3D Visual Activity Assessment Based on Natural Scene Statistics

Kwanghyun Lee, Anush Krishna Moorthy, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*— One of the most challenging ongoing issues in the field of 3D visual research is how to perceptually quantify object and surface visualizations that are displayed within a virtual 3D space between a human eye and 3D display. To seek an effective method of quantification, it is necessary to measure various elements related to the perception of 3D objects at different depths. We propose a new framework for quantifying 3D visual information that we call 3D visual activity (3DVA), which utilizes natural scene statistics measured over 3D visual coordinates. We account for important aspects of 3D perception by carrying out a 3D coordinate transform reflecting the nonuniform sampling resolution of the eye and the process of stereoscopic fusion. The 3DVA utilizes the empirical distortions of wavelet coefficients to a parametric generalized Gaussian probability distribution model and a set of 3D perceptual weights. We conducted a series of simulations that demonstrate the effectiveness of the 3DVA for quantifying the statistical dynamics of visual 3D space with respect to disparity, motion, texture, and color. A successful example application is also provided, whereby 3DVA is applied to the problem of predicting visual fatigue experienced when viewing 3D displays.**

***Index Terms*— 3D visual activity (3DVA), visual natural scene statistic (visual NSS), human visual system (HVS), 3D coordinate transform, stereoscopic video.**

I. Introduction

THERE has been a recent growing demand for stereoscopic/3D content across a wide range of consumer-oriented applications, including digital cinema, gaming, home theatre and mobile video devices [1], [2]. The statistics of 2D content have been intensively studied and the distributions of the essential elements of natural images including luminance, color, contrast, power spectra and motion have been extensively analyzed and modeled [3]–[6]. However, a similar understanding of the statistical properties of perceived 3D content is currently lacking. Further, while many aspects of 2D visual perception have been extensively analyzed and utilized for the quantitative analysis of 2D image content, in many cases, similar studies have not been conducted in the realm of stereoscopic 3D content. For example, while researchers have studied foveation in the 2D spatial domain [7]–[12], the contrast sensitivity function (CSF) in the frequency domain [13]–[15] and motion perception in the temporal domain [16]–[18] for 2D content, similar studies on 3D content are quite sparse [19].

There lately has been significant effort directed towards analyzing and understanding 2D image content based on natural scenes statistics (NSS) [20]–[26]. Analyzing the statistical properties of natural scenes is often viewed as dual to modeling neural responses in visual cortex, which has adapted to natural image statistics over evolutionary (and shorter) time scales [27], and these properties have proved quite useful in advancing methods of automatic image analysis [22]. This duality has been confirmed by strong observed correlations between the statistics of natural scenes and the responses of cortical neurons [28]. 3D NSS have also been recently studied with an eye towards advancing both 3D vision science and image engineering. For example, the proportion of disparity distributions in natural stereoscopic content qualitatively agrees with the distribution of disparity tuning neurons in V1 [29]–[31]. Moreover, according to a study on disparity tuned neurons in area MT [32], the coding area of MT neurons is about 4.7 degrees, which agrees well with natural disparity distributions in stereoscopic content (about 5 degrees). Deeper statistical models of 3D image data have great potential for advancing our understanding of the role of these neural architectures as well as for deepening the design of automatic, perceptually optimized 3D visual analysis and processing systems.

Towards furthering our knowledge in this direction we have developed a new framework for quantitatively analyzing 3D video content. We derive a model of visual activity over space, time and disparity called "3D visual activity (3DVA)", which extracts statistical information from stereoscopic videos. 3DVA has potential usefulness in a multitude of applications such as visual fatigue prediction, visual attention prediction, visual quality assessment etc.

Prior studies have shown that the properties of attention, comfort, fatigue and other factors influencing stereoscopic

Manuscript received February 3, 2013; revised August 21, 2013; accepted November 1, 2013. Date of publication November 12, 2013; date of current version December 12, 2013. This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education under Grant NRF-2013R1A1A2A10011764. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aleksandra Pizurica. *(Corresponding Author: S. Lee.)*

K. Lee and S. Lee are with the Center for Information Technology, Yonsei University, Seoul 120-749, Korea (e-mail: kwangsabu@yonsei.ac.kr; slee@yonsei.ac.kr).

A. K. Moorthy and A. C. Bovik are with the Department of Electrical and Computer Engineering, Laboratory for Image and Video Engineering, University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: anushmoorthy@gmail.com; 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.2013.2290592

perception are correlated with the depth and visual angle subtended by the objects in the scene, as well as by the statistical properties of the scene in space, time, frequency and color-space [33]– [40]. While these properties have been studied in isolation, they have not yet been studied jointly. Here we analyze 3D visual activity using a combination of statistical features extracted from stereoscopic scenes, weighted using perceptual characteristics related to foveation and fusion.

3DVA analyzes a stereoscopic scene by evaluating the structural geometry of the 3D space after aligning objects in accordance with their depths between the human eye and the display. Once objects are aligned in the 3D volume, statistical distributions of the objects are measured with respect to object disparity over the spatial and temporal domains in 3D space. In addition, statistics related to texture and color are also measured over the spatial domain. A visual weight is then defined using models of foveation and stereoscopic fusion.

Foveation refers to the non-uniform sensitivity of the human eye in space, where the resolution decreases as one moves away from the fovea[1]. Fusion refers to the process by which the human visual system (HVS) combines the left and right views of the scene from the two eyes to create a single "fused" stereoscopic image. The region perceived with the highest resolution on the 2D plane along the X and Y axes is called the foveated area, and that on the 3D depth along the Z axis is called *Panum's fusional area*. As one moves away from the foveated area, a sparser distribution of neurons leads to lowered resolution, and hence to perceived 'blurriness'. As one moves away from the Panum's fusional area along the Z-axis, the brain is unable to fuse the two views and this results in the phenomenon of double vision (diplopia).

In this article, we utilize models of foveation and fusion to weight the local statistics of stereoscopic content to better match visual perception. Specifically, a 3D coordinate transform based on visual resolution expressed in terms of foveation and fusion is performed. We have previously demonstrated that a nonuniformly sampled foveated image mapped onto the retina can be analyzed in the uniform domain via a resolution change over virtual curvilinear coordinates (a coordinate transformation) [7]. In the same fashion, we deal with the nonuniform 3D images projected onto the retinas of the two eyes by sequentially mapping it over virtual curvilinear coordinates to a uniform version on the XY plane to account for foveation then on the Z axis to account for fusion. Computed statistical features of color, motion and texture are then weighted according to the computed foveal, fused 3D 'percept.' Combining these statistics produces the proposed 3D visual activity measure. In order to demonstrate the usefulness of 3DVA, we use it to predict the degree of visual fatigue felt when humans view a stereoscopic video.

II. 3D Visual Activity

This section describes the steps involved in the extraction of pertinent 3D visual activity in formation. The algorithm broadly proceeds as follows. First, salient regions in the image are detected and saliency maps are constructed. Once such maps are obtained, a 3D coordinate transform is performed to account for foveation and fusion. A 3D weighting function is then obtained from the computed coordinate transformation based on known perceptual mechanisms. The 3D video is then analyzed and statistical features related to disparity, motion, texture and color are extracted. The obtained feature maps are evaluated in the wavelet domain, and the activities (labelled $\boldsymbol{M}_{\mathrm{D}}$, $\boldsymbol{M}_{\mathrm{M}}$, $\boldsymbol{M}_{\mathrm{T}}$ and $\boldsymbol{M}_{\mathrm{C}}$ respectively) are then perceptually weighted. Combining the features produces the 3D visual activity measure. The overall process is outlined in Fig. 1.

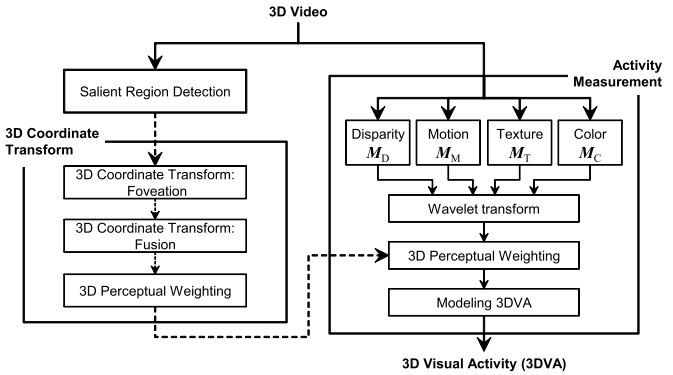


Fig. 1. Block diagram for measuring the 3DVA utilizing the NSS for 3D video.

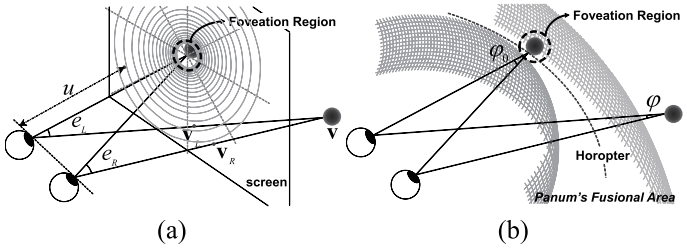


Fig. 2. 3D perceptual properties: foveation and fusion. (a) Geometric representation of objects projected onto a foveated coordinate system. (b) Geometric representation of objects relative to Panum's fusional area.

A. Saliency and 3D Coordinate Transformation

Figure 2 illustrates the two properties of stereoscopic perception that we model – foveation and fusion. In this section we detail the steps involved in extracting salient regions from a stereoscopic image and the foveation and fusion processes that are modeled.

1) Saliency: Visual attention and 3D saliency were studied in [42] and [43]. However, it is generally quite difficult to automatically identify visually salient regions in natural 3D images and videos in a manner that agrees with visual attention or gaze patterns. The authors of [42] proposed a strategy to reduce fixation prediction errors, whereby a scene to be analyzed is first classified, then salient regions are predicted adaptively. This approach to video scene classification also relies on analyzing camera and object motion.

The authors of [42] observed that most subjects directed their attention to foreground objects having large crossed

[1]The fovea centralis is the part of the eye that is located in the center of the retinal macula. The fovea is responsible for sharp central vision, which is necessary for reading, watching television or movies, driving, and any activity where visual detail is of primary importance [41].

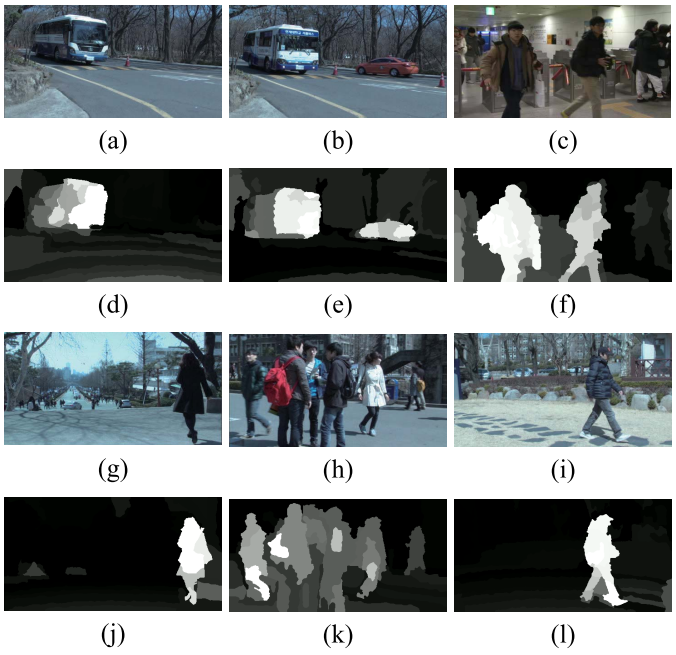


Fig. 3. 3D test sequences and computed salient regions. (a) "Car1" (135^{th} frame). (b) "Car2" (128^{th} frame). (c) "Metro1" (211^{th} frame). (d) Salient regions of "Car1". (e) Salient regions of "Car2". (f) Salient regions of "Metro1". (g) "University1" (78^{th} frame). (h) "University2" (134^{th} frame) (i) "Walking-person8" (155^{th} frame). (j) Salient regions of "University1". (k) Salient regions of "University2". (l) Salient regions of "Walking-person8".

disparities when viewing stereoscopic videos. Moreover, they tended to fixate on moving objects. Further, most subjects also fixated on tracked objects having near zero motion. Therefore, spatial, temporal and disparity data were used to define predictors of visual attention. In their model 3D spatial saliency reflects attributes of luminance, size, density and the presence of depth discontinuities. Temporal and disparity saliencies are predicted using measurements of object motion speed and angular disparity, respectively. We deploy this 3D saliency model in 3DVA to detect candidate salient regions in stereoscopic videos being analyzed.

Figure 3 plots example outputs from the saliency detection algorithm in [42]. These are frames from videos described in [44]. Figures 3 (a)-(c) and (g)-(i) plot the original frames and Figs. 3 (d)-(f) and (j)-(l) plot the predicted salient regions on these frames. In Figs. 3 (a), (b) and (c) the camera is fixed, and moving vehicles and people are identified as salient regions (Figs. 3 (d), (e) and (f)). Figures 3 (g) and (h) illustrate cases where both camera and object motion occur. The most salient regions detected were those where the objects had a motion trajectory opposite to that of the camera motion trajectory (Figs. 3 (j) and (k)). In the case where the camera pans, with the object stationary (Fig. 3 (i)), the object (here, a walking person) is detected as the most salient region (Fig. 3 (l)).

2) Foveation: Visible light from the natural world passes through the optics of the eyes onto photoreceptors which transduce it into neural responses. The distribution of the photoreceptors in the eye is not uniform, and decreases away from the center of the fovea. Since visual acuity is a function of local photoreceptor density, the part of the image that is sampled at the fovea has the highest resolution and hence the highest sensitivity to detail. In the following, we assume that the region with the highest saliency (obtained as described above) falls on the fovea and hence has the highest sensitivity/resolution.

There exist several models that describe 2D visual acuity as a function of the spatial location of the stimulus on the fovea [7], [8], [45]–[47]. In the case of 3D stimuli, the expression for 2D foveation needs to be modified in order to account for the stereoscopic viewing condition.

As depicted in Fig. 2 (a), an object $\mathbf{x}$ on the viewing screen is generally projected onto different regions of the retinas of the two eyes. When a viewer fixates on an object on the screen, so that it falls within the foveation regions, then the distances between the positions of the two eyes causes the foveation regions to differ from each other. Therefore, it is necessary to capture this difference and use it to correctly model the nonuniform resolution in 3D.

Suppose a viewer fixates on a 'foveation region' in the XY plane as shown in Fig. 2 (a). For a given object $\mathbf{x} = (x_1, x_2, x_3)$ (voxels), the local foveated image bandwidth can be calculated as follows. Let $\mathbf{x}_L$ and $\mathbf{x}_R$ be the crossing points on the XY plane from the left and right eyes to $\mathbf{x}$, respectively. The eccentricity $e_L(u, \mathbf{x}_L)$ (or $e_R(u, \mathbf{x}_R)$) can be then found from the distance between the fixation point and the object, and the viewing distance u from the eyes to the XY plane.

Let w represent the local cut-off frequency of the eye. As shown in [13] and [14], w for the left eye (similar for the right eye) is given by

$$w\left(e_L\left(u, \mathbf{x}_L\right)\right) = \frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\alpha\left(e_L\left(u, \mathbf{x}_L\right) + e_2\right)}, \tag{1}$$

where CT_0 is a minimum contrast threshold, e_2 is the half-resolution eccentricity constant and α is a spatial frequency decay constant. The fitting parameters that were found by fitting to experimental data [47] are $\alpha = 0.106$, $e_2 = 2.3$ and $CT_0 = 1/64$. The frequency w in (1) for the two eyes can be calculated by averaging the two cutoff frequencies:

$$w\left(e_L\left(u, \mathbf{x}_L\right), e_R\left(u, \mathbf{x}_R\right)\right) = \frac{2{e_2}^2 \ln\left(\frac{1}{CT_0}\right) + e_2\left(e_L\left(u, \mathbf{x}_L\right) + e_R\left(u, \mathbf{x}_R\right)\right)}{2\alpha\left(e_L\left(u, \mathbf{x}_L\right) + e_2\right)\left(e_R\left(u, \mathbf{x}_R\right) + e_2\right)}. \tag{2}$$

Figure 4 illustrates stereoscopic foveation. Figures 4 (a) and (b) are the left view of the "Cones" stereoscopic image from [48] and the ground truth depth map, respectively. Figure 4 (e) plots a 3D reconstruction using the computed depth information from Fig. 4 (b). Under the assumption that the salient region (and hence the foveation region) of the "Cones" image is the face mask in Fig. 4 (c), the observer perceives the original image in (a) as the foveated image in (c) with a resolution that varies with spatial location.

3) Fusion: The human visual system perceives 3D partly by fusing the two views from the two eyes using computed local disparity. Although the foveated region may have a high spatial resolution, the region where the two eyes are focused in 3D will have the highest *perceived* resolution, which drops off as a function of the distance from the point of focus along

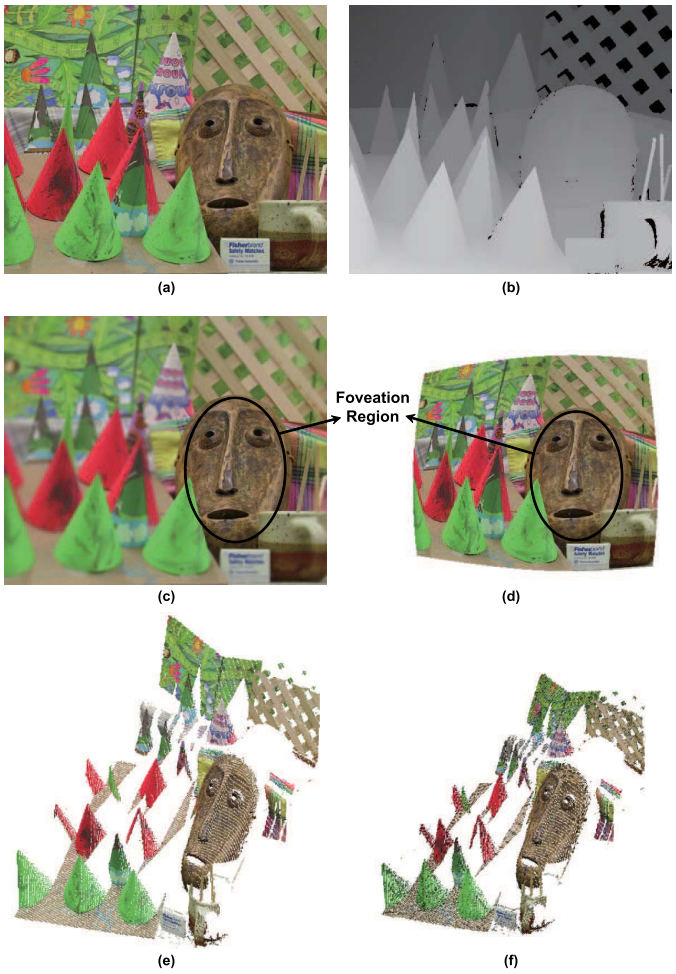


Fig. 4. Original and foveated "Cones" images and spaces. (a) Original image. (b) Ground truth. (c) Foveated image. (d) Foveated coordinate transformation in 2D. (e) Original space. (f) Foveated coordinate transformation in 3D.

the Z-axis. What this implies is that a non-uniform mapping of the Z-axis similar to that performed for the spatial XY plane is necessary to recreate the perceived 3D stimulus.

The region in which the points in the 3D volume have the same angular disparity as the point of fixation is termed the *horopter* (Fig. 2 (b)). The visual resolution is highest along the horopter and decreases with changes in angular disparity. In general, even if an object strays slightly out of the horopter, an observer can fuse it clearly. The area over which a human can easily fuse the image is called *Panum's fusional area*. Objects in Panum's fusional area have the same perceived resolution as those that lie along the horopter. Panum's fusional area extends approximately ±600 arc second (10 arc minutes) on either side of the horopter. It does not have a fixed size, but varies depending on the stimulus conditions [49]. Objects that lie outside Panum's fusional area result in the perception of double images, where the left and right views overlap due to the mismatch in convergence between the two eyes. This phenomenon is termed diplopia. In regions where diplopia occurs, an observer cannot fuse objects completely and hence objects in these regions may be regarded as having significantly lower visual importance than those in Panum's fusional area.

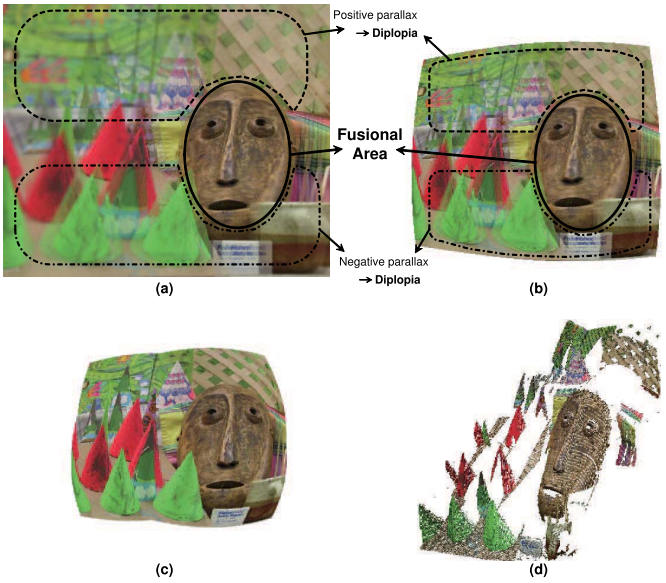


Fig. 5. Fused "Cones" images and space. (a) Fused diplopic image from Fig. 4 (c). (b) Fused diplopic image from on Fig. 4 (d). (c) Fused image following 2D coordinate transformation. (d) Fused 3D percept following 3D coordinate transformation.

Panum's fusional area is frequently referred to in relevant studies, where a model is introduced to define the boundaries of binocular fusion by investigating the greatest amount of horizontal disparity. Similar to foveation, we can model nonuniform resolution along the depth axis. Here, we employ the model from [50]:

$$w'(\Delta\varphi) = \begin{cases} k', & 0 \le \Delta\varphi \le \delta \\ k' \cdot \exp\left(-\frac{\Delta\varphi - \delta}{\epsilon}\right), & \delta < \Delta\varphi \end{cases}, \qquad (3)$$
$$\Delta\varphi = |\varphi - \varphi_0|$$

where φ_0 and φ are the angles of convergence at fixation and at another neighboring region, respectively (Fig. 2 (b)). δ is the threshold that decides the width of the fusional area ($\delta = 0°$ in general), ϵ is a fixed coefficient which has been determined from physiological experiments to be approximately equal to 0.62° [50]. k' is a scaling parameter used for "fusion filtering" as described in Section II-B.1.

Figure 5 (a), which corresponds to Fig. 4 (c) illustrates the experience of diplopia. The left and right images of objects at different depths than at fixation are overlapped.

B. 3D Coordinate Transformation

In this section, we describe the *3D Coordinate Transform* module of Fig. 1 that is applied post- foveation and fusion filtering in our model of stereoscopic perception. Figure 6 illustrates the process. Suppose that there exist coordinate transformations $\mathbf{v} = (v_1, v_2, v_3)^T$ and $\mathbf{v}' = (v_1', v_2', v_3')^T$ for $\mathbf{x} = (x_1, x_2, x_3)^T$ where the subscript T denotes transpose. If a one-to-one correspondence exists among $\mathbf{x}$, $\mathbf{v}$ and $\mathbf{v}'$, where v_1, v_2, v_3, v_1', v_2' and v_3' are continuous and uniquely invertible, then $\mathbf{v}$ and $\mathbf{v}'$ are called 3D curvilinear coordinates.

We use the following notation (Fig. 6). $\mathbf{x}$ represents Cartesian coordinates and $\mathbf{v}$ and $\mathbf{v}'$ are curvilinear coordinates based

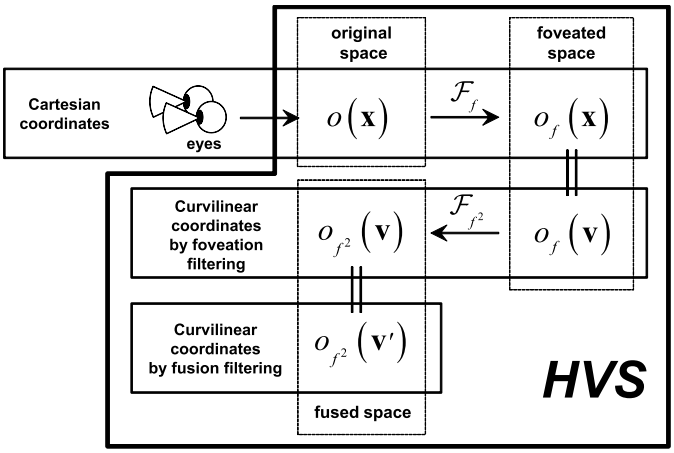


Fig. 6. 3D coordinate transformation for foveation and fusion.

on foveation and fusion filtering, respectively. The 3D space between the observer and the display is called the original space, $o(\mathbf{x})$, and Fig. 4 (e) is an example of $o(\mathbf{x})$. Because of foveation, a human perceives Fig. 4 (a) similar to Fig. 4 (c). In other words, foveation filtering transforms $o(\mathbf{x})$ to the foveated space $o_f(\mathbf{x})$ (Fig. 4 (a) $\rightarrow$ (c)). Then, $o_f(\mathbf{v})$ is the space of $o_f(\mathbf{x})$ in curvilinear coordinates generated by foveation filtering, as shown in Fig. 4 (d) (in 2D) and Fig. 4 (f) (in 3D), respectively. Figure 4 (d) represents the 2D visual information that the visual system receives from the stimulus in Fig. 4 (a). Figure 4 (e) is transformed to Fig. 4 (f) using a 3D coordinate transformation along with foveated weighting. The image retains its original resolution in the vicinity of the facial mask; however, away from fixation there is a loss of 3D resolution. The relationship between the original and foveated spaces is given by $o_f(\mathbf{x}) = \mathcal{F}_f(o(\mathbf{x}))$ and $o_f(\mathbf{x}) = o_f(\mathbf{v})$ where $\mathcal{F}_f$ denotes the process of foveation filtering.

Figure 5 (b), which corresponds to Fig. 4 (d) illustrates diplopia. Fusion filtering transforms $o_f(\mathbf{v})$ to the fused space $o_{f^2}(\mathbf{v})$ (Fig. 4 (d) $\rightarrow$ Fig. 5 (b)). Then, $o_{f^2}(\mathbf{v}')$ becomes $o_{f^2}(\mathbf{v})$ over the curvilinear coordinates as a result of fusion filtering, as shown in Fig. 5 (c) (in 2D) and Fig. 5 (d) (in 3D), respectively. Objects at different depths than at fixation are aligned to the reference disparity using a procedure similar to that for the curvilinear coordinate transform used in the foveation model. Figure 5 (c) plots an example. It is worth noting that the sizes of objects placed at negative and positive parallax are reduced due to depth compensation. Further, the double vision effect apparent in Fig. 4 (a) disappears after the coordinate transformation in Fig. 5 (c). Figure 5 (d) plots Fig. 5 (c) onto a 3D volume, where the object sizes outside the foveation region shrink as a function of their visual importance. The relationship between these spaces is given by $o_{f^2}(\mathbf{v}) = \mathcal{F}_{f^2}(o_f(\mathbf{v}))$ and $o_{f^2}(\mathbf{v}) = o_{f^2}(\mathbf{v}')$ where $\mathcal{F}_{f^2}$ denotes fusion filtering.

To model the perceptual process more precisely, we transform the 3D coordinate twice: (1) The transformation of $o_f(\mathbf{x})$ to $o_f(\mathbf{v})$; (2) The transformation of $o_{f^2}(\mathbf{v})$ to $o_{f^2}(\mathbf{v}')$.

1) Frequency Domain Analysis of Coordinate Transform: Let $\Omega = (\Omega_1, \Omega_2, \Omega_3)^T$ be continuous 3D frequencies. For $\mathbf{x}$, $\Omega \in \mathcal{R}^3$, let $b(\mathbf{x})$ and $\mathcal{B}(\Omega)$ be a 3D signal and its Fourier transform, respectively. When $\mathcal{B}(\Omega)$ is band-limited within a circle of radius Ω_o, $\mathcal{B}(\Omega) = 0$ for $|\Omega| \geq \Omega_o$. Then $b(\mathbf{x})$ is an Ω_o-band-limited signal, i.e., $b(\mathbf{x}) \in B^{\Omega_o}$, where B^{Ω_o} is the space of Ω_o-band-limited signals. Through the operation of $\mathcal{F}_f$ and $\mathcal{F}_{f^2}$, $o_f(\mathbf{x}) \in B^{\Omega(\mathbf{x})}$ and $o_{f^2}(\mathbf{v}) \in B^{\Omega(\mathbf{v})}$ become *locally band-limited signals* with respect to the coordinate systems $\mathbf{v}$ and $\mathbf{v}'$, where $B^{\Omega(\mathbf{x})}$ and $B^{\Omega(\mathbf{v})}$ are the space of locally band-limited signals.

Due to foveation filtering, the original space $o(\mathbf{x})$ in Fig. 4 (a) can be transformed into a locally band-limited signal $o_f(\mathbf{x}) \in B^{\Omega(\mathbf{x})}$ as shown in Fig. 4 (c). The region is transformed from the original space as a function of the local bandwidth. Thus, the region centered at the foveation point expands more than does the periphery. Then, the foveated image over the new coordinates is given by

$$o_f(\mathbf{x}) = \mathcal{F}_f(o(\mathbf{x}), \Omega(\mathbf{x})), \tag{4}$$

where $\Omega(\mathbf{x}) = w(\mathbf{x})$. As shown in Fig. 4 (c), the local bandwidth corresponds to nonuniform sampling of $o_f(\mathbf{x})$. However, the local bandwidth corresponds to uniform sampling of $o_f(\mathbf{v})$ as shown in Fig. 4 (d) (in 2D) and (f) (in 3D).

The procedure of fusion filtering is similar to that of foveation filtering. The foveated space $o(\mathbf{v})$ in Fig. 4 (d) is transformed into a locally band-limited signal $o_{f^2}(\mathbf{v}) \in B^{\Omega(\mathbf{v})}$ as shown in Fig. 5 (b). The fused space is given by

$$o_{f^2}(\mathbf{v}) = \mathcal{F}_{f^2}(o(\mathbf{v}), \Omega(\mathbf{v})), \tag{5}$$

where $\Omega(\mathbf{v}) = w'(\mathbf{v})$. As shown in Fig. 5 (b), the local bandwidth corresponds to nonuniform sampling of $o_{f^2}(\mathbf{v})$, but to uniform sampling of $o_{f^2}(\mathbf{v}')$ as shown in Fig. 5 (c) (in 2D) and (d) (in 3D).

Therefore, the final 3D perceptual weighting is

$$\mathbf{f} = w'(w(\mathbf{x})). \tag{6}$$

C. 3D Visual Activity (3DVA)

This section details the extraction of measures of disparity activity $\mathbf{M}_\mathrm{D}$, motion activity $\mathbf{M}_\mathrm{M}$, texture activity $\mathbf{M}_\mathrm{T}$, and color activity $\mathbf{M}_\mathrm{C}$. Once these features maps are extracted, a wavelet transform of the feature maps is computed and the 3D visual weights, whose computation was detailed in the previous section, are applied to the computed maps in the wavelet domain. A statistical feature extraction process follows, resulting in the final measure of 3D visual activity (3DVA).

1) Activity Measures:

a) Disparity: Research on the accommodation and vergence feedback system of the HVS has shown that the HVS is capable of stereoscopically fusing all of the clearly perceived regions in a scene [33]. When it is difficult to fuse these regions, humans experience visual fatigue [34]. However, while clear 3D content is perceived in the Panum's fusional area, no fusion occurs outside of it. Furthermore large variations in disparity result in increased neural metabolic rates, which can lead to visual fatigue. The features used to define 3DVA are sensitive to these 3D attributes and hence 3DVA can be adapted to capture this phenomenon.

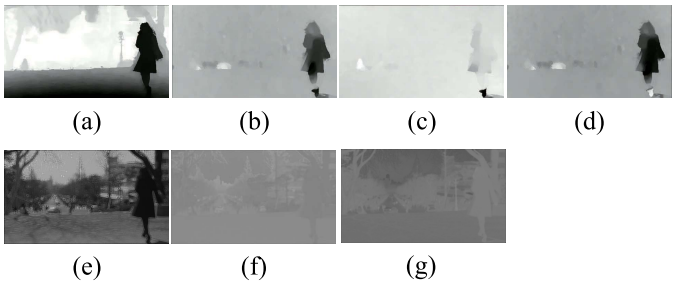


Fig. 7. Activity maps of "University1" at the 78^{th} frame. (a) Disparity map by (7). (b) X-motion map by (9). (c) Y-motion map by (9). (d) Z-motion map by (9). (e) Texture map by (10). (f) Cb-color map by (11). (g) Cr-color map by (11).

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

$$\mathbf{M}_{\mathrm{D}}(t) = \{M_{\mathrm{D}}(x_1, x_2, t) \,|\, 1 \le x_1 \le o_w, 1 \le x_2 \le o_h\} \quad (7)$$

where

$$M_{\mathrm{D}}(x_1, x_2, t) = \arg\min_{d} \left| o_{\mathrm{L}}(x_1, x_2, t) - o_{\mathrm{R}}(x_1 + d, x_2, t) \right|. \quad (8)$$

Here $\mathbf{M}_{\mathrm{D}}(t)$ and $M_{\mathrm{D}}(x_1, x_2, t)$ denote $\mathbf{M}_{\mathrm{D}}$ at time t and its pixel value at (x_1, x_2), respectively. $o_{\mathrm{L}}(x_1, x_2, t)$ and $o_{\mathrm{R}}(x_1, x_2, t)$ denote pixel values at (x_1, x_2) at time t on the original left and right image, respectively; and o_w and o_h denote the width and height, of the original left (or right) image. In (8), d is the disparity at pixel coordinate (x_1, x_2), which can be estimated using a suitable stereo matching algorithm. Here the disparity map was obtained from left and right frame using the depth estimation reference software (DERS)[2]. Figure 7 (a) is an example disparity map computed from the 78^{th} frame of the "University1" sequence.

b) Motion: The HVS is hypothesized to be self-calibrating in its mapping of environmental stimuli onto patterns of neural activity [35]. If vergence fluctuates with variations in depth motion, the eyes may struggle to find a new stable state, leading to visual fatigue. In other words, motion along the depth axis (or the Z-axis) may cause visual fatigue and discomfort [35], [36]. Large motions lead to an increase in the response of 3DVA.

Let $\mathbf{M}_{\mathrm{M}}$ denote the motion map of the voxels on the X, Y and Z axes. For each direction X, Y or Z, the mean value of the square of the position changes over a period of time is

$$\mathbf{M}_{\mathrm{M}}(t) = \sum_{\tau=t-T_{\mathrm{M}}+1}^{t} \left(\frac{\partial o(\mathbf{x}(\tau))}{\partial t} \right)^2 \Big/ T_{\mathrm{M}}, \quad (9)$$

where $\mathbf{M}_{\mathrm{M}}(t)$ and T_{M} denote $\mathbf{M}_{\mathrm{M}}$ as a function of time t and the time interval over which the motion map is extracted, respectively. $\mathbf{M}_{\mathrm{M}}(t)$ is the square of the partial derivative of 3D position (motion) in the original space between times $t - T_{\mathrm{M}} + 1$ and t. Figures 7 (b)-(d) are examples of the X, Y and Z motion maps, for the 78^{th} frame of the "University1" sequence.

c) Texture: The unpleasantness experienced when viewing some patterns can be characterized by their spatial frequency attributes [37]. Unpleasant patterns may even cause an excess of neural excitation, thereby producing anomalous visual effects or in the rare extreme case, clinical seizures. A large concentration of high frequency components can lead to a decrease in visual sensitivity and an increase in visual fatigue and discomfort [37], [38]. 3DVA captures the proportion of spectral energy distributed at high frequency bands and is measured locally in the wavelet domain.

Gabor filters have been widely used as an effective tool to fulfill the feature extraction tasks in many biometric and image processing systems. The frequency and orientation responses of Gabor filters are similar to those of human cortical neurons, and they have been found to be particularly appropriate for achieving perceptually efficient texture representation [53], [54]. Therefore, we use responses of a bank of Gabor filters to construct a texture map, $\mathbf{M}_{\mathrm{T}}$, as a function of time t as

$$\mathbf{M}_{\mathrm{T}}(t) = \mathrm{Gabor}\Big(o_{\mathrm{L}}(\mathbf{x}(t))\Big) \quad (10)$$

where $\mathrm{Gabor}(o_{\mathrm{L}}(\cdot))$ are the responses of Gabor filter bank applied to the left image. The design of the Gabor filter bank is based on [55]. Figure 7 (e) is an example of a computed texture map, on the 78^{th} frame of the "University1" sequence.

d) Color: There has been observed a consistent positive correlation between ratings of visual discomfort and perceptual differences in color [39], [40]. The largest chromatic separations produced the largest haemodynamic responses and the greatest degrees of visual discomfort. The experience of visual discomfort is homeostatic, signifying a large metabolic demand towards reducing the sustained metabolic load on the visual neurons. Again, 3DVA embodies features sensitive to this chromatic behavior.

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

$$\mathbf{M}_{\mathrm{C}}(t) = \mathcal{F}_{\mathrm{M}}\Big(\mathcal{F}_{\mathrm{L}}\Big(o_{\mathrm{L}}(\mathbf{x}(t))\,\big|_{\mathrm{C}}\Big)\Big), \quad \mathrm{C} \in \{\mathrm{Cb}, \mathrm{Cr}\} \quad (11)$$

where $\mathbf{M}_{\mathrm{C}}(t)$ denotes $\mathbf{M}_{\mathrm{C}}$ as a function of time t, and $\mathcal{F}_{\mathrm{M}}$ and $\mathcal{F}_{\mathrm{L}}$ denote median filtering and low pass filtering, and $o_{\mathrm{L}}(\cdot)|_{\mathrm{C}}$ is the color component (Cb and Cr) of the left image. Figures 7 (f)-(g) demonstrate examples of the Cb and Cr color maps, for the 78^{th} frame of the "University1" sequence.

2) Wavelet Transform and 3D HVS Weighting: Once the maps above are computed, a wavelet transform is applied to each of them. We use the steerable pyramid [56] over 3 scales and 3 orientations.

The next step is divisive normalization, which accounts for the non-linear adaptive gain control process over certain populations of cortical neurons [57], [58]. Such normalization could also reduce statistical dependencies between subbands thereby decoupling subband responses to a certain degree [59], [60].

[2]The DERS is the reference software for depth estimation released by the ISO-MPEG 3DV group [51]. The original DERS has been adapted to fit the requirements of the multi-view context. Therefore, we use the modified version of the DERS to fit the two views by using a hole filling algorithm [52].

Here, divisive normalization is implemented as described in [59]. Once each map is normalized, the maps are weighted and then statistically analyzed.

Let $\mathbf{p}_{k,i}(t)$ be the probability mass function (PMF) of the i^{th} subband of factor k as a function of time t. Factor k is one of the maps disparity, motion, texture or color. Let $\mathbf{B}$ be the set of bins of the histogram w.r.t. the wavelet coefficients and $p_{k,i}(j,t)$ be the PMF of the j^{th} bin. Then,

$$\mathbf{p}_{k,i}(t) = \left\{ p_{k,i}(j,t) \,|\forall j \in \mathbf{B} \right\}. \tag{12}$$

If $\mathbf{W}_i(\cdot)$ denotes the wavelet coefficient matrix of the i^{th} subband, $\mathbf{M}_k$ is the map of factor k (k = D, M, T and C), and $\bar{B}$ is the interval between bins, then the values of the wavelet coefficients $\mathbf{W}_i(\cdot)$ belonging to the j^{th} bin fall in the range $j - \bar{B}/2 \leq \mathbf{W}_i(\mathbf{M}_k(t))(w,h) \leq j + \bar{B}/2$. Then

$$p_{k,i}(j,t) = \frac{\sum_{h=1}^{i_h}\sum_{w=1}^{i_w} \Lambda_j(w,h,t)}{\sum_{h'=1}^{i_h}\sum_{w'=1}^{i_w} \left(f_{(i_w,i_h)}(w',h',t)\right)^2}, \tag{13}$$

where

$$\Lambda_j(w,h,t) = \begin{cases} \left(f_{(i_w,i_h)}(w,h,t)\right)^2, \\ \quad \text{if } j - \bar{B}/2 \leq \mathbf{W}_i(\mathbf{M}_k(t))(w,h) \leq j + \bar{B}/2 \\ 0, \\ \quad \text{otherwise} \end{cases} \tag{14}$$

and where i_w and i_h are the width and height of the i^{th} subband, $\mathbf{f}_{(i_w,i_h)}(t)$ is the visual weight $\mathbf{f}$ applied to the i^{th} subband and $f_{(i_w,i_h)}(w,h,t)$ is the $(w,h)^{th}$ element of $\mathbf{f}_{(i_w,i_h)}(t)$.

The weights $f_{(i_w,i_h)}(w,h,t)$ in salient regions take higher values than those in non-salient regions. Thus, statistics computed from salient regions are given greater weight in the curvilinear coordinate system.

Applying (12)-(14) results in a situation where the number of binned samples in the salient region becomes relatively large. We have observed that the distribution of these coefficients can be well-modeled as following a generalized Gaussian distribution (GGD). In the Appendix, we provide experimental validation for this choice.

3) Generalized Gaussian Fits for Wavelet Data: The computed empirical distributions are modeled as GGD. The GGD has been used in numerous studies of image wavelet coefficients [61], [62]. We use the reliable method of [63] to extract the parameters of the GGD. The GGD is defined as

$$\text{GGD}\left(x : \mu, \sigma^2, \gamma\right) = a e^{-((x-\mu)/s)^\gamma}, \tag{15}$$

where μ, σ^2 and γ are the mean, variance, and shape parameter of the distribution, respectively. The positive constants a and s are given by

$$a = s\gamma / 2\Gamma(1/\gamma) \tag{16}$$

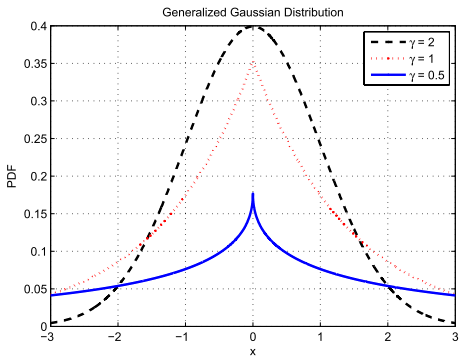


Fig. 8. Example generalized Gaussian distributions (GGD).

and

$$s = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}. \tag{17}$$

Figure 8 plots various GGDs with varying γ. Small values of γ result in peakier distributions with heavier tails, and larger values result in more uniform distributions. The value $\gamma = 2$ results in a Gaussian distribution, while $\gamma = 1$ results in a Laplacian distribution. Thus, the shape parameter is indicative of the distribution of the energy in the computed map in the wavelet-frequency domain, and this parameter is used to characterize the 3D visual activity.

We denote the GGD fit to $\mathbf{p}_{k,i}(t)$ as

$$\overline{\text{GGD}}\left(\mathbf{p}_{k,i}(t) : \mu_{k,i}(t), \sigma_{k,i}(t)^2, \gamma_{k,i}(t)\right), \tag{18}$$

where $\mu_{k,i}(t)$, $\sigma_{k,i}(t)^2$ and $\gamma_{k,i}(t)$ are the estimated mean, variance and shape parameter of $\mathbf{p}_{k,i}(t)$.

Given these parameters, 3DVA is computed as follows: First, the "visual activity" occurring within the i^{th} subband of factor k at time t is obtained using double sigmoid normalization [64], [65]:

$$\mathcal{A}_{k,i}(t) = \begin{cases} 1 - \dfrac{1}{1+\exp\left(-2\left(\left(\bar{\gamma}_{k,i}(t) - c_k\right)/c_{k,D}\right)\right)}, \\ \qquad \text{if } \bar{\gamma}_{k,i}(t) < c_k \\ 1 - \dfrac{1}{1+\exp\left(-2\left(\left(\bar{\gamma}_{k,i}(t) - c_k\right)/c_{k,U}\right)\right)}, \\ \qquad \text{otherwise} \end{cases} \tag{19}$$

where

$$\bar{\gamma}_{k,i}(t) = \begin{cases} \gamma_{k,i}(t) \cdot \bar{\delta}(t), & k = \text{D} \\ \gamma_{k,i}(t), & k \in \{\text{M,T,C}\}. \end{cases} \tag{20}$$

In (19), c_k is the reference operating point for each factor k, and $c_{k,D}$ and $c_{k,U}$ are the trailing and leading edges of the region over which (19) is approximately linear, respectively, as experimentally determined as explained in Section III-B. $\gamma_{k,i}(t)$ is the shape parameter of the i^{th} subband of factor k at time t. $\bar{\delta}(t)$ is a value that depends on the average disparity at time t, as explained next.

In order to account for conflicts between convergence and the accommodation, the disparity computed is expressed

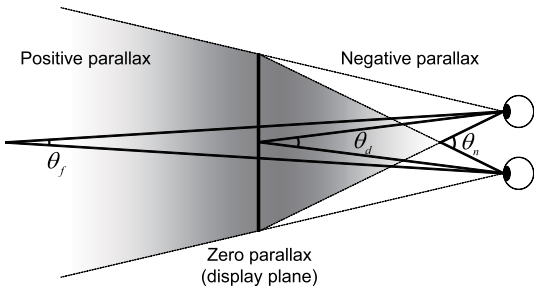


Fig. 9. Angles between the eyes and the farthest point, the disparity plane and the nearest point, respectively. The gradation in gray-scale indicates the available region to form images of 3D objects. The darker the region, the greater the presumed visual comfort.

as angular disparity. For small absolute angular disparities, $\bar{\delta}(t)$ tends towards 1 (at 0 angular disparity, we define $\bar{\delta}(t) = 1$). For larger absolute angular disparities, $\bar{\delta}(t)$ tends towards 0.

As shown in Fig. 9, if the average angular disparity $\delta(t)$ is $\theta_d - \theta_n$ (at the nearest point) or is θ_d (assume that θ_f has converged to 0), then $\bar{\delta}(t) = 0$, where θ_n, θ_d and θ_f are the angles between the eyes and the nearest point, the display plane and the farthest point, respectively. $\bar{\delta}(t)$ is given by

$$\bar{\delta}(t) = \begin{cases} (\theta_n - \theta_d + \delta(t))/(\theta_n - \theta_d), & \text{Negative parallax} \\ (\theta_d - \delta(t))/\theta_d, & \text{Positive parallax} \end{cases} \tag{21}$$

Finally, the overall (summed, or pooled) 3DVA index is

$$\mathcal{A} = \sum_t \mathcal{A}(t)/T \tag{22}$$

where

$$\mathcal{A}(t) = \sum_k \omega_k \left(\sum_i \mathcal{A}_{k,i}(t)/I \right), \quad k \in \{\text{D,M,T,C}\}. \tag{23}$$

In (22) and (23), T is the overall temporal duration of the 3D content, $\mathcal{A}(t)$ is the temporal 3DVA at time t, I is the number of subbands, and ω_k is a parameter that is used to adjust the relative importance of the four factors. For simplicity, we shall take $\omega_k = 1$ ($\forall k \in$ {D,M,T and C}).

III. Simulation Results

A. Dataset

To evaluate the performance of 3DVA, the 3D test sequences in [44] were employed. These 3D sequences were captured using Panasonic AG-3DA1, Sony HDR-TD20/S and Sony HDCP1 stereoscopic cameras, and include content containing a highly diverse distribution disparities, motions, textures and colors as well as camera motion.

Figures 10 plots exemplar frames from the 3D test sequences that were used for validation. The lengths of the sequences "Library5", "Library6", "Market1", "Metro1", "Metro2", "Metro3", "Statue1", and "Street2" are 30 seconds, and those of the sequences "Library2" and "Library7" are 60 seconds while the rests are 10 seconds. The frame-rate and the resolution of all sequences are $30fps$ and 1920×1080, respectively.

B. Fitting and Normalization

1) Fitting to GGD: We list the shape parameter γ computed by the GGD fit to the weighted wavelet responses of the activity maps of disparity, motion, texture and color for each subband in Tables I-II for the "Car1" and "University2" sequences. To conserve space, we tabulate these results for only two sequences. In these tables, γ is the mean shape parameter over the sequence duration T.

2) Normalization of 3DVA: As mentioned before, it is necessary to normalize the measured values of visual activity by the range of values that the shape parameter can take: c_k, $c_{k,U}$ and $c_{k,D}$ ($k \in$ {D,M,T,C}) in (19). We have a large testset that is composed of around 13,800 images (the total number of frames in the test set). Using [65], the normalization parameters thus obtained are: $(c_{\text{D}}, c_{\text{D},U}, c_{\text{D},D}) = (1.0068, 0.9670, 0.2094)$, $(c_{\text{M}}, c_{\text{M},U}, c_{\text{M},D}) = (1.1288, 0.4527, 0.3306)$, $(c_{\text{T}}, c_{\text{T},U}, c_{\text{T},D}) = (1.1322, 0.1338, 0.2904)$, and $(c_{\text{C}}, c_{\text{C},U}, c_{\text{C},D}) = (0.9535, 0.1697, 0.1492)$.

C. Measuring 3DVA

Figure 11 plots temporal 3DVA for the videos in Fig. 10. We show the results obtained for eight representative sequences. Figures 11 (a)-(h) plot the temporal activities of disparity and motion while those of texture and color are shown in Figs. 11 (i)-(p) for "Car1", "Car2", "Library5", "Library6", "Metro3", "Street2", "Restaurant1", and "University2". The temporal 3DVAs are shown in Figs. 11 (q)-(t).

When a stereopair exhibits a wide variety of disparities as in "Library5", "Metro3", and "University2", the average measured disparity activity was, in each case: (Library5: 0.76, Metro3: 0.76, University2: 0.92). By contrast, when there was narrow disparity distribution as in "Library6" and "Street2" ($500^{th} - 899^{th}$ frames), the average measured disparity activity was: (Library6: 0.16, Street2: 0.01).

Regarding activity in foveated regions, in the sequences "Car1", "Car2", "Library6" ($135^{th} - 899^{th}$ frames), and "Restaurant1", the measured motion activity was distinctly modified by the appearance of salient objects. When objects labelled as 'salient' approached the camera, the measured motion activity was found to remain high (Car1: 0.66, Car2: 0.71, Library6: 0.67, Restaurant1: 0.72).

When there was camera motion or random motions of people or other salient objects as in "Library5", "Metro3", "Street2" ($1^{st} - 320^{th}$ frames), and "University2", then the measured motion activity was found to fluctuate a great deal.

For the indoor and static camera scenes; "Library6" and "Restaurant1", the distances between the objects in the scene and camera is less than in the outdoor scenes. Further, when the camera is static, texture and color properties in the images are less distributed. Since the textures and colors of objects are captured with greater detail, the measured texture and color activities remain high. The average measured texture and color activities were: (Library6: 0.65 and 0.66, Restaurant1: 0.62 and 0.76).

Figures 12 and 13 plot the measured visual activities as well as the frames that correspond to the lowest and highest points

Fig. 10. Example frames of the test sequences. (a) 112^{th} frame of "Car1". (b) 129^{th} frame of "Car2". (c) 86^{th} frame of "Crosswalk2". (d) 201^{st} frame of "Library2". (e) 64^{th} frame of "Library4". (f) 117^{th} frame of "Library5". (g) 560^{th} frame of "Library6". (h) 494^{th} frame of "Library7". (i) 6^{th} frame of "Marathon1". (j) 1^{st} frame of "Market1". (k) 211^{th} frame of "Metro1". (l) 307^{th} frame of "Metro2". (m) 292^{nd} frame of "Metro3". (n) 66^{th} frame of "Restaurant1". (o) 142^{nd} frame of "Sidewalk-lateral1". (p) 317^{th} frame of "Statue1". (q) 255^{th} frame of "Street2". (r) 76^{th} frame of "University1". (s) 131^{st} frame of "University2". (t) 204^{th} frame of "Walking-person8".

TABLE I

SHAPE PARAMETER γ FOR THE "CAR1" SEQUENCE

	Scale	Horizontal	Vertical	Diagonal		Scale	Horizontal	Vertical	Diagonal
	1	0.4689	0.4815	0.3337		1	0.6309	0.6449	0.4792
Disparity	2	1.1889	1.0620	0.7425	Motion	2	1.3071	1.3115	1.1331
	3	1.5113	0.8679	0.9550		3	1.1714	1.1252	1.2530
	ave	1.0563	0.8038	0.6771		ave	1.0365	1.0272	0.9551
	Scale	Horizontal	Vertical	Diagonal		Scale	Horizontal	Vertical	Diagonal
	1	0.8635	0.8656	0.9927		1	0.5191	0.5411	0.5489
Texture	2	1.2148	1.2082	1.1287	Color	2	1.0827	1.0114	1.0634
	3	1.3679	1.1730	1.0219		3	1.0325	0.9288	1.0007
	ave	1.1488	1.0823	1.0478		ave	0.8781	0.8271	0.8710

on the graphs in Fig. 11. Figures 12 (a)-(d) show the frames that correspond to the lowest and highest measured disparity activities in sequences "Street2" and "University2". The 727^{th} frame of "Street2" has a low disparity activity (0.01), while the 46^{th} frame of "University2" has a high disparity activity (0.96).

Figures 12 (e)-(h) show the frames containing the lowest and highest measured motion activities in "Library6" and "Metro3". The 442^{nd} frame of "Library6" has a low motion activity (0.19), while the 287^{th} frame of "Metro3" has a motion activity of 0.81 because of large camera motion.

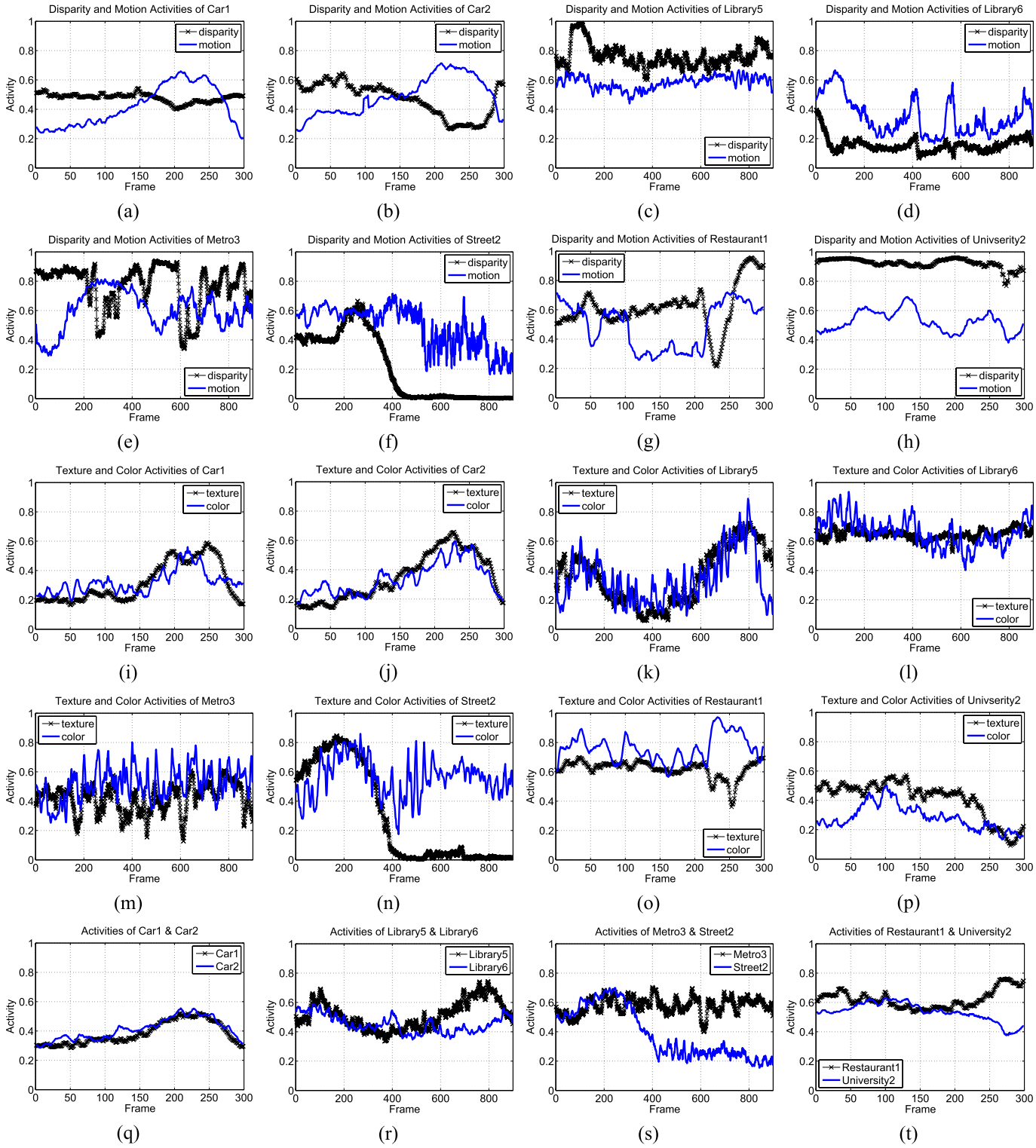


Fig. 11. 3DVA plotted against time. (a) Disparity and motion activities of "Car1", (b) "Car2", (c) "Library5", (d) "Library6", (e) "Metro3",(f) "Street2" (g) "Restaurant1", and (h) "University2". (i) Texture and color activities of "Car1", (j) "Car2", (k) "Library5", (l) "Library6", (m) "Metro3", (n) "Street2" (o) "Restaurant1", and (p) "University2". (q) 3DVA of "Car1", "Car2", (r) "Library5", "Library6", (s) "Metro3", "Street2" (t) "Restaurant1", and "University2".

Figures 13 (a)-(d) show the frames having the lowest and highest texture activities in "Car1" and "Car2". The 50^{th} frame of "Car1" and the 228^{th} frame of "Car2" have a low texture activity of 0.17 and a relatively high texture activity of 0.66, respectively. Figures 13 (e)-(h) show the 876^{th} frame of "Library5" and the 233^{rd} frame of "Restaurant1", which have a low color activity of 0.10 and a high color activity of 0.97, respectively.

TABLE II

SHAPE PARAMETER γ FOR THE "UNIVERSITY2" SEQUENCE

	Scale	Horizontal	Vertical	Diagonal		Scale	Horizontal	Vertical	Diagonal
Disparity	1	0.3421	0.3637	0.2271	Motion	1	0.5489	0.5562	0.4217
	2	0.9332	0.7913	0.4820		2	1.2035	1.1944	0.9962
	3	0.9943	0.6248	0.6042		3	0.9988	1.0466	1.2644
	ave	0.7565	0.5933	0.4378		ave	0.9170	0.9324	0.8941
	Scale	Horizontal	Vertical	Diagonal		Scale	Horizontal	Vertical	Diagonal
Texture	1	0.8322	0.8086	0.8464	Color	1	0.5243	0.5296	0.5359
	2	1.2203	1.2063	0.9948		2	1.0224	1.0707	1.0496
	3	1.3658	1.2102	0.9406		3	1.0243	1.0352	1.0227
	ave	1.1395	1.0750	0.9272		ave	0.8570	0.8785	0.8694

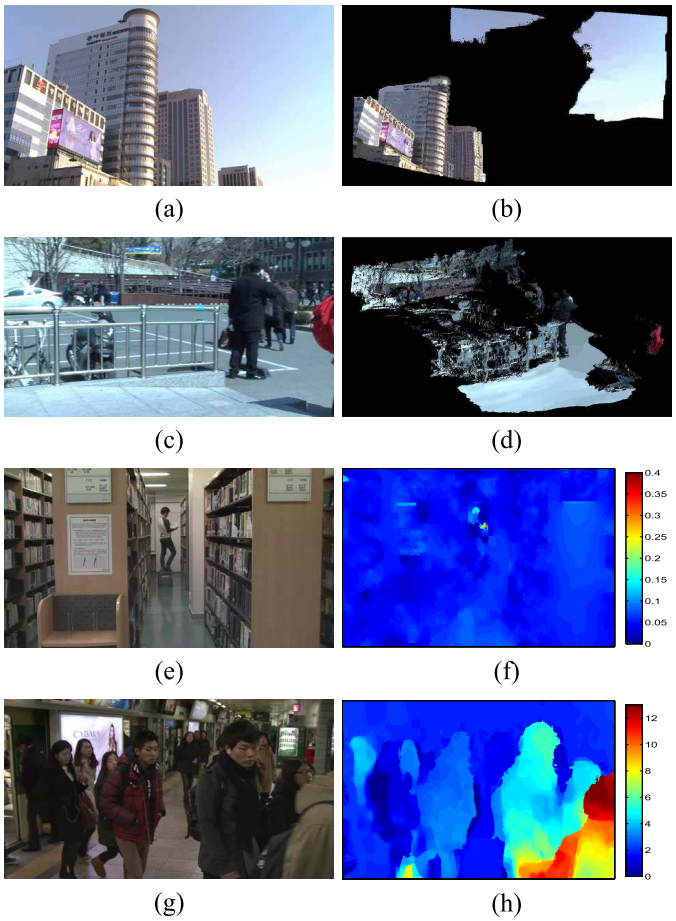


Fig. 12. Sample frames corresponding to the results in Fig. 11. (a) 727^{th} frame of "Street2". (b) 3D reconstruction of (a) showing low disparity activity. (c) 46^{th} frame of "University2". (d) 3D reconstruction of (c) showing high disparity activity. (e) 442^{nd} frame of "Library6". (f) Magnitude of motion in (e) showing low motion activity. (g) 287^{th} frame of "Metro3". (h) Magnitude of motion of (g) showing high motion activity.

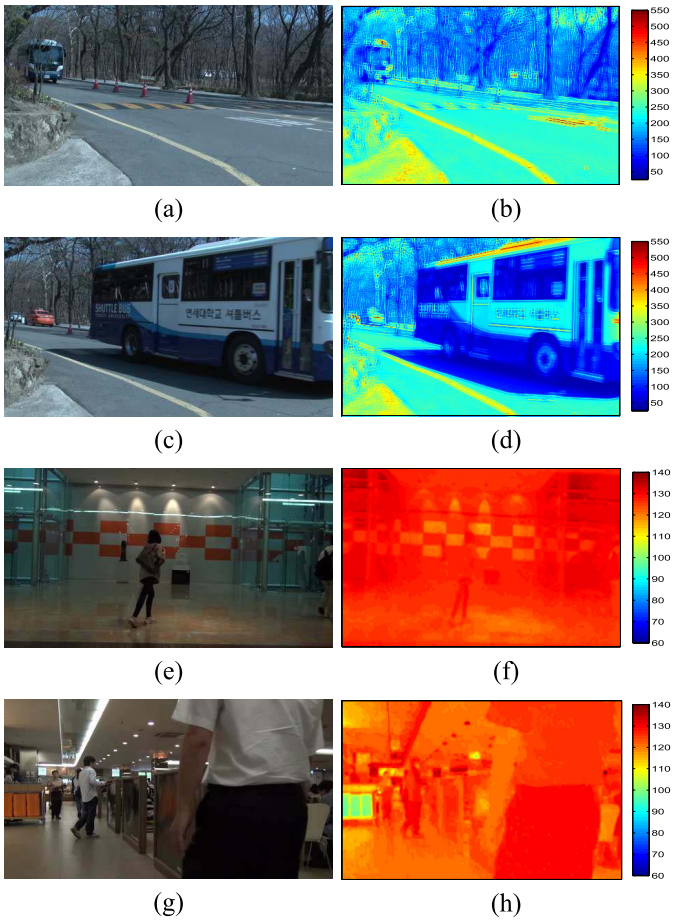


Fig. 13. Sample frames corresponding to the results in Fig. 11. (a) 50^{th} frame of "Car1". (b) Texture map of (a) showing low texture activity. (c) 228^{th} frame of "Car2". (d) Texture map of (c) showing high texture activity. (e) 876^{th} frame of "Library5". (f) Average of color components of (e) showing low color activity. (g) 233^{rd} frame of "Restaurant1". (h) Average of color components of (g) showing high color activity.

D. Application of 3DVA

As an example application of 3DVA, we use 3DVA to measure visual fatigue. When viewing 3D content, a viewer receives two images corresponding to the left and right views of the scene, and the convergence of the human visual system allows for the creation of a fused image. As we have seen, there exists a conflict between convergence and the accommodation, which causes visual fatigue [34]. Visual fatigue manifests in a wide range of visual symptoms, including tiredness, headaches, ocular pain and so on [66]. Highly 'active' 3D content often leads to increased visual fatigue [33]– [40], and hence, the use of a perceptual activity measure could be used to predict fatigue.

We conducted a subjective study to evaluate the relation between the human perception of visual fatigue and the proposed 3DVA measure. Forty subjects (14 females, 26 males) between the ages of 20 and 50 years old participated in the study. Fifteen of the subjects were researchers in video processing, while the rest were naive observers. Five of the researchers and ten of the non-researchers fell in the age group $31 - 50$, while the remaining twenty-five fell in the

TABLE III
CORRELATION RESULTS BETWEEN 3DVA AND SUBJECTIVE VISUAL FATIGUE

	PLCC	SROCC
Ages $20-30$	0.6867	0.6737
Ages $31-50$	0.5707	0.5385
Researcher	0.7482	0.7295
Non-researcher	0.5686	0.5517
Overall	0.7499	0.7344

age group $20-30$. All subjects were found to have corrected visual acuity of better than 1.25 (the Landolt C-test) and good stereoscopic acuity of less than 60 arc (on the RANDOT stereo test). The video sequences used in the subjective test were drawn from [44] and are shown in Fig. 10. The experiment was conducted using a Miracube 46" polarized Stereoscopic display. The viewing distance was set at 2.29m, which is four times the screen height.

We performed the subjective assessment by using a multimodal interactive continuous scoring of quality (MICSQ) technique [67] which helps engage and focus the subject on his/her task. For studies which attempt to ferret out relatively subtle measures such as fatigue, it is necessary to be able to deploy reliable methodologies that measure viewer's subjective experience. MICSQ is composed of a device interaction process between the 3D display and a separate device (PC, tablet, etc.) used as an assessment tool; and a human interaction process between the subject and the device. The scoring process is multimodal and uses aural and tactile cues. The authors in [67] found that assessment using MICSQ yields consistent, highly-reliable human responses and allows for a wide range of visual content to be graded, as compared to conventional single stimulus continuous quality evaluation (SSCQE).

Subjects were required to rate the amount of visual fatigue they experienced when they viewed the stereoscopic content. Subjects rated the videos continuously (i.e., as a function of time/on every frame) on a scale of 0-1, where a score of 0 corresponds to "most comfortable" and 1 corresponds to "most fatigued". The subject rejection procedure described in the ITU-R BT.500 [68] was applied to the scores obtained which rejected 5 of the 40 subjects. The remaining scores were then averaged to produce a mean visual fatigue score. We evaluated 3DVA as an indicator of visual fatigue by comparing the algorithm output to the visual fatigue scores.

To combine the four factors with appropriate weights ω_k ($k \in \{\mathrm{D,M,T,C}\}$) in (23), we applied a support vector regression (SVR). Specifically, we used the SVR to estimate the relative importance of the four quantities being weighted. The weights so computed turned out to be $\omega_{\mathrm{D}} = 0.3557$, $\omega_{\mathrm{M}} = 0.4471$, $\omega_{\mathrm{T}} = 0.1159$, and $\omega_{\mathrm{C}} = 0.1961$.

To better understand the factors that affect the relationship between visual fatigue and predictions of it using 3DVA, we also separately analyzed performance by age category (20 to 30 vs. 31 to 50) and profession (researcher vs. non-researcher). Figure 14 plots recorded subjective visual fatigue scores against computed 3DVA index values. Figures 14 (a) to (e) are scatter plots for researchers, for non-researchers, for

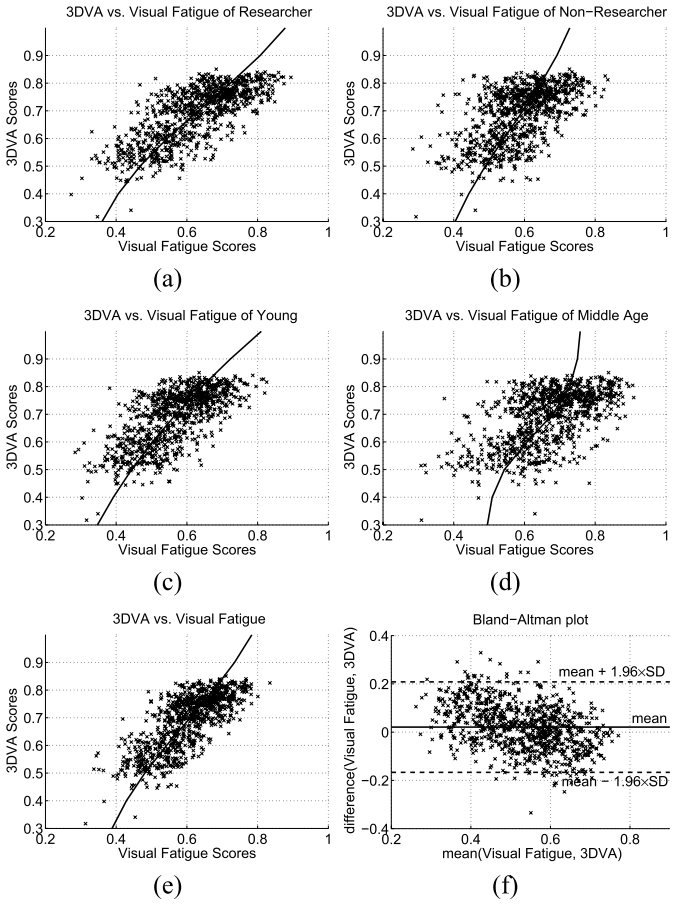


Fig. 14. Scatter plots between 3DVA prediction scores and subjective visual fatigue scores (a) over researchers, (b) of non-researchers, (c) of age group 20-30, (d) of age group 31-50, and (e) of all subjects. (f) Bland-Altman plots over all subjects.

age group $20-30$, for age group $31-50$, and for all subjects, respectively, while Fig. 14 (f) is a Bland-Altman plot over all subjects. The figures demonstrate qualitatively the degree and nature of the correlation between 3DVA predictions and subjective visual fatigue scores. The 3DVA predictions were compared with subjective fatigue scores at a rate of $2Hz$ ($2\ samples/s$) following [68]. Thus, the 3DVA predictions were averaged over each set of 15 frames. In order to measure quantitative correlation between 3DVA and subjective fatigue, we employed two commonly used performance measures: the Pearson linear correlation coefficient (PLCC) and the Spearman rank-order correlation coefficient (SROCC) between the subjective visual fatigue results and the fitted 3DVA scores, obtained by the regression method in [69]. A four-parameter, monotonic logistic function was used to fit the predicted fatigue predictions to the subjective fatigue scores

$$Q'_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-(Q_j - \beta_3/|\beta_4|)}} \tag{24}$$

using nonlinear least squares optimization. The correlation results are given in Table III. The computed PLCC and SROCC all subjects indicates that 3DVA functions quite reasonably well as a predictor of visual fatigue. There are differences in the results for the different subject age ranges.

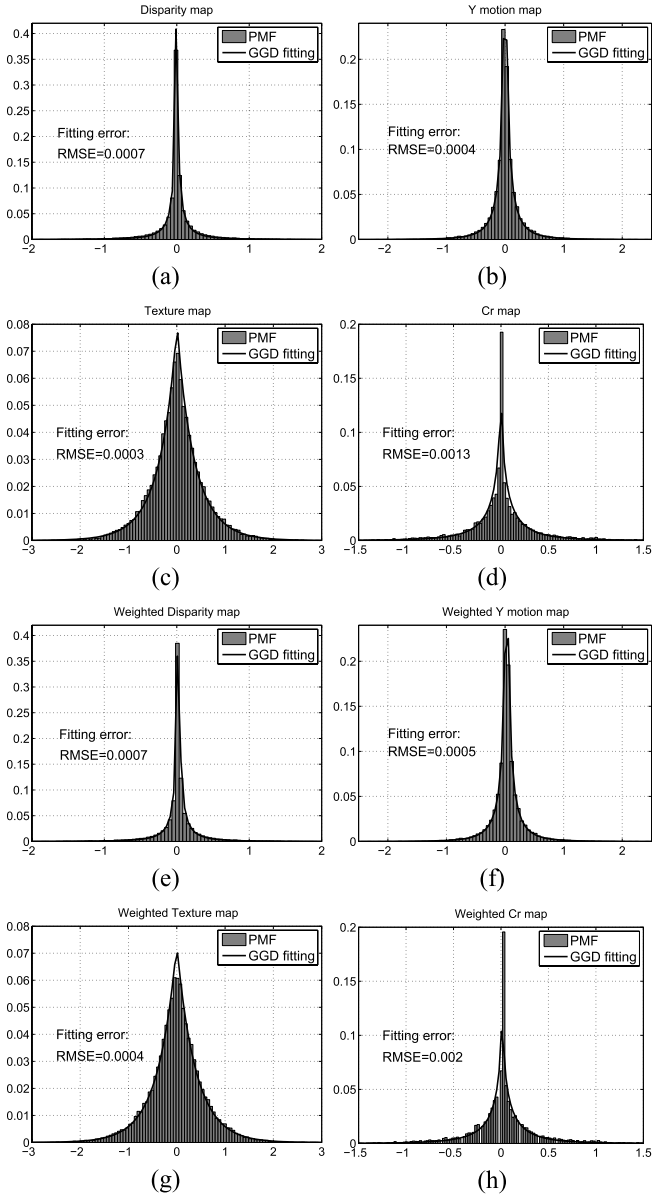


Fig. 15. Examples of GGD fitting for the wavelet data of "University 1" at the 78$^{\text{th}}$ frame (Fig. 7). All graphs correspond to the 3$^{\text{rd}}$ scale and vertical orientation. (a) Disparity map (Fig. 7 (a)). (b) Y-motion map (Fig. 7 (c)). (c) Texture map (Fig. 7 (e)). (d) Cr-color map (Fig. 7 (g)). (e) Weighted disparity map by (12)-(14). (f) Weighted Y-motion map. (g) Weighted texture map. (h) Weighted Cr-color map.

First, the correlation between 3DVA and subjects is higher in the younger age group. There was also a tendency among the older viewers to experience a greater degree of visual fatigue (0.7138 and 0.6011 were the average subjective visual fatigue scores recorded by the older and younger age groups, respectively). One possible explanation for this could be decreased visual sensitivity with age which is not accounted for in 3DVA. Moreover, the 3DVA correlated higher with the researcher scores than with the non-researcher scores, possibly indicating a bias derived from prior experience.

A Bland-Altman plot can provide useful information with regards to the ranges of values over which the two results are most concordant or discordant [70]. It is common to compute 95% limits of agreement for each comparison (average difference ± 1.96 standard deviation of the difference), which indicates how far apart the two results were likely to be for most individuals. As shown in Fig. 14 (f), the region in which the differences were mostly located was $(-0.17 \sim 0.21)$.

The analysis in this section serves as a demonstration of the application power of 3DVA. In the future we plan to adopt more sophisticated machine-learning based approaches [71], [72] using 3DVA features (disparity, motion, texture and color) that mirror our recent design philosophy in the field of no-reference image quality assessment [73], [74].

TABLE IV
FITTING ERRORS (RMSE) OF PMF AND WEIGHTED PMF

	Disparity	Motion	Texture	Color
PMF	0.0008	0.0005	0.0007	0.0011
Weighted PMF	0.0009	0.0007	0.0009	0.0015

IV. CONCLUSION

When analyzing 3D content, it is essential to be able to quantify visual importance over the perceived 3D visual space. The human visual response is highly reliant on the size, position and motion of objects aligned along the depth axis. Along these lines, we proposed a new framework for analyzing 3D content termed 3D visual activity (3DVA). 3DVA measures the statistical dynamics of objects in 3D space. To achieve this, we accounted for the nonuniform sampling resolution of the eye (foveation) and for 3D stereoscopic fusion by casting the problem in a coordinate-transformed space. Over nonuniform transformed coordinates, we fitted wavelet coefficients to a generalized Gaussian distribution model on disparity, motion, texture and color data. The new 3D visual activity index called 3DVA was formulated by quantifying the statistical variance and randomness of the 3D content. Randomness, activity and complexity are important for understanding 3D content, e.g., when analyzing and predicting visual fatigue experienced when viewing 3D visual content.

APPENDIX

In this appendix, we experimentally verify that the GGD model supplies good experimental fits to each PMF, $\mathbf{p}_{k,i}$. This is importance since, while the GGD model is well-established for luminance data, it has not been applied to the other forms of visual data studied here.

As an example, Fig. 15 plots the results of GGD fitting to the empirical histograms of the disparity, motion, texture and color maps of "University 1" at the 78$^{\text{th}}$ frame as shown in Fig. 7. Figures 15 (a)-(d) are the PMF and GGD fits of the un-weighted wavelet data and (e)-(h) are those of the weighted wavelet data. As the Figure demonstrates, applying the 3D HVS weight changes (narrows) the shape (spread) of the distribution, however, the modified shape is still well modeled using a GGD.

In Table IV, we tabulate the root mean-square-error (RMSE) between actual histogram values and the best GGD fit for both the weighted and un-weighted PMFs. As the table illustrates, the fitting errors for both these cases are comparable and extremely small, thus verifying that the GGD is a good model in both cases.

REFERENCES

[1] E. Lantz, "Future directions in visual display systems," *ACM SIGGRAPH Comput. Graph.*, vol. 31, no. 2, pp. 38–42, May 1997.

[2] L. Steinbach, "3D or not 3D is that a question," *Museum J.*, vol. 54, no. 1, pp. 41–54, Jan. 2011.

[3] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*, 2nd ed. Pacific Grove, CA, USA: Brooks/Cole, 1999.

[4] C.-W. Ngo, T.-C. Pong, and H.-J. Zhang, "Motion analysis and segmentation through spatio-temporal slices processing," *IEEE Trans. Image Process.*, vol. 12, no. 3, pp. 341–355, Mar. 2003.

[5] H. Greenspan, C. H. Anderson, and S. Akber, "Image enhancement by nonlinear extrapolation in frequency space," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1035–1048, Jun. 2000.

[6] T. Bose, F. Meyer, and M. Q. Chen, *Digital Signal and Image Processing*. New York, NY, USA: Wiley, 2004.

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

[8] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.

[9] S. Lee and A. C. Bovik, "Fast algorithms for foveated video processing," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 2, pp. 149–162, Feb. 2003.

[10] H. Lee and S. Lee, "Visual entropy gain for wavelet image coding," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 553–556, Sep. 2006.

[11] H. Ha, T. Oh, and S. Lee, "Macroblock-based frequency selective weighting for visual scalable video coding of H.264/AVC," *IEEE Trans. Broadcast.*, vol. 55, no. 3, pp. 559–568, Sep. 2009.

[12] H. Ha, J. Park, S. Lee, and A. C. Bovik, "Perceptually scalable extension of H.264," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 21, no. 11, pp. 1667–1678, Nov. 2011.

[13] C. J. van den Branden Lambrecht and M. Kunt, "Characterization of human visual sensitivity for video imaging applications," *Signal Process.*, vol. 67, pp. 255–269, May 1996.

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

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

[16] F. Yang, S. Wan, Y. Chang, and H. R. Wu, "A novel objective noreference metric for digital video quality assessment," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 685–688, Oct. 2005.

[17] ITU-T Recommendation BT.500-10, *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Geneva, The Switzerland: International Telecommunication Union, 2000.

[18] 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.

[19] A. Boev, M. Poikela, A. Gotchev, and A. Aksay, "Modeling of the stereoscopic HVS," MOBILE3DTV, Tech. Rep. D5.3, Jul. 2009.

[20] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[21] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

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

[23] S. Osindero, M. Welling, and G. E. Hinton, "Topographic product models applied to natural scene statistics," *Neural Comput.*, vol. 18, no. 2, pp. 381–414, 2006.

[24] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.

[25] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature, Neurosci.*, vol. 4, pp. 819–825, Aug. 2001.

[26] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[27] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Ann. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[28] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Ann. Rev. Psychol.*, vol. 59, no. 1, pp. 167–192, Jan. 2008.

[29] G. F. Poggio and B. Fischer, "Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey," *J. Neurophysiol.*, vol. 40, no. 6, pp. 1392–1405, Nov. 1977.

[30] S. Prince, B. G. Cumming, and A. J. Parker, "Range and mechanism of encoding of horizontal disparity in macaque V1," *J. Neurophysiol.*, vol. 87, no. 1, pp. 209–221, Jan. 2002.

[31] S. J. D. Prince, A. D. Pointon, B. G. Cumming, and A. J. Parker, "Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms," *J. Neurophysiol.*, vol. 87, no. 1, pp. 191–208, Jan. 2002.

[32] 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, Feb. 2003.

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

[34] S. Pastoor, *COST230 Final Report Working Group 1 Psychooptics*. Roma, Italy: Fondazinoe Ugo Bordoni, 1998, pp. 17–32.

[35] C. W. G. Clifford, "Perceptual adaptation: Motion parallels orientation," *Trends Cognit. Sci.*, vol. 6, no. 3, pp. 136–143, Mar. 2002.

[36] F. Speranza W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," *Proc. SPIE*, vol. 6055, pp. 94–103, Jan. 2006.

[37] A. Wilkins, I. N. Smith, A. Tait, C. Mcmanus, S. D. Sala, A. Tilley, *et al.*, "A neurological basis for visual discomfort," *Brain*, vol. 107, no. 4, pp. 989–1017, 1984.

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

[39] S. M. Haigh, L. Barningham, M. Berntsen, L. V. Coutts, E. S. T. Hobbs, J. Irabor, *et al.*, "Discomfort and the cortical haemodynamic response to coloured gratings," *Vis. Res.*, vol. 89, pp. 47–53, Aug. 2013.

[40] I. Juricevic, L. Land, A. J. Wilkins, and M. A. Webster, "Visual discomfort and natural image statistics," *Perception*, vol. 39, no. 7, pp. 884–899, 2010.

[41] M. Iwasaki and H. Inomata, "Relation between superficial capillaries and foveal structures in the human retina," *Invest. Ophthalmol. Vis. Sci.*, vol. 27, no. 12, pp. 1698–1705, 1986.

[42] H. Kim, S. Lee, and A. C. Bovik, "Saliency prediction on stereoscopic videos," *IEEE Trans. Image Process.*, to be published.

[43] Y. Liu, L. K. Cormack, and A. C. Bovik, "Dichotomy between luminance and disparity features at binocular fixations," *J. Vision*, vol. 10, no. 12 pp. 1–17, 2010.

[44] (2008). *IEEE Standards Association Stereoscopic Database* [Online]. Available: http://grouper.ieee.org/groups/3dhf/

[45] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vis. Res.*, vol. 21, no. 3, pp. 409–418, 1981.

[46] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: Limited imposed by optics, photoreceptors and receptor pooling," *J. Opt. Soc. Amer.*, vol. 8, no. 11, pp. 1775–1787, 1991.

[47] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low bandwidth video communication," *Proc. SPIE*, vol. 3299, pp. 1–8, Jul. 1998.

[48] (2012). *Middlebury Stereo* [Online]. Available: http://vision.middlebury.edu/stereo/data/

[49] A. Coltekin, "Foveation for 3D visualization and stereo imaging," Ph.D. dissertation, Dept. Sci. Technol., Helsinki Univ. Technol., Helsinki, Finland, 2006.

[50] T. Ohshima, H. Yamamoto, and H. Tamura, "Gaze-directed adaptive rendering for interacting with virtual space," in *Proc. IEEE VRAIS*, Apr. 1996, pp. 103–110.

[51] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," New York, NY, USA: Springer-Verlag, Apr. 2008.

[52] K.-J. Oh, S. Yea, and Y.-S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3D video," in *Proc. Picture Coding Symp.*, May 2009, pp. 1–4.

[53] 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.

[54] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, "Ensemble of local and global information for finger-knuckle-print recognition," *Pattern Recognit.*, vol. 44, no. 9, pp. 1990–1998, Sep. 2011.

[55] R. Gao and W. F. Bischof, "Detection of linear structures in remote-sensed images," in *Image Analysis and Recognition*. New York, NY, USA: Springer-Verlag, 2009, pp. 896–905.

[56] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. pp. 444–447, Oct. 1995.

[57] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, no. 2, pp. 181–198, 1992.

[58] Z. Wang and A. C. Bovik, "Reduced and no reference visual quality assessment- The natural scene statistic model approach," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 29–40, Nov. 2011.

[59] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Adv. Neural Inf. Process. Syst.*, vol. 12, no. 1, pp. 855–861, 2000.

[60] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, "Natural image statistics and divisive normalization: Modeling nonlinearities and adaptation in cortical neurons," in *Probabilistic Models of the Brain: Perception and Neural Function*. Cambridge, MA, USA: MIT Press, Feb. 2002, pp. 203–222.

[61] J. Huang, A. Lee, and D. Mumford, "Statistics of range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2000, pp. 324–331.

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

[63] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[64] A. Jaina, K. Nandakumara, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.

[65] R. Cappelli, D. Maio, and D. Maltoni, "Combining fingerprint classifiers," in *Proc. 1st Int. Workshop MCS*, Jun. 2000, pp. 351–361.

[66] *Image Safety—Reducing the Incidence of Undesirable Biomedical Effects Caused by Visual Image Sequences*, Geneva, The Switzerland, ISO, 2005.

[67] T. Kim, J. Kang, S. Lee, and A. C. Bovik, "Multimodal interactive continuous scoring of subjective 3D video quality of experience," *IEEE Trans. Multimedia*, to be published.

[68] ITU, "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Tech. Rep. BT.500-13, 2012.

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

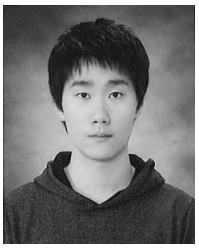
[70] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, no. 8476, pp. 307–310, Feb. 1986.

[71] B. Scholkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," *Neural Computat.*, vol. 12, no. 5, pp. 1207–1245, 2000.

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

[73] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[74] 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. 20, no. 12, pp. 3350–3364, Mar. 2012.



Kwanghyun Lee received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2008 and 2010, respectively, where he is currently pursuing the Ph.D. degree in 2010. His research interests include quality assessment of 2D and 3D image and video, 3D video coding, cross-layer optimization, and wireless multimedia communications.

Anush Krishna Moorthy received the B.E. degree in electronics and telecommunication from the University of Pune, Pune, India, in 2007, the M.S. degree in electrical engineering from the University of Texas at Austin in 2009, and the Ph.D. degree from the University of Texas at Austin in 2012.

He joined the Laboratory for Image and Video Engineering (LIVE), University of Texas, Austin, in 2008, and was the Assistant Director of LIVE from 2008 to 2012. He is a recipient of the Continuing Graduate Fellowship from The University of Texas at Austin from 2010 to 2011, the Professional Development Award in 2009 and 2010, and the Center for Perceptual Systems Travel Grant in 2010 and the TATA scholarship for higher education abroad. He was an Advanced Imagining Engineer with Texas Instruments, Dallas, TX, USA, from 2012 to 2013, and is currently a Senior Video Systems Engineer with Qualcomm, Inc., San Diego, CA, USA.

His research interests include image and video quality assessment, image and video compression, and computational vision.

Sanghoon Lee (M'05–SM'12) received the B.S. degree in electrical engineering from Yonsei University in 1989 and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology in 1991. From 1991 to 1996, he was with Korea Telecom. He received the Ph.D. degree in electrical engineering from the University of Texas at Austin in 2000. From 1999 to 2002, he was with Lucent Technologies on 3G wireless and multimedia networks. In 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, where he is a Full Professor. He has been an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2010 and 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 served as the General Chair of the 2013 IEEE IVMSP workshop and a Guest Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING in 2013. He has received the 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 is the Curry/Cullen Trust Endowed Chair Professor with the University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering. He is a Faculty Member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published more than 650 technical articles in these areas and holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

He has received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award in 2009; the Education Award in 2007; the Technical Achievement Award in 2005, and the Meritorious Service Award in 1998. He was named recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, received the SPIE Technology Achievement Award for 2012, and was the IS&T/SPIE Imaging Scientist of the Year for 2011. He received the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana in 2008, the IEEE Third Millennium Medal in 2000, and the two Journal Paper Awards from the International Pattern Recognition Society in 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, including: a Board of Governors, IEEE Signal Processing Society from 1996 to 1998; a co-founder and Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002; an Editorial Board of *The Proceedings of the IEEE* from 1998 to 2004; a Series Editor for *Image, Video, and Multimedia Processing* (Morgan and Claypool Publishing Company, 2003); and a Founding General Chairman, the First IEEE International Conference on Image Processing, Austin, TX, in 1994.

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