized empirical disparity distributions (histograms) along with descriptions of the statistical features computed from them.

MT neurons, as depicted on the right side of Fig. 1. Since neurons in area MST, which initiate vergence eye movements, receive most of their inputs from area MT [48], it appears that the horizontal disparity-selective MT neurons play a substantial role in the control of vergence eye movements. Further processes involved in vergence eye movements are summarized as follows. Since areas MT/MST have reciprocal connections with the frontal eye field (FEF), it is thought that the signals that guide vergence eye movements emanate from area MST to the FEF [50]. In addition, it has been suggested that area MST is also involved in early stages of processing visual signals for depth pursuit, while the FEF plays a primary role in the control of vergence eye movements by generating motor control signals, which are carried to the premotor neurons of the supra-oculomotor area (SOA) and the superior colliculus (SC) located in the brain stem. The SOA and the SC produce ocular motor signals that drive fast and slow vergence, respectively [51]–[54]. Finally, the eyeballs converge or diverge by action of the extraocular (rectus) muscles, which are controlled by premotor control circuits in the brain stem and cerebellum, which compute the final motor signals that drive vergence eye movements [54].

III. 3D VISUAL DISCOMFORT PREDICTOR

The overall processing flow of the 3D Visual Discomfort Predictor (3D-VDP) is depicted in Fig. 2. Two types of information are computed from the estimated horizontal

disparity map to form a feature vector that is predictive of visual discomfort. The first type derives from a statistical analysis of horizontal disparity. The second type extracts a predictive measure of neural activity in a brain center that is heavily implicated in both horizontal disparity processing and vergence eye movement control. The extracted features are learned, along with subjective S3D image discomfort scores recorded in a large human study using a support vector regressor (SVR). An aggregate visual discomfort score is computed using this predictive model trained on the IEEE Standard Association (IEEE-SA) stereo image database, which is publicly available at [55].

A. Statistical Analysis of Horizontal Disparity Maps

Horizontal disparity maps may present a variety of empirical distributions, for example, the idealized histograms depicted in plots A to F in Fig. 3. In the figure, α is the angle between the two eyes when verged at a fixation point on the display screen and β is the angle between projections onto the retina from points nearer or further from the viewer than the point of fixation. When the horizontal pixel disparity is zero, the angular disparity is zero, as depicted by the dashed line in Fig. 3. A stereo image may contain negative (crossed, $\alpha - \beta < 0$), or positive (uncrossed, $\alpha - \beta > 0$) disparities at points appearing in front of or behind the screen, respectively.

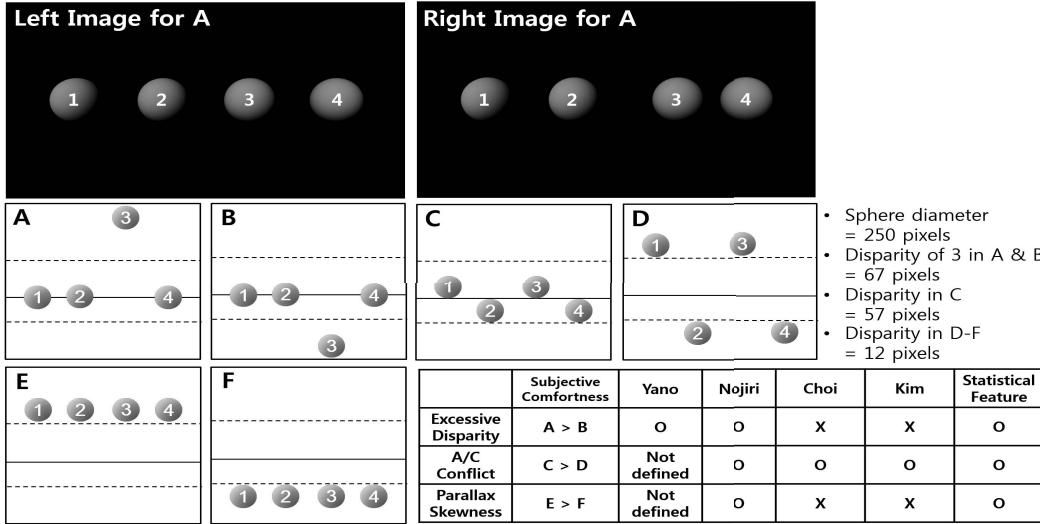


Fig. 4. Presentation used in a simple subjective test to compare statistical horizontal disparity features. The view is ‘from above’ in panels A-F. The top two panels are the S3D stimulus for case A as viewed by the subjects. The configurations A-F correspond to the distributions A-F in Fig. 3.

Excessively large, discomfort producing disparities can appear at either end of the horizontal disparity range. For example, in Fig. 3, the hypothetical distributions A and B present excessively large positive and negative disparities, respectively. Horizontal disparity events near both ends of the distribution may be good candidate features for describing excessive horizontal disparity. In addition, following the results in [16]–[20] and the experiment described in Fig. 4 (and in detail later), excessive negative disparities generally produce more discomfort than excessive positive disparities of the same magnitudes.

We use these observations as follows. Generally, it is known that the most severe local distortions have a large effect on the perceived quality of 2D images and videos [17], [18]. Likewise, we can assume that the most excessive disparities exert a significant effect on the degree of visual discomfort that is experienced. Therefore, compute the p^{th} -percentiles of both the left (lower) and right (higher) sides of the distribution:

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

$$f_2 = \frac{1}{d_{max}} \cdot \left(\frac{1}{N_p^r} \sum_{n > N \cdot (100-p)/100} d(n) \right), \quad (2)$$

where N is the total number of horizontal disparity values, N_p^l and N_p^r are the number of disparities within the lower and upper p^{th} -percentiles, respectively (p could be 5% or 10%, for example), $d(n)$ is the n^{th} disparity among the rank-ordered horizontal disparity values, and d_{max} is the maximum horizontal disparity. Since most of the disparities processed by area MT fall within the range -2° and $+2^\circ$ [35], we shall use $d_{max} = 2^\circ$. If the mean of the lower or upper p^{th} -percentile of horizontal disparity values is larger than d_{max} (lower than $-d_{max}$), we set $f_1 = 1$ or $f_2 = 1$ ($f_1 = -1$ or $f_2 = -1$), respectively.

AV conflicts occur when there are inconsistencies between the distances implied by vergence eye movements and those for accommodation to screen distance. Most non-zero disparities compel vergence eye movements, which can cause AV conflicts. Yet, it is not easy to predict the degree of an AV conflict precisely, since many internal and external factors influence the processes of accommodation and vergence, such as visual acuity, pupil size, age, luminance, contrast and accommodation-vergence coupling [4], [5]. However, there is a certain tendency that the greater the dispersion of the horizontal disparity distribution from zero, the more likely that an AV conflict occurs. A simple measure of dispersion relative to zero is:

$$f_3 = \frac{1}{d_{max}} \sqrt{\frac{1}{N} \sum_n d(n)^2}, \quad (3)$$

where, if $f_3 > 1$, set $f_3 = 1$. The distributions C and D in Fig. 3 have similar means but very different dispersion relative to zero disparity, which implies that a stereo image corresponding to D could induce a more severe AV conflict than one corresponding to C.

The distributions E and F have similar dispersions but different skewness of the horizontal disparity distributions. As mentioned above, negative disparities tend to induce greater degrees of visual discomfort than do positive disparities. Thus define a simple measure of skewness to capture the influence of the horizontal disparity distribution, f_4 :

$$f_4 = \frac{\sum_n d(n)}{\sum_n |d(n)|}. \quad (4)$$

If the horizontal disparity distribution is more concentrated on the negative (or positive) side of zero disparity, f_4 approaches -1 (or 1). The sign and magnitude of f_4 captures horizontal disparity skewness relative to zero disparity. In cases C and D, the disparities are symmetrically distributed around

zero disparity, hence $f_4 \approx 0$, and horizontal disparity skewness has little influence.

In order to better understand the role of statistical horizontal disparity features on experienced visual discomfort, we conducted a simple subjective study. Consider four numbered spheres laterally arranged along the horizontal as depicted in the top left and right images of Fig. 4. The four numbered spheres are variously positioned with disparities corresponding to the panels A-F in Fig. 4. The stimuli are 1920×1080 resolution S3D images containing spheres of diameter 250 pixels (about 13 centimeters in the display). The horizontal pixel disparities of the third balls in A and B were set to 67 pixels (angular disparities of 1.2°), 57 pixels for the spheres in C (angular disparity of 1°), and 12 pixels for the spheres in D-F (angular disparities are 0.2°). Panels A and B in Fig. 4 depict cases of large positive and negative excessive disparities, respectively. C and D in Fig. 4 demonstrate instances of very different disparity dispersions relative to zero disparity, corresponding to possible AV conflicts. Panels E and F show cases where a negatively skewed distribution of horizontal disparity incurs a greater degree of visual discomfort than does a positively skewed disparity. Panels A-F correspond to possible realizations of the distributions A-F in Fig. 3. In Fig. 4, the solid line represents the line of zero disparity, while the dotted lines represent the comfort zone used by Yano *et al.* [26] and Kim *et al.* [22]. The third spheres from the left in A and B have the same absolute disparity, while all of the spheres in E and F have the same absolute disparity. The subjective study was conducted using the same experimental environment described in Section IV. Sixteen subjects participated in the test. The subjects were asked to select the most comfortable stimulus amongst A against B, C against D, and E against F. All subjects consistently selected A, C, and E as more comfortable views than B, D, and F, respectively.

We calculated the features used in [20]–[22] and [26], to compare performance relative to features f_1 - f_4 . As shown in Fig. 4, only f_1 - f_4 were able to discriminate all of the differences. We also compared features used in previous studies. The feature used by Yano [26] is only applicable to cases A, B, and C. Since the feature is calculated as the sum of disparities outside the comfort zone, without disparities within the comfort zone, the feature cannot be defined for cases D, E, and F due to numerical instability. Since the features used by Choi [21] include the variance and absolute mean of disparity, it is difficult to discriminate between negative and positive disparities. The features used in Kim [22] include the disparity range and the sum of absolute maximum disparities, which also cannot distinguish between negative and positive disparities. The features of Nojiri [20] do allow for all the cases. However, the results obtained when correlating the features against subjective scores are not very good, as shown in Section IV.

B. Features From the Neural Population Coding Model

The neural interaction of accommodation and vergence in the midbrain can be modeled as a cross-coupled feedback

system [56]. A change of accommodation naturally alters vergence via the accommodation-vergence (AV) cross-link. Likewise, retinal disparity also modifies accommodation through the vergence-accommodation (VA) cross-link. However, when viewing a stereo image on a flat stereoscopic display, accommodation decisions produced in the midbrain conflict with horizontal disparity inferences produced by neural activity in area MT that guide vergence eye movements as a function of retinal disparity. Thus, we use a model of neural activity in area MT to derive features that can be used to automatically predict visual discomfort induced by AV conflicts. Specifically, we use a model of the responses of neurons in visual area MT that appear to be dedicated to both stereo perception and control of vergence eye movements.

Neural coding is a field of computational neuroscience concerned with identifying the relationship between a stimulus and the electrical responses of neurons [57]. In order to guide motor actions based on sensory information, neurons propagate signals in the form of electrical pulses called action potentials or spikes. The information contained within the signal is encoded as a pattern of action potentials in response to each input stimulus. The relationship between the stimuli and the responses of neurons in area MT can be modeled using population coding [46], [57], [58] whereby information is encoded based on the aggregate activity of populations of neurons [59].

Neural population codes are based on the neurophysiological finding that individual neurons selectively respond to particular variables underlying each stimulus. The selectivity is described by a tuning function representing the mean firing rate of the cell as a function of the variable. In [35], the authors formulated models of the tuning curves of visual area MT as functions of the amplitude of horizontal disparity. Gabor functions [60], [61], or Gaussian kernel functions modulated by sinusoidal carrier waves, were used to fit the curves, as depicted in the plots on the right side of Fig. 1. As described in [35], the curve-fit parameters were obtained by displaying moving random-dot stereograms containing a range of different disparities to each of three alert macaques and by quantifying the resulting measured MT neuron responses [35] (the visual system of monkeys closely resembles that of humans, and they perceive stereoscopic depth much as humans do [39]). The parameters of 13 exemplar tuning curves (from [35]) are given in Table I. The tuning function of the i^{th} typical MT neuron can be modeled as:

$$R_i(d) = R_0^i + A_i \cdot e^{-0.5((d-d_0^i)^2/\sigma_i^2)} \cdot \cos(2\pi f_i(d - d_0^i) + \Phi_i), \quad (5)$$

where d is horizontal disparity, R_0^i is the baseline response, A_i is the amplitude of the Gaussian kernel, d_0^i is the center of the Gaussian, σ_i is the width of the Gaussian, f_i is frequency, and Φ_i is the phase. We consider 13 representative neurons deemed typical of a much larger population of 501, and whose curve-fit parameters are given in [35].

Since MT cells are also selective for other variables such as velocity, in addition to horizontal disparity, it is assumed that the neurons are intrinsically noisy, hence the

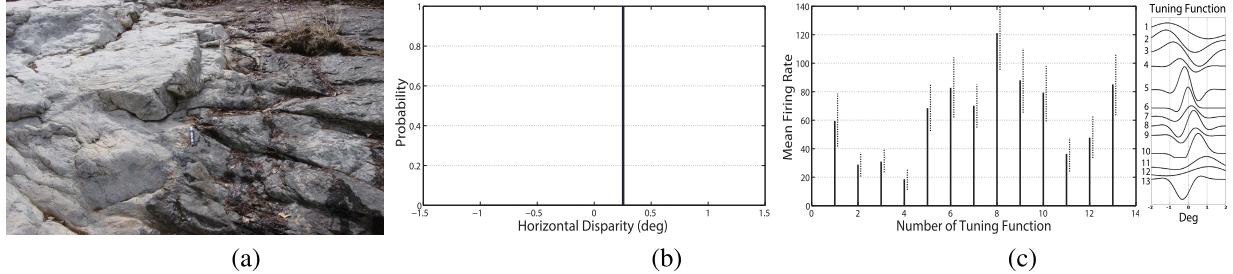


Fig. 5. The right image is obtained by locally shifting the left image using horizontal disparity values. (a) Left image. (b) Probability of horizontal disparity distribution where only one horizontal disparity exists. (c) Mean firing rate for each of a set of tuning functions assuming a poisson distribution of the population responses.

TABLE I
CURVE-FIT PARAMETERS FOR THE TUNING
FUNCTIONS OF FIG. 1 GIVEN IN [35]

Cell No. (i)	R_0^i	A_i	d_0^i	σ_i	f_i	Φ_i
1	72	44	-0.23	1.86	0.19	74
2	77	67	-0.46	1.16	0.25	86
3	81	73	0.15	1.07	0.28	123
4	41	42	-0.11	0.62	0.43	73
5	75	110	-0.04	0.53	0.51	40
6	32	124	-0.16	0.31	0.37	-51
7	24	51	-0.02	0.62	0.42	-38
8	51	77	0.04	0.67	0.50	-55
9	59	46	-0.01	0.57	0.49	-92
10	18	121	0.24	0.52	0.30	-61
11	16	49	0.81	1.01	0.21	-19
12	33	31	1.6	2.10	0.19	38
13	101	92	-0.23	0.56	0.33	-162

population coding model is approached using a probabilistic framework [58], [59], [62], [63]. The probability mass function of the firing rate r_i of the i^{th} neuron is often modeled as Poisson:

$$P[r_i|d] = e^{-R_i(d)} \frac{(R_i(d))^{r_i}}{r_i!}. \quad (6)$$

If there is a single horizontal disparity, as depicted in Fig. 5 (b) (left image is Fig. 5 (a), right image is the disparity shifted left image where horizontal disparity is d), and where d tunes a set of mean firing rates for the 13 typical MT neurons, using the tuning functions (5). Fig. 5 (c) shows firing rates obtained using the tuning functions of typical MT neurons when the input horizontal disparity is as shown in Fig. 5 (b). The actual spikes would be Poisson distributed about the mean firing rate as depicted by the dotted lines in Fig. 5 (c).

In (6), the firing rate r_i is probabilistically described using only a single horizontal disparity value d . However, sampled, discrete-space stereoscopic images contain multiple possible disparities, e.g., as shown in the horizontal disparity maps of Figs. 6 (e) and (f), whose left images

are Figs. 6 (a) and (b). An alternative model is required to deal with multiple disparities. The input disparities in Figs. 6 (e) and (f) can be modeled as realizations of a probability distribution, $P[d]$, as shown in Figs. 6 (i) and (j), respectively. A more comprehensive encoding model can be obtained using the extended Poisson model in [58]:

$$P[r_i|P[d]] = e^{-E[r_i]} \frac{E[r_i]^{r_i}}{r_i!}, \quad (7)$$

where $E[r_i]$ is the expected mean firing rate given the horizontal disparity probability distribution, $P[d]$:

$$E[r_i] = \sum_d P[d] \cdot R_i(d). \quad (8)$$

It should be noted that horizontal disparities are dependent on eccentricity in the retinal images. However, since we do not model the exact firing rate for a specific fixation point or for each position on the retina, but instead stochastically estimate the mean firing rate using the overall distribution of disparities, we do not consider the effect of eccentricity. Figs. 6 (m) and (n) show the estimated mean firing responses activated by the stereo images in Figs. 6 (a) and (b), respectively.

The expected mean firing rate in (8) is the shape parameter of the Poisson distribution of the action potentials. We calculate normalized neural features from the expected mean firing rates:

$$f_{i+4} = \frac{E[r_i]}{R_{\max}}, \quad 1 \leq i \leq 12, \quad (9)$$

where R_{\max} is the maximum MT neuron response. In the experimental data of [35], the response of the fifth cell exhibited the largest response at preferred disparity -0.2 among all MT neuronal responses, so we use $R_{\max} = R_5(-0.2)$ to normalize the feature values between $[0, 1]$.

Figures 6 (c) and (d), which show the left stereo images ‘OSL3_100’ and ‘ISS8_25’ in the IEEE-SA database, respectively, have similar expected means firing rates as in Figs. 6 (m) and (n), as shown in Figs. 6 (o) and (p), respectively. Although the spatial arrangement of action potentials would be different in real MT neurons, the distributions of expected action potentials are quite similar when comparing Figs. 6 (m) and (o). They have roughly similar horizontal disparity distributions as those in Figs. 6 (i) and (j), as shown in Figs. 6 (k) and (l), respectively. However, other elements,

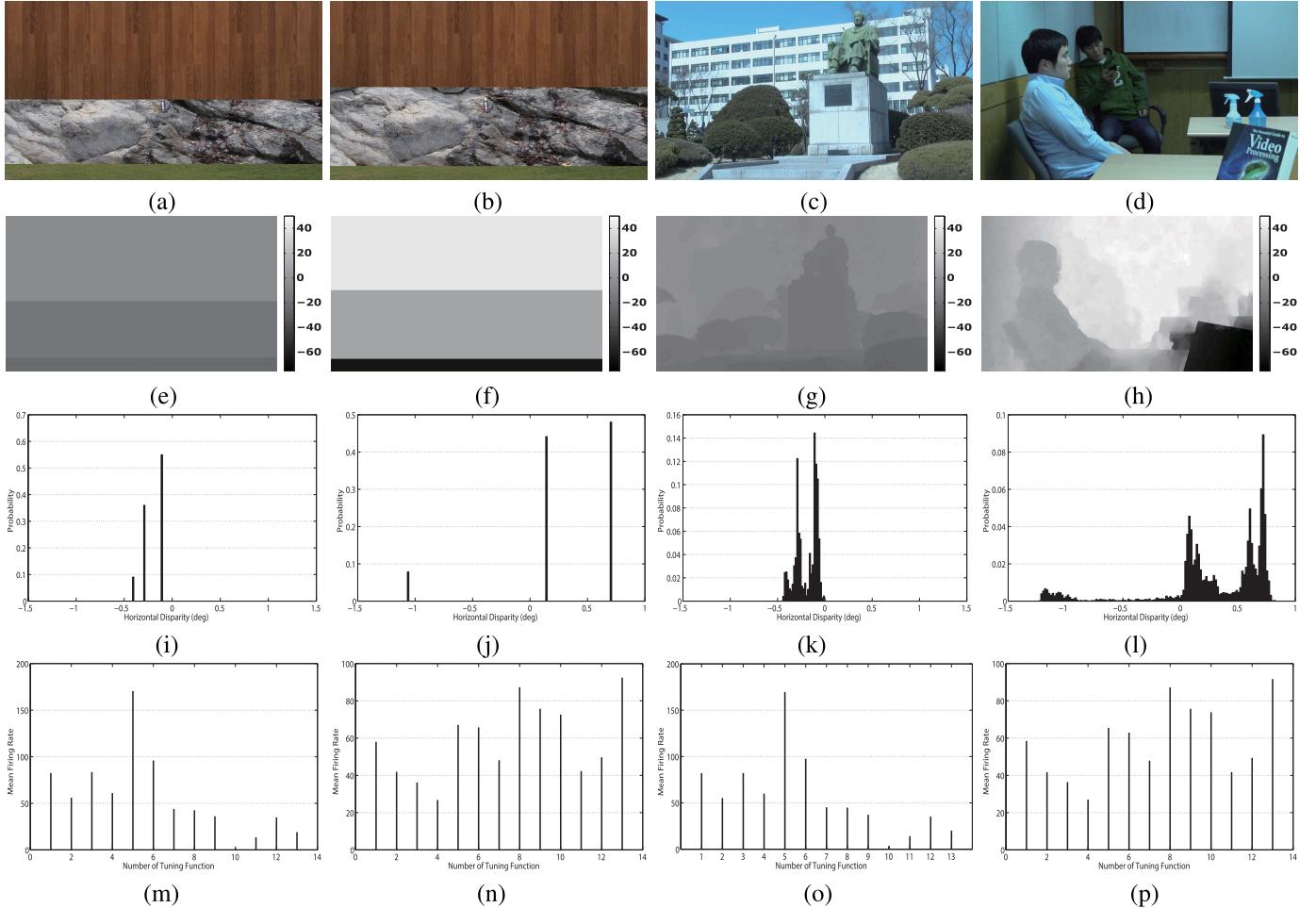


Fig. 6. Probability distribution of horizontal disparity and population responses. (a)-(b) Left stereo images composed of image patches having diverse disparity distributions. (c) Left image of the stereo image ‘OSL3_100’. (d) Left image of the stereo image ‘ISS8_25’. (e)-(h), (i)-(l) and (m)-(p) Horizontal disparity maps, probability distributions of horizontal disparity and estimated mean firing rates of the stereo images (a)-(d), respectively.

such as the horizontal disparity maps and other characteristics of the image, are quite different. Yet in the subjective tests, discomfort (MOS) values 3.9148, 1.7234, 3.8462, and 1.7692 were obtained for the stereo images in Figs. 6 (a)-(d), respectively. The test environment was as described in Section IV.

Fig. 7 shows examples where neural features are used to supplement statistical features. As can be seen in Figs. 7 (a) and (c), the statistical features are unable to discriminate between stereo images whose MOS are different. However, as may be seen in Figs. 7 (b) and (d), since the neural features more finely represent the distribution of disparities in the same way that MT neurons produce action potentials, the neural features discriminate between the different stereo images.

Fig. 8 shows the average mean firing rate after dividing the IEEE-SA database into bins of MOS of visual discomfort. The circle, rectangle, cross and triangle symbols denote the average mean firing rates for stereo images whose MOS are in the 0%–25%, 25%–50%, 50%–75% and 75%–100% bins, respectively. It may be observed that stereo images associated with low MOS tend to produce relatively high mean firing rates on MT neurons whose preferred horizontal disparity is

crossed, and vice versa. Since, in our model, stereo images that induce similar MT action potentials produce similar levels of subjective visual discomfort, the distribution of the action potentials presents a promising feature for predicting visual discomfort. Here, the important thing is that we extract reliable features based on a good model of the action potential that is generated when a human viewer perceives depth. Towards this end, the classic Gabor tuning function model is quite suitable [35]. The typical tuning functions shown in Table 1 clearly demonstrate the feasibility of using horizontal disparity tuned MT neural data to predict the degree of visual discomfort experienced when humans view S3D images. In Section V, it is demonstrated that these fine neural features effectively complement the coarse statistical features, giving rise to considerable performance improvement when predicting visual discomfort.

IV. IEEE-SA STEREO IMAGE DATABASE

In order to test 3D-VDP and other models that we and others are developing, we built the IEEE-SA stereo image database and conducted a subjective discomfort experiment [55].

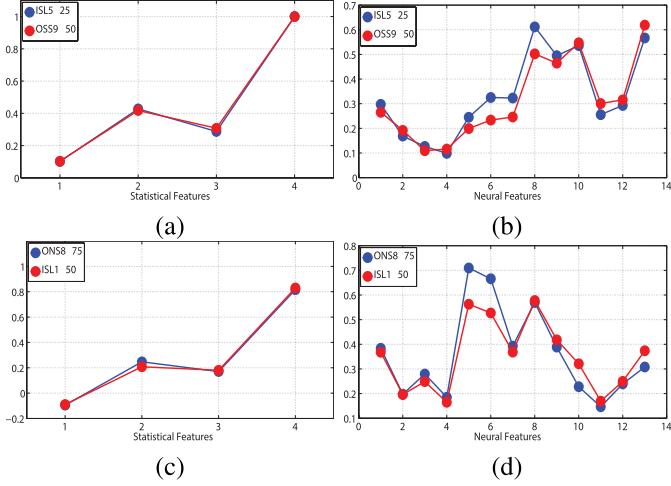


Fig. 7. Statistical (coarse) and neural (fine) features of the stereo images ‘ISL5_25’, ‘OSS9_50’, ‘ONS8_75’ and ‘ISL1_50’, whose MOSs are 2.8461, 3.4615, 2.6923 and 3.5384, respectively. (a) Statistical features of the stereo images ‘ISL5_25’ and ‘OSS9_50’ (b) Neural features of the stereo images ‘ISL5_25’ and ‘OSS9_50’ (c) Statistical features of the stereo images ‘ONS8_75’ and ‘ISL1_50’ (d) Neural features of the stereo image ‘ONS8_75’ and ‘ISL1_50’.

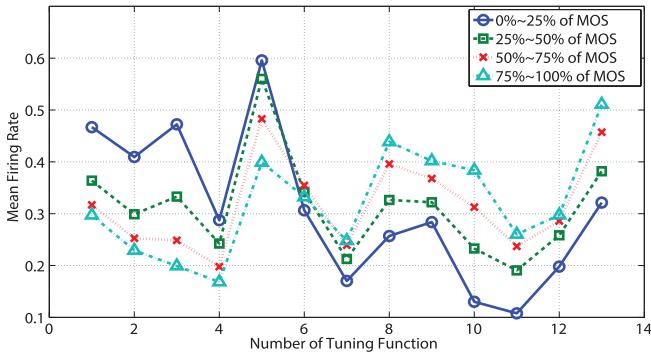


Fig. 8. Average of mean firing rate as a function of recorded subjective visual discomfort on the IEEE-SA database.

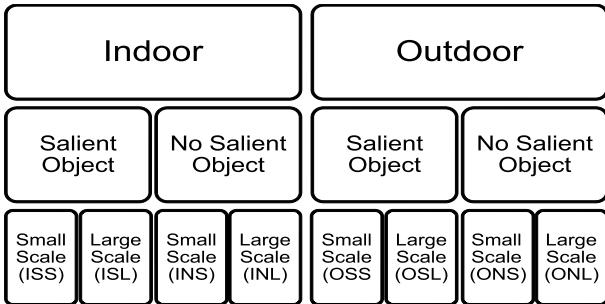


Fig. 9. Categories in the IEEE-SA database. The abbreviations of the 8 categories come derive the first letters of each category level. For example, ISS denotes the category ‘indoor - salient object - small scale’.

We divided the collected stereoscopic scenes into eight categories encompassing a diversity of shapes and depths, which are reasonably representative and challenging, as shown in Fig. 9. The scenes were divided into indoor and outdoor categories. Each category was then divided again according to



Fig. 10. Example images from the IEEE-SA Stereo Image Database. From top row to bottom row: ISS, ISL, INS, INL, OSS, OSL, ONS and ONL.

whether they contain ‘salient’ objects, such as people, dolls, cars, bikes, books, or sculptures. Finally, scene depth was estimated as the shooting distance, then category was again subdivided by the range of object depths in the scene. The categorization and labeling scheme is shown in Fig. 9. The stereo images in the categories ISS and INS were captured in small spaces (rooms, small offices and hallways), while category ISL and INL stereo pairs were captured in larger spaces, such as lobbies and large hallways. Category OSS and the OSL stereo pairs were distinguished by distances from the nearest salient object (OSS if closer than about 3 m, and OSL if farther). The ONS and ONL categories were roughly distinguished by the distance from the background in the scene (OSS if closer than about 5 m, and OSL if farther). Figure 10 shows example images from the IEEE-SA stereo image database, where each row corresponds to the eight categories, ranging from ISS to ONL as depicted in Fig. 9.

The IEEE-SA stereo image database includes a total of 800 stereo image pairs of high-definition (HD) resolution (1920×1080 pixels). The database was enriched by using multiple evenly separated convergence points on each scene. The convergence point was adjusted by shifting the sensors in the integrated twin-lens 3D camcorder, a PANASONIC AG-3DA1, thereby modifying the relative depth distribution between the observer and the screen. The apparatus was not toed-in, instead horizontal disparity was obtained by a parallel setup thereby avoiding keystone distortions [65]. Additionally, the captured S3D images are absent of vertical disparities because of the built-in precision aligned twin-lens system. The IEEE-SA stereo image database is composed of 160 such convergence-sampled sets so that each content category contains 20 sets.

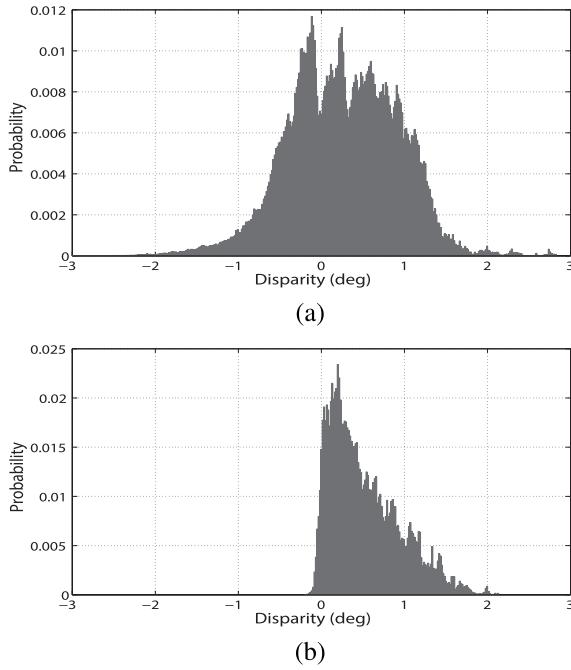


Fig. 11. Distribution of horizontal disparity (deg). (a) IEEE-SA stereo image database (b) EPFL stereo image database.

The IEEE-SA stereo image database includes highly diverse disparities. Figure 11 (a) shows that the overall horizontal disparity distribution over all 800 stereo image pairs is approximately normally distributed with a mean near zero, ranging from extremes of around -3° to $+3^\circ$. For simplicity, we obtained the horizontal disparity maps using the optical flow software from [66], available at [67]. We use the horizontal component of the computed ‘motion vectors’ computed between the left and right images as horizontal disparity. The choice of the optical flow software is motivated by the fact that this tool delivers competitive prediction of horizontal disparity as compared to the state of the art on the Middlebury Stereo Evaluation table [68], [69]. Since the optical flow algorithm does not assume an epipolar constraint [70], the computational complexity is somewhat higher than otherwise, but with the advantage of computing possibly better disparities.

Figure 11 (b) shows the total horizontal disparity distribution of the EPFL stereo image database [64]. The EPFL stereo image database consists of stereo images having resolution 1920×1080 pixels, with associated subjective opinion scores. Nine different scenes were captured using a rig-based 3D system with six cameras at varying distances ranging from $10 - 60$ cm, leading to a total of 54 stereo image pairs. Notice that the distribution is nearly one-sided, with mostly positive disparities.

The subjective discomfort assessment experiment was conducted in a laboratory environment, commensurate with standardized recommendations for subjective evaluation of picture quality [71]. The ratio of the luminance of an inactive screen to the peak luminance was below 0.02. The ratio of the luminance of the screen when displaying only black level in a completely dark room to that corresponding to peak white was about 0.01. The ratio of the luminance of the background

behind the picture monitor to the peak picture luminance was about 0.15. Otherwise, the room illumination was low. A 46-inch polarized stereoscopic monitor with HD (1920×1080) resolution was used to display the test stereo images. Each subject viewed the test stereo images at a distance of about 170 cm, or about three times the height of the monitor, as suggested in [72]. Twenty-eight subjects participated in the subjective test, with ages ranging from 22 to 38 years and an average of 28 years, which is nearly double the level recommended in ITU-R BT.500 [71]. All were non-experts in the fields of 3D image processing and quality assessment.

Each subject was asked to assign a visual discomfort score to each stereo test image using a Likert-like scale: 5 = very comfortable, 4 = comfortable, 3 = mildly comfortable, 2 = uncomfortable, and 1 = extremely uncomfortable. Due to the large number of test images in the IEEE-SA stereo image database, we divided the tests into nine separate sessions, one for training and eight for testing. During the training session, the subjects were instructed regarding the methodology of the test and the general range of comfort levels by showing them 20 stereo images broadly spanning the range of parameters in the database. In each session, the subjects each assessed 100 stereo image pairs, by first randomly shuffling the 800 stereo images in the IEEE-SA stereo image database, then evenly dividing them into eight sessions. A rest period of 10 minutes was inserted between sessions in order to reduce accumulated visual fatigue. Also, each subject participated in only four test sessions on a given day, and the remaining four sessions on another day. After completing the subjective tests, we discarded four outlier subjects that were detected according to the guideline described in [71]. Thus, MOS was computed using the results on 24 valid subjects.

V. STATISTICAL PERFORMANCE EVALUATION

3D-VDP is learned using a regression tool that maps feature vectors to predicted discomfort scores. Test and training sets were drawn from the IEEE-SA database along with the corresponding MOS. Regression was conducted using SVR [73], [74], which performs well on high-dimensional regression problems, and has been successfully utilized in previous NR-QA algorithms [14]. The libSVM package [75] was utilized to implement the SVR using the linear kernel, whose parameter was estimated by cross-validation during the training session. Since we used the linear kernel, there is only one parameter (i.e., the penalty parameter of the error term in the linear kernel).

We rigorously tested and compared 3D-VDP against the state of the art on the IEEE-SA stereo image database. We computed the Spearman rank order correlation coefficient (SROCC), Pearson linear correlation coefficient (LCC), and root mean square error (RMSE) between predicted and subjective scores to evaluate the discomfort prediction power of all of the compared algorithms. The database was subdivided into 80% of the stereo pairs for each training set and 20% for test set (every training set and subsequent test set were made to be entirely content-separate). Specifically, since each

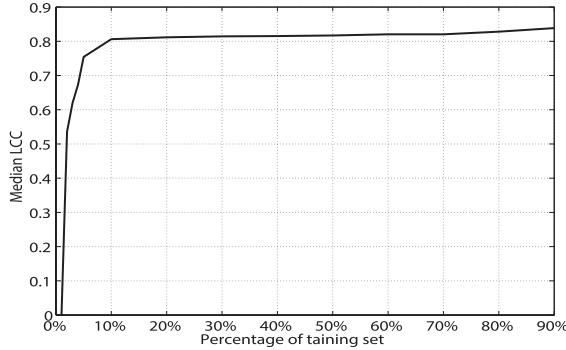


Fig. 12. Median LCC of Visual Discomfort Predictor as a function of the percentage of the IEEE-SA stereo image database comprised by the training set (over 2000 iterations).

TABLE II
LCC OVER 2000 TRIALS OF RANDOMLY CHOSEN
TRAIN AND TEST SETS ON IEEE-SA DATABASE

	Mean	Median	Standard Deviation
Yano [26]	0.3988	0.4045	0.0748
Nojiri [20]	0.6854	0.6935	0.0788
Choi [21]	0.6509	0.6565	0.0783
Kim [22]	0.7018	0.7113	0.0771
Statistical 3D-VDP	0.7930	0.7996	0.0543
Neural 3D-VDP	0.7337	0.7391	0.0547
3D-VDP	0.8310	0.8388	0.0526

TABLE III
SROCC OVER 2000 TRIALS OF RANDOMLY CHOSEN
TRAIN AND TEST SETS ON IEEE-SA DATABASE

	Mean	Median	Standard Deviation
Yano [26]	0.3363	0.3384	0.0732
Nojiri [20]	0.6108	0.6155	0.0732
Choi [21]	0.5851	0.5909	0.0798
Kim [22]	0.6151	0.6195	0.0700
Statistical 3D-VDP	0.7110	0.7151	0.0493
Neural 3D-VDP	0.6392	0.6397	0.0502
3D-VDP	0.7534	0.7514	0.0498

category contains 20 sets of stereo image pairs, 18 sets were chosen for training and 2 for testing, respectively, for each category. In order to ensure that the results were not built on a specific train-test separation, we iterated the train-test sequence 2000 times using randomly chosen training and test sets. In addition, to determine whether the discomfort prediction models were dependent on the training data, we also found the median LCC as a function of the percentage of the overall dataset that the training set comprised over the 2000 trials, as shown in Fig. 12. This percentage was varied from 1% to 90%. While the LCC decreased with decreasing training set percentage, the reduction in performance was not significant until the training set fell below 10% of the overall database.

TABLE IV
RMSE OVER 2000 TRIALS OF RANDOMLY CHOSEN
TRAIN AND TEST SETS ON IEEE-SA DATABASE

	Mean	Median	Standard Deviation
Yano [26]	0.7602	0.7612	0.0730
Nojiri [20]	0.6088	0.6048	0.0838
Choi [21]	0.7096	0.7106	0.0839
Kim [22]	0.5838	0.5804	0.0594
Statistical 3D-VDP	0.4602	0.4540	0.0566
Neural 3D-VDP	0.5204	0.5199	0.0586
3D-VDP	0.4489	0.4378	0.0557

TABLE V
LCC OVER 2000 TRIALS BY COMBINING FEATURES
OF THE PROPOSED AND PREVIOUS MODELS

	Yano [26]	Nojiri [20]	Choi [21]	Kim [22]
Neural 3D-VDP	0.7318	0.7364	0.7721	0.7880
Statistical 3D-VDP	0.7915	0.8061	0.8036	0.8092
3D-VDP	0.8120	0.8221	0.8233	0.8248

TABLE VI
LCC, SROCC AND RMSE OF COMPARED
MODELS ON EPFL DATABASE

	LCC	SROCC	RMSE
Yano [26]	0.8475	0.8626	0.6918
Nojiri [20]	0.8589	0.8721	0.4823
Choi [21]	0.7731	0.8231	0.7397
Kim [22]	0.8693	0.8720	0.4660
Statistical 3D-VDP	0.8697	0.8787	0.5004
Neural 3D-VDP	0.8132	0.7776	0.5629
3D-VDP	0.8882	0.8896	0.4346

The mean, median, and standard deviations of the LCC, SROCC, and RMSE computed across the 2000 train-test trials is tabulated in Tables II-IV for all of the discomfort prediction models considered. SVR was utilized to train all of the models to achieve a fair comparison. In the Tables, “3D-VDP” is used as a shorthand for the 3D Visual Discomfort Predictor, while “Statistical 3D-VDP” uses only the features explained in Section III-A, “Neural 3D-VDP” uses only the features developed in Section III-B and “3D-VDP” uses both the neural and statistical features. Clearly, 3D-VDP delivers significantly better predictive performance than the other models in terms of both correlation and reliability. Moreover, while Neural 3D-VDP does not supply standout performance when used alone, the complementary information it contributes, when combined with statistical 3D-VDP, leads to considerable performance improvement. In addition, in Table V, we measured the efficacy of the neural and statistical features by applying them to conventional models. It was observed that the

TABLE VII
RESULTS OF THE F-TEST PERFORMED ON THE RESIDUALS BETWEEN OBJECTIVE VISUAL
DISCOMFORT PREDICTIONS AND MOS VALUES AT A SIGNIFICANCE LEVEL OF 99.9%

	Yano	Nojiri	Choi	Kim	Neural 3D-VDP	Statistical 3D-VDP	3D-VDP
Yano	—	0	—	—	0	0	0
Nojiri	1	—	—	—	—	0	0
Choi	—	—	—	—	—	0	0
Kim	1	—	—	—	—	0	0
Neural 3D-VDP	1	—	—	—	—	0	0
Statistical 3D-VDP	1	1	1	1	1	—	—
3D-VDP	1	1	1	1	1	—	—

LCC values were significantly improved compared to those of Table II. We obtained the LCC values over the 2000 trials by combining the features of the proposed models with the previous model. However, these levels did not exceed the performance reached by 3D-VDP alone, suggesting that no reverse improvement occurs.

In order to demonstrate the database independence of 3D-VDP and that the training process is only a calibration, we performed additional testing on the EPFL stereo image database. We trained 3D-VDP on the entire IEEE-SA database, then tested the trained model on the EPFL database. The performance results and comparisons with the other models are given in Table VI. Since the distribution of horizontal disparity is strongly biased toward positive disparity on this database, and since the number of stereo images is small and spans a smaller range of vergence angles and disparities, the performance results of all the models are inflated. Nevertheless, the performance of 3D-VDP is quite competitive, although the capture system, the horizontal disparity distributions, and the visual content of the EPFL database are different from those of the IEEE-SA database.

Table VII shows the results of F-tests conducted to assess the statistical significance of the errors between the MOS scores and the model predictions on the IEEE-SA database. The residual error between the predicted score of a discomfort prediction model and the corresponding MOS value in the IEEE-SA database can be used to test the statistical efficacy of the model against other models. The residual errors between the model predictions and the MOS values are

$$R = \{Q_i - \text{MOS}_i, \quad i = 1, 2, \dots, N_T\} \quad (10)$$

where Q_i is the i^{th} objective visual discomfort score and MOS_i is the corresponding i^{th} MOS score. The F-test was used to compare one objective model against another objective model at the 99.9% significance level (i.e., at a p-level of 0.001 and critical F-value of 1.6378 when the degrees of freedom were 159 for both numerator and denominator). Table VII is the result of the F-test. A symbol value of “1” indicates that the statistical performance of the model in the row is superior to that of the model in the column, while “0” indicates the performance in the row is inferior to that in the column, and “-” indicates equivalent performance. The results indicate that 3D-VDP achieves better performance than the prior models with statistical significance.

VI. CONCLUSIONS

The 3D Visual Discomfort Predictor extracts two kinds of features: coarse statistical features computed from a horizontal disparity map, and fine features indicative of likely induced neural activity in a central processing stage of horizontal disparity perception and vergence eye movement. In the future, we plan to generalize measures of “3D naturalness” on stereoscopic images to improve the process of visual discomfort prediction, by including other factors such as geometrical distortions and window violations. The idea of that direction of inquiry is that stereo pairs associated with “natural” reconstructions, e.g., that closely conform to data-driven 3D natural scene models [76], [77], will be comfortable to view (assuming a human viewing geometry).

REFERENCES

- [1] T. W. Dillon and H. H. Emurian, “Some factors affecting reports of visual fatigue resulting from use of a VDU,” *Comput. Human Behaviour*, vol. 12, no. 1, pp. 49–59, 1996.
- [2] M. Emoto, T. Niida, and F. Okana, “Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television,” *J. Display Technol.*, vol. 1, no. 2, pp. 328–340, 2005.
- [3] J. S. Cooper, C. R. Burns, S. A. Cotter, K. M. Daum, J. M. Griffin, and M. M. Scheiman, “Optometric clinical practice guideline care of the patient with accommodative and vergence dysfunction,” Amer. Optometric Assoc., St. Louis, MO, USA, Tech. Rep., 2001.
- [4] 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.
- [5] T. Fukushima, M. Torii, K. Ukai, J. S. Wolfsohn, 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. 2008.
- [6] 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, Art. ID 11.
- [7] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, “A survey of perceptual evaluations and requirements of three-dimensional TV,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 381–391, Mar. 2004.
- [8] F. L. Kooi and A. Toet, “Visual comfort of binocular and 3D displays,” *Displays*, vol. 25, nos. 2–3, pp. 99–108, 2004.
- [9] C. W. Tyler, L. T. Likova, K. Atanassov, V. Ramachandra, and S. Goma, “3D discomfort from vertical and torsional disparities in natural images,” *Proc. SPIE*, vol. 8291, pp. 82910Q-1–82910Q-9, Feb. 2012.
- [10] F. Liu, Y. Niu, and H. Jin, “Keystone correction for stereoscopic cinematography,” in *Proc. IEEE Workshop 3D Cinematograph.*, Jun. 2012, pp. 1–7.
- [11] Y. Jung, H. Sohn, S.-I. Lee, and Y. Ro, “Visual comfort improvement in stereoscopic 3D displays using perceptually plausible assessment metric of visual comfort,” *IEEE Trans. Consum. Electron.*, vol. 60, no. 1, pp. 1–9, Apr. 2014.

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

[13] T. Bando, A. Iijima, and S. Yano, "Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review," *Displays*, vol. 33, no. 2, pp. 76–83, Apr. 2012.

[14] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

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

[16] S. Ide, H. Yamanoue, M. Okui, and F. Okano, M. Bitou, and N. Terashima, "Parallax distribution for ease of viewing in stereoscopic HDTV," *Proc. SPIE*, vol. 4660, pp. 38–45, May 2002.

[17] 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.

[18] 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.

[19] Y. Nojiri, H. Yamanoue, S. Ide, S. Yano, and F. Okano, "Parallax distribution and visual comfort on stereoscopic HDTV," in *Proc. IBC*, 2006, pp. 373–380.

[20] 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.

[21] J. Choi, D. Kim, S. Choi, and K. Sohn, "Visual fatigue modeling and analysis for stereoscopic video," *Opt. Eng.*, vol. 51, no. 1, pp. 017206-1–017206-11, Jan. 2010.

[22] 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.

[23] M. Wopking, "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," *J. Soc. Inf. Display*, vol. 3, no. 3, pp. 101–103, Dec. 1995.

[24] Y. Nojiri, H. Yamanoue, A. Hanazato, M. Emoto, and F. Okano, "Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing," *Proc. SPIE*, vol. 5291, pp. 303–313, Jan. 2004.

[25] Y.-Y. Yeh and L. D. Silverstein, "Limits of fusion and depth judgment in stereoscopic color displays," *Human Factors*, vol. 32, no. 1, pp. 45–60, Feb. 1990.

[26] 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, Jun. 2002.

[27] 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, Oct. 2004.

[28] M. Emoto, Y. Nojiri, and F. Okano, "Changes in fusional vergence limit and its hysteresis after viewing stereoscopic TV," *Displays*, vol. 25, nos. 2–3, pp. 67–76, Aug. 2004.

[29] M.-J. Chen, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Optimizing 3D image display using the stereoacuity function," in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 617–620.

[30] F. Speranza and L. M. Wilcox, "Viewing stereoscopic images comfortably: The effects of whole-field vertical disparity," *Proc. SPIE*, vol. 4660, pp. 18–25, May 2002.

[31] W. A. IJsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 225–233, Mar. 2000.

[32] A. J. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proc. SPIE*, vol. 1915, pp. 36–48, Sep. 1993.

[33] B. T. Backus, D. J. Fleet, A. J. Parker, and D. J. Heeger, "Human cortical activity correlates with stereoscopic depth perception," *J. Neurophysiol.*, vol. 86, no. 4, pp. 2054–2068, Oct. 2001.

[34] P. Neri, "A stereoscopic look at visual cortex," *J. Neurophysiol.*, vol. 93, no. 4, pp. 1823–1826, 2005.

[35] G. C. DeAngelis and T. Uka, "Coding of horizontal disparity and velocity by MT neurons in the alert macaque," *J. Neurophysiol.*, vol. 89, no. 2, pp. 1094–1111, 2003.

[36] A. W. Roe, A. J. Parker, R. T. Born, and G. C. DeAngelis, "Disparity channels in early vision," *J. Neurosci.*, vol. 27, no. 44, pp. 11820–11831, Oct. 2007.

[37] L. R. Squire, Ed., *Fundamental Neuroscience*. San Diego, CA, USA: Academic, 1999.

[38] L. M. Martinez and J.-M. Alonso, "Complex receptive fields in primary visual cortex," *Neuroscientist*, vol. 9, no. 5, pp. 317–331, Oct. 2003.

[39] J. Read, "Early computational processing in binocular vision and depth perception," *Progr. Biophys. Molecular Biol.*, vol. 87, no. 1, pp. 77–108, 2005.

[40] B. G. Cumming and G. C. DeAngelis, "The physiology of stereopsis," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 203–238, 2001.

[41] G. C. DeAngelis, B. G. Cumming, and W. T. Newsome, "Cortical area MT and the perception of stereoscopic depth," *Nature*, vol. 394, no. 6694, pp. 677–680, Aug. 1998.

[42] L. G. Ungerleider and M. Mishkin, *Analysis of Visual Behavior*. Cambridge, MA, USA: MIT Press, 1982.

[43] P. Janssen, R. Vogels, and G. A. Orban, "Three-dimensional shape coding in inferior temporal cortex," *Neuron*, vol. 27, no. 2, pp. 385–397, Aug. 2000.

[44] 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.

[45] J. D. Nguyenkim and G. C. DeAngelis, "Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons," *J. Neurosci.*, vol. 23, no. 18, pp. 7117–7128, Aug. 2003.

[46] R. T. Born and D. C. Bradley, "Structure and function of visual area MT," *Annu. Rev. Neurosci.*, vol. 28, pp. 157–189, Mar. 2005.

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

[48] J. P. Roy, H. Komatsu, and R. H. Wurtz, "Disparity sensitivity of neurons in monkey extrastriate area MST," *J. Neurosci.*, vol. 12, no. 7, pp. 2478–2492, 1992.

[49] A. Takemura, Y. Inoue, K. Kawano, C. Quaia, and F. A. Miles, "Single-unit activity in cortical area MST associated with disparity vergence eye movements: Evidence for population coding," *J. Neurophysiol.*, vol. 85, no. 5, pp. 2245–2266, 2001.

[50] T. Akao, M. J. Mustari, J. Fukushima, S. Kurkin, and K. Fukushima, "Discharge characteristics of pursuit neurons in MST during vergence eye movements," *J. Neurophysiol.*, vol. 93, no. 5, pp. 2415–2434, May 2005.

[51] A. Wong, *Eye Movement Disorders*. London, U.K.: Oxford Univ. Press, 2007.

[52] P. D. R. Gamlin, "Neural mechanisms for the control of vergence eye movements," *Ann. New York Acad. Sci.*, vol. 956, no. 1, pp. 264–272, Apr. 2002.

[53] U. Buttner and J. A. Buttner-Ennever, "Present concepts of oculomotor organization," *Prog. Brain Res.*, vol. 151, pp. 1–42, 2006.

[54] U. Schwarz, "Neuroophthalmology: A brief vademecum," *Eur. J. Radiol.*, vol. 49, no. 1, pp. 31–63, 2004.

[55] J. Park, H. Oh, and S. Lee. (2012). *IEEE-SA Stereo Image Database*. [Online]. Available: <http://grouper.ieee.org/groups/3dhf/>

[56] P. D. R. Gamlin, "Subcortical neural circuits for ocular accommodation and vergence in primates," *Ophthalmic Physiol. Opt.*, vol. 19, no. 2, pp. 81–89, 1999.

[57] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, Eds., *Principles of Neural Science*. New York, NY, USA: Elsevier, 1990.

[58] R. S. Zemel, P. Dayan, and A. Pouget, "Probabilistic interpretation of population codes," *Neural Comput.*, vol. 10, no. 2, pp. 403–430, 1998.

[59] T. D. Sanger, "Neural population codes," *Current Opinion Neurobiol.*, vol. 13, no. 2, pp. 238–249, 2003.

[60] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.

[61] M. Clark and A. C. Bovik, "Experiments in segmenting texton patterns using localized spatial filters," *Pattern Recognit.*, vol. 22, no. 6, pp. 707–717, 1989.

[62] S. Wu, S. Amari, and H. Nakahara, "Population coding and decoding in a neural field: A computational study," *Neural Comput.*, vol. 14, no. 5, pp. 999–1026, 2002.

[63] L. Paninski, J. Pillow, and J. Lewi, "Statistical models for neural encoding, decoding, and optimal stimulus design," *Prog. Brain Res.*, vol. 165, pp. 493–507, Aug. 2007.

[64] L. Goldmann, F. De Simone, and T. Ebrahimi, "Impact of acquisition distortion on the quality of stereoscopic images," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron. (VPQM)*, 2010.

- [65] F. Zilly, J. Kluger, and P. Kauff, "Production rules for stereo acquisition," *Proc. IEEE*, vol. 99, no. 4, pp. 590–606, Apr. 2011.
- [66] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2432–2439.
- [67] D. Sun, S. Roth, and M. Black. (2010). *Optical Flow Software*. [Online]. Available: <http://www.cs.brown.edu/~black/>
- [68] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7–42, Apr. 2002.
- [69] D. Scharstein and R. Szeliski. *Middlebury Stereo Evaluation—Version 2*. [Online]. Available: <http://vision.middlebury.edu/stereo/>
- [70] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [71] "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Geneva, Switzerland, Tech. Rep. BT.500-11, 2002.
- [72] "Subjective assessment of stereoscopic television pictures," ITU-R, Geneva, Switzerland, Tech. Rep. BT.1438, 2000.
- [73] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.
- [74] 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.
- [75] C. Chang and C. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: <http://www.csie.ntu.edu.tw/cjlin/libsvm/>
- [76] 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.
- [77] 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.
- [78] K. Umeda, S. Tanabe, and I. Fujida, "Representation of stereoscopic depth based on relative disparity in macaque area V4," *J. Neurophysiol.*, vol. 98, no. 1, pp. 241–252, 2007.



Jincheol Park was born in Korea in 1982. He received the B.S. degree in information and electronic engineering from Soongsil University, Seoul, Korea, in 2006, and the M.S. and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, in 2008 and 2013, respectively. He was a Visiting Researcher under the guidance of Prof. A. C. Bovik with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA, from 2010 to 2011. His current research interests include 2D and 3D video quality assessment.



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. He is currently pursuing the Ph.D. degree since 2012. His research interests include 2D/3D image and video processing based on human visual system, and quality assessment of 2D/3D image and video.



Sanghoon Lee (M'05–SM'12) received the B.S. in electrical engineering from Yonsei University, Seoul, Korea, in 1989, and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 1991. From 1991 to 1996, he was with Korea Telecom, Seongnam, Korea. He received the Ph.D. degree in electrical engineering from the University of Texas at Austin, Austin, TX, USA, in 2000. From 1999 to 2002, he was with Lucent Technologies Korea Ltd., Seoul, where he was involved in 3G wireless and multimedia networks. In 2003, he joined the Department of Electrical and Electronics Engineering, Yonsei University, Seoul, Korea, as a Faculty Member, where he is currently a Full Professor. He was an Associate Editor of the *IEEE TRANSACTION ON IMAGE PROCESSING* (2010–2014). He has been an Associate Editor of the *IEEE SIGNAL PROCESSING LETTERS* since 2014, an Editor of the *Journal of Communications and Networks* since 2009, and the Chair of the IEEE P3333.1 Quality Assessment Working Group since 2011. He has served as the Technical Committee of the IEEE IVMSP 2014, the Technical Program Co-Chairs of International Conference on Information Networking in 2014, Global 3D Forum 2012, 2013, the General Chair of the 2013 IEEE IVMSP Workshop, and the Guest Editor of the *IEEE TRANSACTION ON IMAGE PROCESSING* in 2013. He has received a 2012 Special Service Award from the IEEE Broadcast Technology Society and the 2013 Special Service Award from the IEEE Signal Processing Society. His research interests include image/video quality assessments, medical image processing, cloud computing, wireless multimedia communications, and wireless networks.



Alan Conrad Bovik (S'80–M'81–SM'89–F'96) is currently the Curry/Cullen Trust Endowed Chair Professor with the University of Texas at Austin, Austin, TX, USA, where he is the Director of the Laboratory for Image and Video Engineering. He is a Faculty Member in the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has authored over 650 technical articles in these areas and holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including the Best Paper Award (2009), the Education Award (2007), the Technical Achievement Award (2005), and the Meritorious Service Award (1998). He was a recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, 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 Urbana–Champaign (2008), the IEEE Third Millennium Medal (2000), and two journal paper awards from the International Pattern Recognition Society (1988 and 1993). He is a Fellow of the Optical Society of America, the Society of Photo-Optical and Instrumentation Engineers, and the American Institute of Medical and Biomedical Engineering. He has been involved in numerous professional society activities. He has been on the Board of Governors of the IEEE Signal Processing Society from 1996 to 1998, the Co-Founder and Editor-in-Chief of the *IEEE TRANSACTIONS ON IMAGE PROCESSING* from 1996 to 2002, on the Editorial Board of the *Proceedings of the IEEE* from 1998 to 2004, a Series Editor of the *Image, Video, and Multimedia Processing* (Morgan and Claypool) since 2003, and the Founding General Chairman of the First IEEE International Conference on Image Processing, held in Austin, TX, in 1994.

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