

# Color and Depth Priors in Natural Images

Che-Chun Su, Lawrence K. Cormack, and Alan C. Bovik, *Fellow, IEEE*

**Abstract**—Natural scene statistics have played an increasingly important role in both our understanding of the function and evolution of the human vision system, and in the development of modern image processing applications. Because range (egocentric distance) is arguably the most important thing a visual system must compute (from an evolutionary perspective), the joint statistics between image information (color and luminance) and range information are of particular interest. It seems obvious that where there is a depth discontinuity, there must be a higher probability of a brightness or color discontinuity too. This is true, but the more interesting case is in the other direction—because image information is much more easily computed than range information, the key conditional probabilities are those of finding a range discontinuity given an image discontinuity. Here, the intuition is much weaker; the plethora of shadows and textures in the natural environment imply that many image discontinuities must exist without corresponding changes in range. In this paper, we extend previous work in two ways—we use as our starting point a very high quality data set of co-registered color and range values collected specifically for this purpose, and we evaluate the statistics of perceptually relevant chromatic information in addition to luminance, range, and binocular disparity information. The most fundamental finding is that the probabilities of finding range changes do in fact depend in a useful and systematic way on color and luminance changes; larger range changes are associated with larger image changes. Second, we are able to parametrically model the prior marginal and conditional distributions of luminance, color, range, and (computed) binocular disparity. Finally, we provide a proof of principle that this information is useful by showing that our distribution models improve the performance of a Bayesian stereo algorithm on an independent set of input images. To summarize, we show that there is useful information about range in very low-level luminance and color information. To a system sensitive to this statistical information, it amounts to an additional (and only recently appreciated) depth cue, and one that is trivial to compute from the image data. We are confident that this information is robust, in that we go to great effort and expense to collect very high quality raw data. Finally, we demonstrate the practical utility of these findings by using them to improve the performance of a Bayesian stereo algorithm.

**Index Terms**—Bayesian stereo algorithm, color images, Gabor filter bank, natural scene statistics (NSS).

## I. INTRODUCTION

**G**IVEN continuous, rapid advances in three-dimensional (3D) imaging and display technology, the quantity and quality of 3D and stereoscopic data, e.g., image, video,

Manuscript received March 26, 2012; revised November 5, 2012; accepted February 4, 2013. Date of publication February 26, 2013; date of current version April 12, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mary Comer.

C.-C. Su and A. C. Bovik are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: ccsu@utexas.edu; bovik@ece.utexas.edu).

L. K. Cormack is with the Department of Psychology, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: cormack@mail.utexas.edu). Digital Object Identifier 10.1109/TIP.2013.2249075

movies, geographic models, etc., has increased dramatically. A substantial amount of research has been conducted towards better understanding the perception of 3D content, with the aim of improving the quality of visual experience delivered by 3D technologies and products. For example, impairments in viewing and comfort when using 3D displays has been studied towards developing auto-stereoscopic 3D displays [1], [2]. There are numerous sources of distortion and visual discomfort that can be experienced when viewing 3D content. Understanding how the depth sensation is affected by improper geometry (stereography) and by signal distortion are crucial open problems [3]–[5]. We believe that reliable statistical models of natural depth/range and luminance/chrominance image information can be used to improve these processes. We develop such models here and show that joint color/depth statistical models derived from natural image and range data can be used to improve statistical approaches to the classic 3D problem of resolving binocular correspondences.

The evolution of the human vision apparatus has involved many different factors and driving forces, such as natural scene statistics, the computational resources available in the human brain, and the kinds of tasks that humans need to perform [6]. Natural scene statistics (NSS) have been proven to be important ingredients towards understanding both the evolution of the human vision system and the design of image processing algorithms [7]. Given disparity and other visual cues, the visual brain is able to reconstruct the geometry of the 3D visual space so quickly, effortlessly and seamlessly that an individual rarely feels how difficult and ill-posed this problem can be. Given that humans have evolved this capability by millions of years of adaptation to the structure and statistics of the visual environment, it becomes important to understand these statistics in order to understand the human vision system.

Extensive work has been conducted towards understanding the luminance statistics of natural scenes [8]–[11], and the link between natural scene statistics and neural processing of visual stimuli [12], [13]. It has been discovered that the distributions of local quantities such as luminance contrast are scale invariant, and that the power spectra of natural images vary as  $1/f^2$  with radial spatial frequency  $f$ . This has been successfully used to explain and predict early visual processing in both insects and higher vertebrates [8], [9], and [14]. The statistics of natural images have been found to exhibit non-Gaussian behavior, but when projected onto appropriate multi-scale spaces, e.g., using wavelet bases [15], or 2D Gabor decompositions [8], the resulting coefficients are found to obey regular statistical models, such as Gaussian scale mixtures [16]. These statistical models have been successfully applied in a variety of image and video applications, such as image de-noising and restoration [17], and image quality

assessment [7], and [18]–[20]. Moreover, it has also been suggested that the spatial receptive fields of the simple cells in mammalian primary visual cortex (V1) can be characterized as being localized, oriented, and bandpass, which are comparable with the basis functions of wavelet or Gabor transforms [21]. It is also widely believed that the goal of the early stages of visual signal processing is to transform and encode the input stimuli from natural images into a sparse, efficient representation to utilize the available computation resources of neurons [12], [22]. This sparse and efficient coding strategy along with its over-complete basis leads to non-linear relationships between visual stimuli and neural responses, which can be used to help understand higher stages of cortical processing in human vision systems [23]. In this paper, we leverage these psychophysical and physiological findings by performing multi-scale, multi-orientation decompositions on color image data with optimally localized and perceptually relevant Gabor filter banks. We do this to expand our understanding of the statistical relationships between cortical neural responses to luminance, color, and depth/range data in natural images.

Very little work has been done on analyzing the joint statistics of luminance and range in natural scenes, and we haven't found any relating the statistics of color and range. One major reason for the lack of studies on color and range statistics has been limited access to high quality databases of color images and associated ground-truth range maps. Potetz *et al.* [24] constructed a database of co-registered 2D color images and range maps, and discovered that there is a correlation between range and intensity of luminance in natural scenes. This negative range-luminance correlation merely reflects the fact that nearer objects tend to appear brighter than far objects, on average. The authors also deployed a few convex range filters, selected for specific structural properties relevant to computer vision, to filter both range and luminance images. Using a canonical correlation analysis, they found a relatively low degree of strictly linear correlation between the "structure-filtered" luminance and range patches. In a later study on the same dataset, Potetz *et al.* [25] examined the relationships between luminance and range over multiple scales and applied their results to shape-from-shading problems. In [26], Yang *et al.* explored the statistical relationships between luminance and disparity in the wavelet domain using a public co-registered database of range maps and luminance natural images from [24], and applied the derived models to a Bayesian stereo algorithm. The authors found that the correlations between bandpass luminance and bandpass disparity are stronger in coarser scales, and also showed that the statistical models of 3D natural scenes improve the quality of computed disparity maps. Su *et al.* [27] very recently constructed the LIVE Color+3D Database [28], a large co-registered database of high-quality 2D color images and high-resolution ground-truth range maps ( $1280 \times 720$ ), and explored simple statistical relationships between band-pass responses to luminance/chrominance and range gradients in natural scenes.

Color is an important and dense natural visual cue that is used by the brain to reconstruct both low-level and high-level visual percepts. The cone photoreceptors, which are densely

distributed in the fovea centralis of the retina, capture and convey rich information both in space and time. While the cones themselves do not encode color, they do come in three types that have different spectral sensitivities. Hence, comparisons of the outputs of the different cone types by the retinal ganglion cells allow dense spatiotemporal chromatic information to be transmitted from the retina to the primary visual cortex (V1). Likewise, color can be used at later processing stages to help infer large-scale shape information to better solve visual tasks by both humans and machine algorithms [29].

Moreover, it has been demonstrated that the perception of color and depth are related [30], and that chromatic information can be used to improve the solution of stereo correspondence problems [31], [32]. Therefore, interaction and correlation between color and depth need further examination. Towards obtaining a better understanding of the statistical relationships between color and range, we studied both the marginal and joint statistics of color and range using the co-registered color images and ground-truth range maps in the LIVE Color+3D Database [28]. To better approximate color perception in human vision systems, all color images in RGB were transformed into the more perceptually relevant CIELAB color space. We use CIELAB since it is optimized for quantifying perceptual color difference and it better corresponds to human color perception than does the perceptually nonuniform RGB space [33].

An important stereoscopic cue, disparity, comes from the angular difference between the two different retinal images received by the two frontally placed, horizontally separated eyes. It has been verified that there exist simple and complex neurons tuned to binocular disparity in V1 [34], [35], and the human vision system has very fine stereo acuity, which falls between 2 arcsec to 6 arcsec under the best conditions [36]. The visual system also has a large upper disparity limit, which reaches  $7^\circ$  for crossed disparities and  $12^\circ$  for uncrossed disparities [37]. The excellent acuity and broad operating range of stereopsis indicate that disparity is extensively used for depth perception. Liu *et al.* [38] studied the disparity distributions of natural scenes by converting forest range maps to disparity maps. They found that the disparity distributions at eye level are centered at zero, non-Gaussian, but well modeled as generalized Gaussian. A similar study on indoor range maps showed similar results. Moreover, the authors correlated disparity sensitivity with naturally available disparities by showing that the proportion of near- and far-tuned disparity distributions qualitatively agrees with the distribution of disparity-tuning neurons in V1 [39]. This suggests that the human vision system may use the rich disparity cues both in near- and far-viewing distances to recover the depth information in natural scenes.

We now proceed as follows. Section II briefly introduces the database used in the statistical analysis and modeling, and Section III explains methods that were used to pre-process both image and range data. The detailed statistical modeling of depth/range and luminance/chrominance data is then explained in Sections IV and V. Section VI develops a Bayesian stereo algorithm using the derived statistical models, which delivers good results as shown by the simulation results in

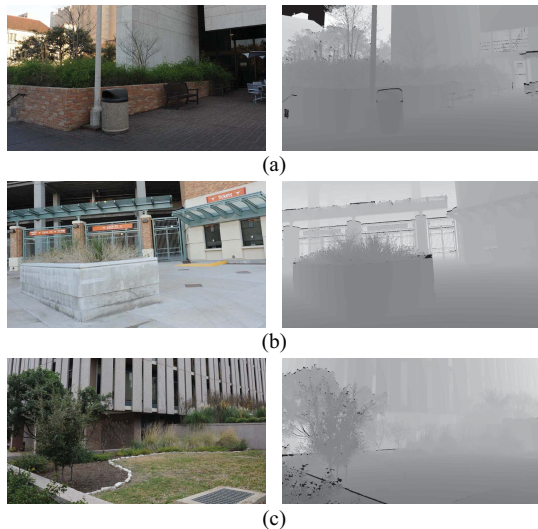


Fig. 1. Three examples of the natural scenes, 2D color images on the left and aligned 2D range maps on the right. (a) Scene 1. (b) Scene 2. (c) Scene 3.

Section VII. Finally, Section VIII concludes with thoughts for future work.

## II. DATA ACQUISITION

To derive the natural scene statistics of image and range data in a more realistic and wider viewing angle, we use the LIVE Color+3D Database, which contains 12 sets of color images with corresponding ground-truth range maps at a high-definition resolution of  $1280 \times 720$  [27], [28]. The image and range data in the LIVE Color+3D Database were collected using an advanced range scanner, RIEGL VZ-400, with a Nikon D700 digital camera mounted on top of it [40]. Since there are inevitable translational and rotational shifts when mounting the camera onto the range scanner, calibration is performed before data acquisition. The mounting calibration is done manually using the RIEGL RiSCAN PRO software, which is designed for scanner operation and data processing [41]. Next, to acquire the image and range data in natural scenes, the device obtains distances by lidar reflection and waveform analysis as it rotates, and then the digital camera takes an optical photograph with the same field of view. The acquired range data are exported from the range scanner as point clouds with three-dimensional coordinates and range values, while the image data are stored in the digital camera. Finally, to obtain the aligned 2D range map with the 2D image, the 3D point clouds are projected and transformed into the 2D range map by applying the pinhole camera model with lens distortion [42].

The natural environments where the image and range data were collected include areas around Austin, Texas, including the campus at The University of Texas and the Texas State Capitol. Fig. 1 shows three examples of acquired natural scenes with 2D color images and aligned 2D range maps. Note that in our database, the unit of the range data acquired by using the RIEGL VZ-400 range scanner is meter (m), with precision up to 3 (mm), and for display purposes, the range maps shown in Fig. 1 are first re-scaled using the

logarithm, i.e.,  $R' = \log(1 + R)$ , and normalized to  $[0, 255]$ . The black holes, i.e., range values equal to zero, are locations where the range scanner was not able to measure the reflected laser beams and report range values. The reasons include the material of the object surface, wind disturbance, limitation of the scanner, etc. These invalid range values are carefully avoided during the filtering process in the statistical analysis and modeling.

## III. DATA PRE-PROCESSING

Human vision systems extract abundant information from natural environments by processing visual stimuli through different levels of decomposition and interpretation. Since we want to learn and explore the statistical relationships between luminance/chrominance and range/disparity and how these statistics might be implicated in visual processing, and subsequently used in image processing algorithms, some pre-processing was performed on both the 2D color images and the co-registered 2D ground-truth range maps.

### A. Color Space Conversion and Gabor Filter Bank

All color images were transformed into the perceptually relevant CIELAB color space having one luminance ( $L^*$ ) and two chrominance ( $a^*$  and  $b^*$ ) components. CIELAB color space is optimized to quantify perceptual color differences and better corresponds to human color perception than does the perceptually nonuniform RGB space [33]. The coordinate  $L^*$  of the CIELAB space represents the lightness of the color,  $a^*$  represents its position between red/magenta and green, and  $b^*$  represents its position between yellow and blue. Moreover, the nonlinear relations for  $L^*$ ,  $a^*$ , and  $b^*$  mimic the nonlinear responses of human eyes, starting from the cone cells (L, M, and S) in the retina. Each image was then decomposed by a 2D Gabor filter bank over multiple scales and orientations, which serves to mimic the receptive fields of simple cells in V1 [8], [43]–[45]. Both the luminance and chrominance components of the transformed color images and the converted disparity maps were filtered by the same 2D Gabor filter bank.

Before discussing the statistical analysis and modeling, we wish to briefly motivate them by discussing the formation of receptive fields in V1 neurons and their relevance to understanding natural scene statistics. From physiological evidence [46], it is known that the simple cells in V1 process visual signals received from LGN (Lateral Geniculate Nucleus) neurons. The simple cells can be linearly modeled as having elongated, center-surround receptive fields that are highly selective in spatial frequency and orientation, remarkably like a Gabor filter. Thus, the physical statistics of the natural environment are manifested in the spectral responses of neurons in the visual cortex [12], [13], and likely in the responses of disparity-tuned neurons as well [38].

A complex 2D Gabor filter can be written

$$G(x, y, \sigma_1, \sigma_2, \zeta_x, \zeta_y, \theta) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2} \left[ \left( \frac{R_1}{\sigma_1} \right)^2 + \left( \frac{R_2}{\sigma_2} \right)^2 \right]} e^{i(x\zeta_x + y\zeta_y)} \quad (1)$$

where  $R_1 = x \cos \theta + y \sin \theta$  and  $R_2 = -x \sin \theta + y \cos \theta$ ,  $\sigma_1$  and  $\sigma_2$  are the standard deviations of an elliptical Gaussian envelope along the rotated axes,  $\zeta_x$  and  $\zeta_y$  are the spatial center frequencies of the complex sinusoidal carrier, and  $\theta$  is the orientation.

Since physiological evidence shows that visual neurons in primary visual cortex usually have an elliptical Gaussian envelope with an aspect ratio of 0.25–1.0, with propagating direction along the short axis of the elliptical Gaussian envelope [47], we use complex 2D Gabor filters of the form

$$G(x, y, \gamma, \sigma, \omega, \theta) = \frac{1}{2\pi\gamma\sigma^2} e^{-\frac{1}{2}\left[\left(\frac{R_1}{\sigma}\right)^2 + \left(\frac{R_2}{\gamma\sigma}\right)^2\right]} e^{i\omega R_1} \quad (2)$$

where  $\gamma = \sigma_y/\sigma_x$  is the aspect ratio of the elliptical Gaussian envelope,  $\sigma = \sigma_x$ , and  $\omega = \sqrt{\zeta_x^2 + \zeta_y^2}$  is the radial center frequency. To create a suitable set of Gabor filter banks which can cover most of the frequency domain, the two parameters of the elliptical Gaussian envelope need to be chosen properly, including the aspect ratio,  $\gamma$ , and the standard deviation,  $\sigma$  [48]. Here, six spatial center frequencies, 0.84, 1.37, 2.22, 3.61, 5.87, and 9.53 (cycles/degree) are used, with four different sinusoidal grating orientations for each spatial frequency: horizontal (0-deg), diagonal-45 (45-deg), vertical (90-deg), and diagonal-135 (135-deg) [49], [50]. The aspect ratio,  $\gamma$ , is chosen to be 1.0 [47]. The spatial frequency bandwidth of each sub-band is 0.7 (octave), and neighboring filters intersect at half-power point, i.e., 3-dB point [49], [51].

### B. Range and Disparity

In the following sections, we study the statistics of decomposed range data and explore the statistical relationship of range with decomposed luminance and chrominance in natural images. First, we examine the conditional distribution of range gradient data given luminance/chrominance. Since the depth information acquired by the human vision system is more relative than absolute, which means that we know which objects are further and which are closer, but we are not sure about the exact distance of each object from us, the disparity serves to be an important stereoscopic cue and its statistics is of most interest towards understanding depth perception. Therefore, in order to be able to examine statistical correlations between range/depth and luminance/chrominance information in the presence of stereoscopic fixation, we also convert ground-truth range maps into disparity maps under a parallel-viewing model. Fig. 2 depicts the geometry of the parallel-viewing model. The disparity values are computed as:

$$\frac{d_p}{f_c} = \frac{d_{io}}{R} \Rightarrow d_p = f_c \frac{d_{io}}{R} \quad (3)$$

where  $d_p$  is disparity,  $f_c$  is the focal length of the camera,  $d_{io}$  is the inter-ocular distance, and  $R$  is ground-truth range. Finally, the converted disparity maps are decomposed by the same multi-scale, multi-orientation 2D Gabor filter bank.

## IV. STATISTICAL ANALYSIS

The statistics of 2D and 3D natural scenes have previously been learned by the human visual apparatus over the eons.

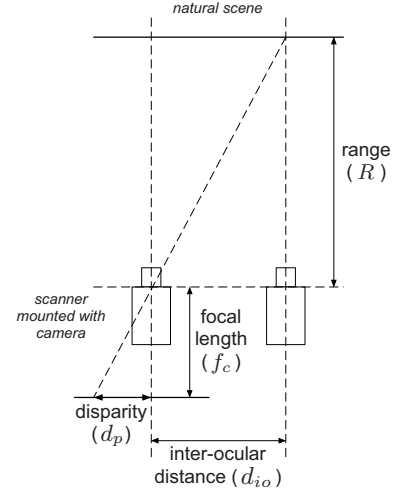


Fig. 2. Geometry of the parallel-viewing model.

These powerful, physically and perceptually relevant constraints form priors which can be applied to solve visual tasks. Since acquiring geometric knowledge about the surrounding 3D environment is a basic element of human visual activity, accurate perception and consistent interpretation of natural range/depth information is an essential processing role of the early visual processing pathway.

Towards understanding the statistical basis of such computations, we first examine the marginal statistics of luminance and chrominance Gabor responses, and the conditional statistics of range gradients given measurements of these responses. Our analysis is performed on the (demodulated) magnitude responses of the Gabor quadrature functions, expressed as root mean square (RMS) values of the sine and cosine responses [8]. Based on these measurements, we form models of the prior marginal and conditional distributions towards leveraging them in solving model-based visual processing problems.

### A. Marginal Statistics

All color images were first transformed into the perceptually relevant CIELAB color space, then decomposed by a multi-scale, multi-orientation 2D Gabor filter bank. This serves the dual role of supplying optimally conjoint spatio-spectral decompositions of the data, while also providing a reasonable approximation of area V1 responses. As a first point of study, we computed the mean and standard deviations of the luminance and chrominance Gabor magnitude responses against spatial frequency and orientation. Specifically, we found the Gabor magnitude responses on the  $L^*$  channel for all 12 test images in the LIVE Color+3D Database. We plot mean and standard deviation of these responses as a function of spatial frequency and orientation in Figs. 3(a) and (b). The same computation was performed on the  $a^*$  and  $b^*$  channels as well, and is plotted in Figs. 3(c)–(f).

As generally expected, mean magnitude responses fall off approximately as  $1/f^2$  with spatial frequency  $f$ . This agrees with findings that the power spectra of natural images varies

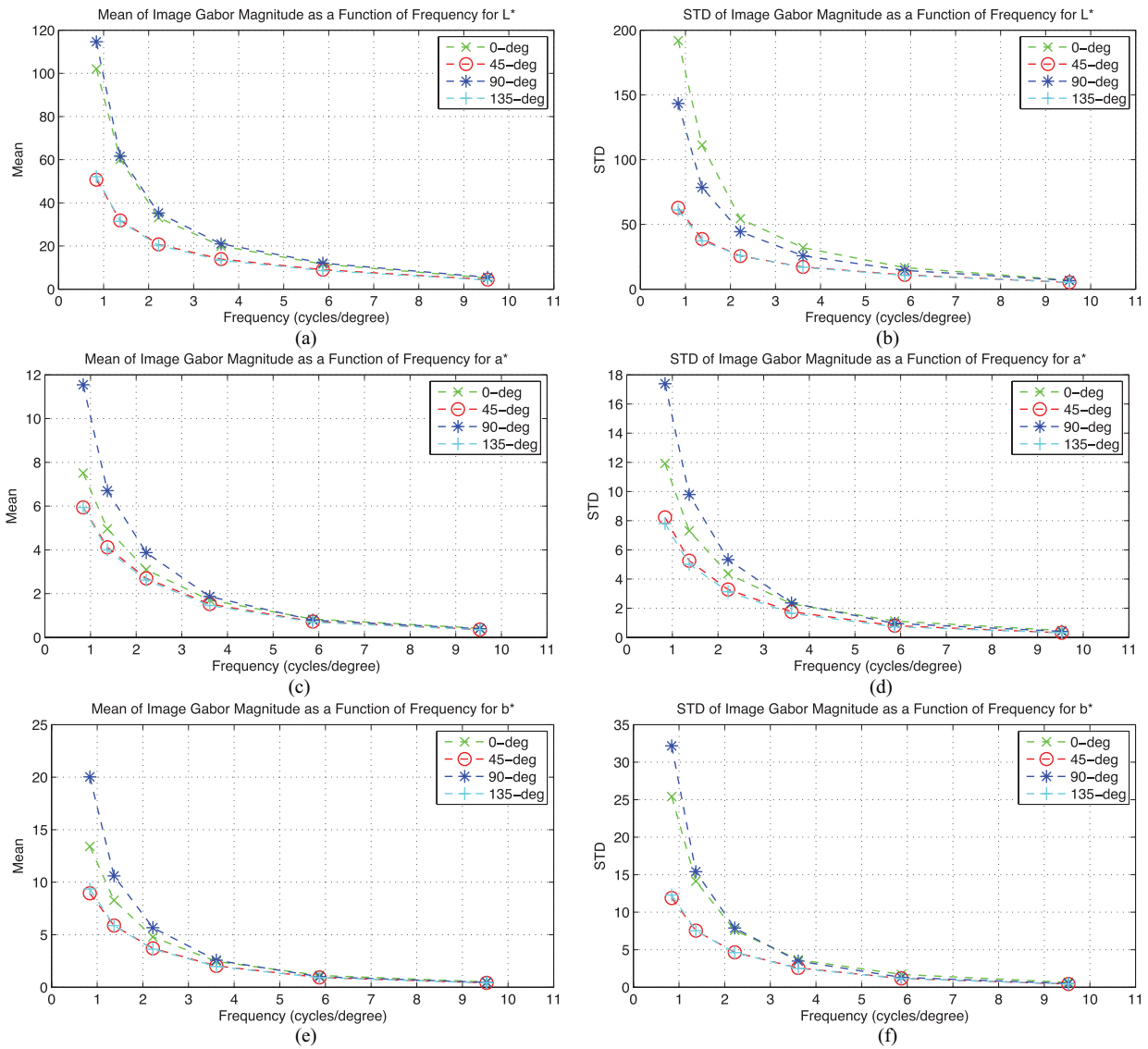


Fig. 3. Mean and standard deviation (STD) of Gabor magnitude responses against radial spatial frequency (a) mean for L\*, (b) STD for L\*, (c) mean for a\*, (d) STD for a\*, (e) mean for b\*, and (f) STD for b\*.

as  $1/f^2$  with spatial frequency [8], [9], and [52]. This fundamental model has been successfully used to explain and predict certain early stages of visual processing in insects and higher vertebrates [14], [53]–[55]. An interesting observation is that the curves for the diagonal-45 and diagonal-135 orientations nearly overlap in all three channels. For the L\* channel, the curves for the horizontal and vertical orientations also overlap, while for the a\* and b\* channels, the curves are distinct.

The standard deviations of the magnitude responses also follows a  $1/f^2$  shape for both luminance and chrominance channels. As observed for the mean magnitude responses, the curves for diagonal orientations overlap across all three channels. However, the standard deviation curves for horizontal and vertical orientations are distinct for both luminance and chrominance channels.

The  $1/f^2$  distribution of mean and standard deviation of the Gabor responses implies equal energy within equal (octave) bandwidths, and also equal variation of energy within equal

bands over different orientations in natural images. Moreover, the distribution of spectral energy contained in the luminance channel is different from that carried by the chrominance channels in natural environments. These findings can potentially be utilized to better understand and explain various stages of visual processing.

### B. Conditional Statistics

It may be observed from natural scenes that there is substantial co-occurrence of luminance/chrominance edges and range/depth discontinuities. For example, if there is a discontinuity in the range/depth map, it is highly likely that an edge of the same orientation is co-located in the corresponding color image. However, the intuition of the other direction is rather weaker; many image edges must exist without corresponding discontinuities in range due to the plethora of shadows and textures in the natural environment. In order to examine the



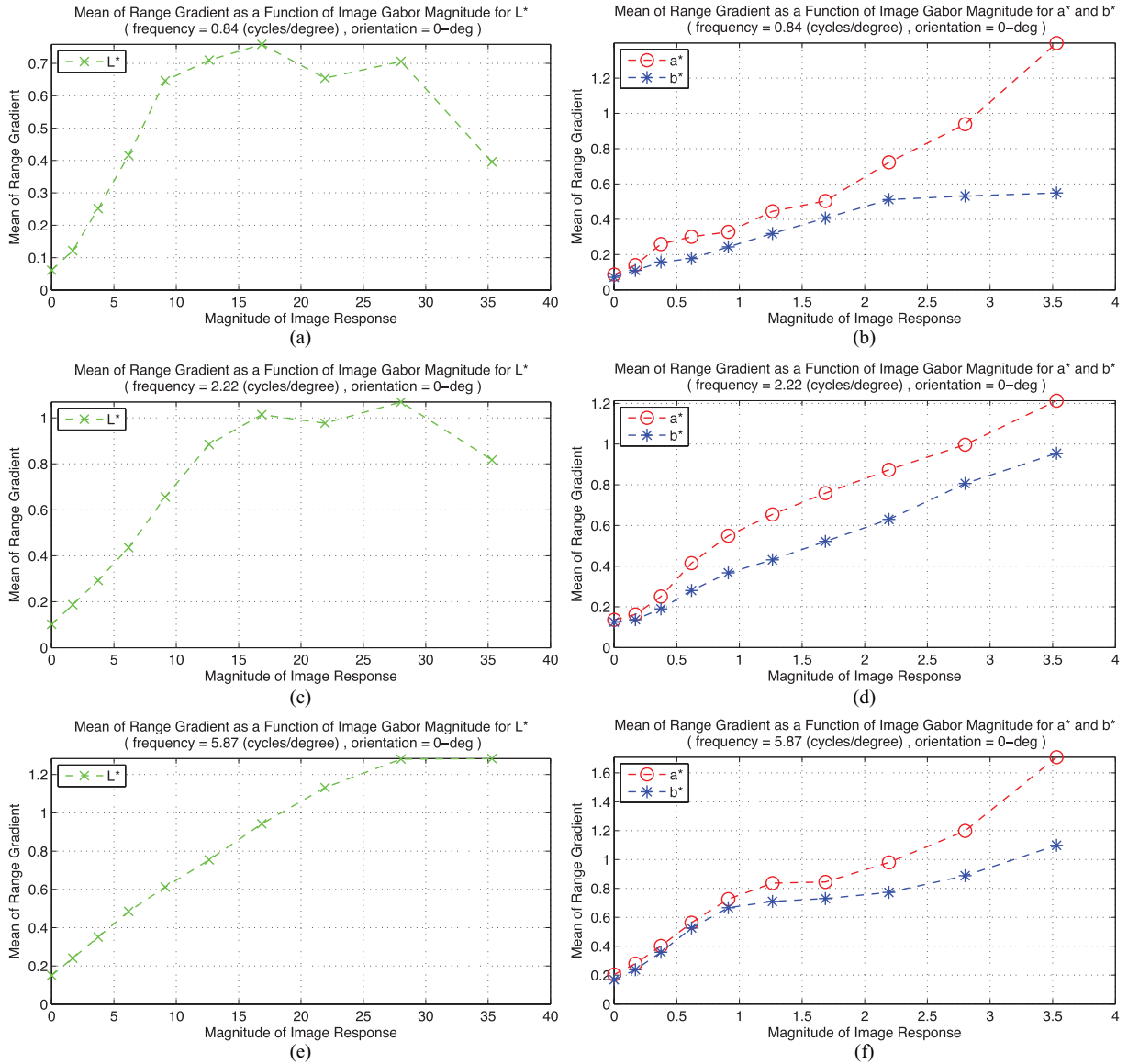


Fig. 4. Means of range gradient magnitudes against Gabor magnitude responses over different spatial frequencies with the same horizontal (0-deg) orientation (a), (c), and (e) for  $L^*$  and (b), (d), and (f) for  $a^*$  and  $b^*$ .

relationship between chromatic image and range singularities, we studied the conditional statistics of range gradients given Gabor image responses.

Again, all color images were first transformed into the perceptually relevant CIELAB color space, then passed through the multi-scale, multi-orientation 2D Gabor filter bank, and the magnitude responses computed. We also computed the gradient magnitude matrix,  $R_g$ , of each range map,  $R$ , for each corresponding color image in the database, which is given by

$$R_g(i, j) = \|\nabla R(i, j)\| = \sqrt{\left(\frac{\partial R(i, j)}{\partial x}\right)^2 + \left(\frac{\partial R(i, j)}{\partial y}\right)^2} \quad (4)$$

where

$$\nabla R(i, j) = \left[ \frac{\partial R(i, j)}{\partial x}, \frac{\partial R(i, j)}{\partial y} \right]^T$$

$$\frac{\partial R(i, j)}{\partial x} = \frac{R(i+1, j) - R(i-1, j)}{2}$$

$$\frac{\partial R(i, j)}{\partial y} = \frac{R(i, j+1) - R(i, j-1)}{2}. \quad (5)$$

To obtain conditional statistics, we binned the magnitude responses across all images for each luminance and chrominance channel. Within each bin, we also collected the corresponding gradient magnitudes for all range maps. Finally, we computed the conditional mean and standard deviations of magnitude of the range gradients given the Gabor magnitude responses for the  $L^*$ ,  $a^*$ , and  $b^*$  channels. Figs. 4 and 5 plot the mean and standard deviation, respectively, of the range gradient magnitudes against Gabor magnitude responses for luminance and chrominance channels for an exemplar sub-band. Very similar curves and results are observed at different sub-bands.

The six panels in Fig. 4 plot the conditional statistics of mean range gradient magnitude given magnitude responses of

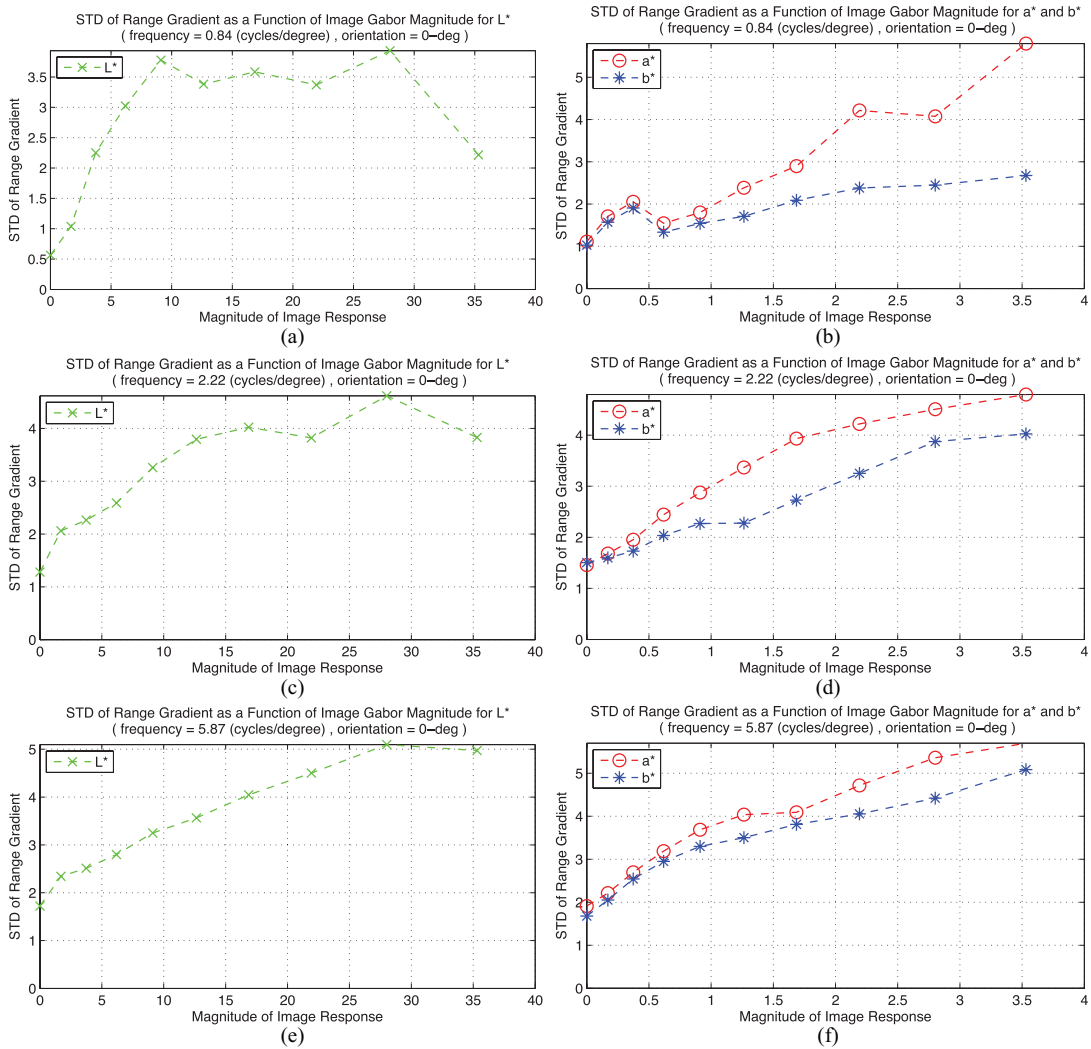


Fig. 5. Standard deviations of range gradient magnitudes against Gabor magnitude responses over different spatial frequencies with the same horizontal (0-deg) orientation (a), (c), and (e) for  $L^*$  and (b), (d), and (f) for  $a^*$  and  $b^*$ .

horizontal Gabors at three different spatial frequencies. For the luminance channel, the range gradient magnitude increases monotonically with small Gabor responses, but saturates with larger Gabor responses over all frequencies. On the other hand, the magnitude of range gradients increases monotonically with Gabor magnitude responses at all frequencies for both chromatic channels, and there is no saturation for large Gabor responses. Similar trends are observed for the standard deviation of range gradients conditioned on the magnitude of horizontal Gabor responses at three different spatial frequencies, as shown in Fig. 5. These monotonic relationships between range gradients and luminance and chrominance Gabor responses demonstrate a high correlation between these quantities, while also strengthening the intuition that, if there are strong variations in a natural image, i.e. large Gabor responses, there is a high likelihood of co-located large variations, i.e. large range gradients, in the corresponding range map. Moreover, the luminance channel carries information that is different from that carried by the chrominance channels in the sense that the means and standard deviations of range gradient magnitudes both saturate given large luminance

Gabor magnitude responses, which implies that the chromatic components in natural images can also possibly be utilized in depth perception, a concept that is supported by our prior human study [30].

To further validate the existence of strong correlations between range gradients and image Gabor responses, we performed a simple hypothesis test on the sample correlation coefficients between range gradients and luminance/chrominance Gabor responses using the same sub-band as in Figs. 4(c) and 5(c). The  $t$ -score is given by

$$t = \frac{r\sqrt{n-2}}{1-r^2} \quad (6)$$

where  $r$  is the sample correlation coefficient and  $n$  is the number of samples. Since there are millions of points within each sub-band, we iterated 100 times, taking 1000 random samples per iteration to compute the correlation coefficient. The final  $t$ -score was obtained by finding the average correlation coefficient over 100 iterations with the sample size  $n$  equal to 1000. Table I lists the  $t$ -scores of the color channels and the corresponding decisions using a two-sided level of significance

TABLE I

HYPOTHESIS TEST OF CORRELATION ON CONDITIONAL STATISTICS  
[FREQUENCY = 5.87 (CYCLES/DEGREE) AND ORIENTATION = 0-DEG]

Channel	t-score	p-value ( $10^{-3}$ )	Decision
L*	3.342	0.205	Reject $H_0$
a*	2.933	7.681	Reject $H_0$
b*	3.343	0.357	Reject $H_0$

$\alpha = 0.05$ . It can be seen from Table I that the null hypothesis,  $H_0$ , that there is no correlation between range gradients and image Gabor responses was rejected for all three channels. Very similar results were obtained for all other sub-bands. Note that this corresponds to a very conservative test. Testing the full data set would yield minuscule  $p$ -values because of the large  $n$ . Rather, we did the test using approximately the number of samples that would be available instantaneously over about  $0.005$  ( $\text{mm}^2$ ) on the primate retina [or about  $0.056$  ( $\text{deg}^2$ ) of visual angle].

## V. STATISTICAL MODELS

The statistical analysis described thus far discovers a link between range/depth variations and co-located luminance and chrominance variations in natural images, and by extension, in the neural responses of luminance/chrominance in primary visual cortex (V1). In this section, we seek to quantitatively model the statistical relationships between co-located luminance/chrominance and range/depth information in natural images. The Gabor magnitude responses were computed as in the previous analysis on both luminance and chrominance channels. To acquire statistics of the important stereoscopic cue, disparity, used in the perception of depth by the human vision system, the ground-truth range maps were converted into disparity maps using the parallel viewing model described in Section III-B. These converted disparity maps were also subjected to a multi-scale, multi-orientation Gabor decomposition, from which the disparity Gabor magnitude responses were computed. Since we want to derive statistical models that relate disparity and color image data, and demonstrate their usefulness, the marginal distributions of the Gabor magnitude responses to luminance/chrominance and disparity, and the conditional distributions of luminance/chrominance given disparity Gabor magnitude responses are of great interest.

### A. Marginal Distributions

In order to examine the marginal distributions of luminance, chrominance, and disparity processed by different sub-bands, the Gabor responses were first collected across all scenes in the database. The empirical marginal distributions for all quantities of interest within each sub-band were obtained as histograms computed by binning all of the Gabor magnitude responses within each channel at that sub-band.

Fig. 6 shows the marginal distributions of the luminance, chrominance, and disparity Gabor magnitude responses for one sub-band. The circle-dotted, square-dotted, cross-dotted, and triangle-dotted lines depict the best (least-squares) generalized log-normal, Weibull, Rayleigh, and exponential distribution

fits, respectively, to each marginal distribution. In particular, the generalized log-normal distribution is given by

$$p_g(x) = \begin{cases} \frac{\beta_g}{2x\alpha_g\Gamma\left(\frac{1}{\beta_g}\right)} \exp\left[-\left(\frac{|\ln(x)-\mu_g|}{\alpha_g}\right)^{\beta_g}\right], & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (7)$$

where  $\Gamma(\cdot)$  is the gamma function,  $\mu_g$ ,  $\alpha_g$ , and  $\beta_g$  are the location, scale, and shape parameters, respectively. The general Weibull distribution is given by

$$p_w(x) = \begin{cases} \frac{\beta_w}{\alpha_w} \left(\frac{x}{\alpha_w}\right)^{(\beta_w-1)} e^{-\left(\frac{x}{\alpha_w}\right)^{\beta_w}}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (8)$$

where  $\alpha_w$  and  $\beta_w$  represent the scale and shape parameters, respectively, which allows the model to include exponential ( $\beta_w = 1$ ) and Rayleigh ( $\beta_w = 2$ ) distributions as special cases depending on the shape parameter. The characteristic shape of the marginal distributions is quite different from the symmetric Gaussian-like distributions used in other studies. These previous models have captured the statistics of the luminance channel of natural images using band-pass wavelet filter-banks, e.g., steerable pyramid decompositions, but without finding the magnitude (envelope) responses. Here we have used the Gabor filter-bank to match the receptive fields of simple neurons in primary visual cortex (V1), and computed the Gabor magnitude responses to mimic the energy exchange in neural signal communication. It can be seen that the generalized log-normal fits better capture the shapes of all four marginal distributions, while the Weibull and exponential fits are not able to match the positive-skewed bell shapes, and the Rayleigh fits fail to model the heavy tails. In other words, with the three model parameters, location, scale and shape, the generalized log-normal function can flexibly adjust both its peak location and its variance, i.e., the width of distribution, to better fit the characteristic shape of the marginal distribution of image and disparity magnitude responses. We also performed a numerical comparison of different distribution fits at the same sub-band using the sum of squared error, as shown in Table II. In accordance with the visual comparison in Fig. 6, the generalized log-normal functions yield the best fits among all four marginal distributions, i.e. fixing the parameters gives much worse fits. Note that the marginal distributions of the filtered luminance, chrominance, and disparity over different sub-bands all share similar shapes. For reference, we list the best-fit generalized log-normal parameters for the marginal distributions of luminance, chrominance, and disparity Gabor magnitudes at all sub-bands in [28].

### B. Conditional Distributions

Similarly, the conditional distributions of luminance and chrominance given disparity at different sub-bands were obtained by first computing and collecting the filtered luminance, chrominance, and disparity magnitude responses across all scenes, then computing the histograms for each sub-band. For each sub-band, the conditional histograms of luminance and chrominance given disparity were computed by first binning the disparity magnitude responses, then binning the



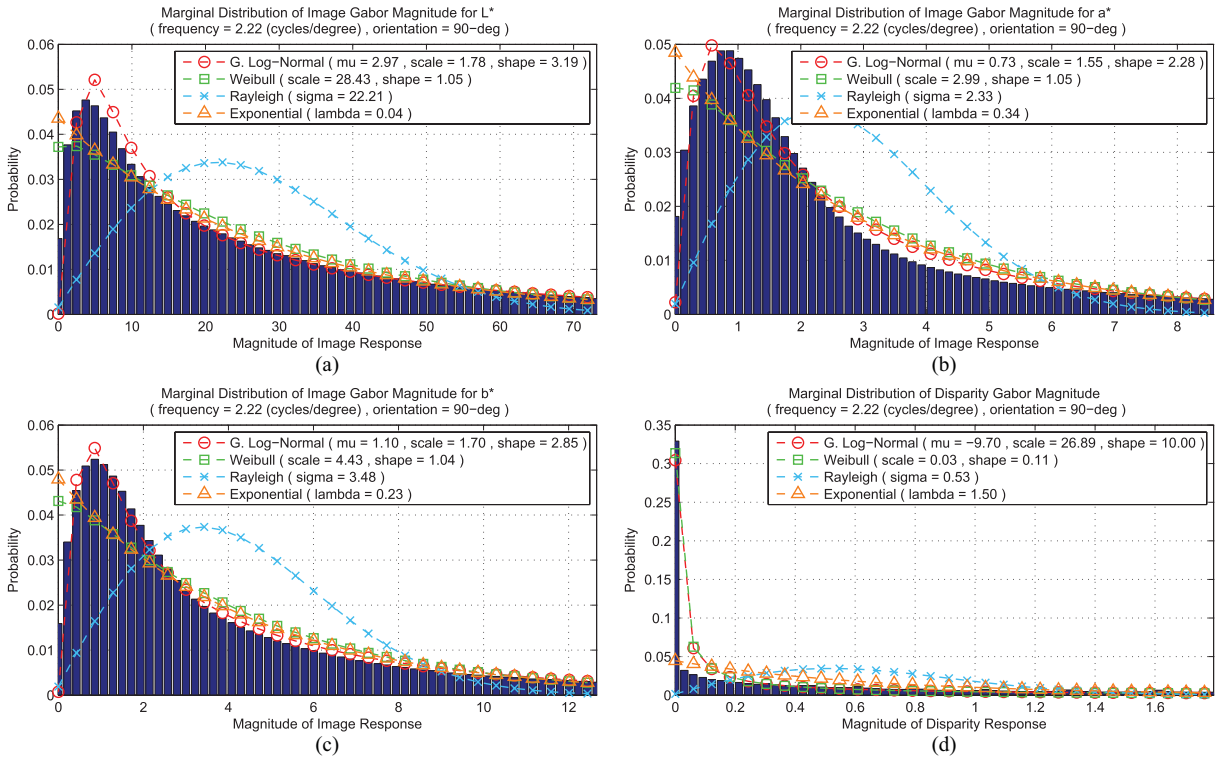


Fig. 6. Marginal distributions of Gabor magnitude responses at one sub-band for (a)  $L^*$ , (b)  $a^*$ , (c)  $b^*$ , and (d) disparity.

TABLE II  
COMPARISON OF MARGINAL DISTRIBUTION FITS USING SUM OF SQUARED ERROR ( $10^{-2}$ )

	Marginal Distribution			
	$L^*$	$a^*$	$b^*$	Disparity
Exponential fit	0.116	0.245	0.222	8.569
Rayleigh fit	1.143	1.141	1.292	12.008
Weibull fit	0.103	0.202	0.197	0.823
G. Log-normal fit	<b>0.078</b>	<b>0.057</b>	<b>0.041</b>	<b>0.785</b>

filtered luminance/chrominance magnitude responses within each disparity bin. Fig. 7 shows the conditional distributions of all three luminance and chrominance components (solid lines), as well as their corresponding best-fit generalized log-normal distributions (dotted lines) and their model parameters. It can be seen that the conditional distributions of luminance and chrominance given disparity are well-fitted by the generalized log-normal distribution. As discussed in Section V-A, the flexible generalized log-normal function is a better model than other specific fits, such as Weibull, exponential, and Rayleigh functions. For the conditional distributions of luminance given disparity, the location parameter ( $\mu_g$ ) of the fitted generalized log-normal model increases monotonically and linearly as the disparity magnitude response increases, while both the scale ( $\alpha_g$ ) and shape ( $\beta_g$ ) parameters decrease monotonically and linearly across the disparity magnitude responses. On the other hand, for the conditional distributions of chrominance given disparity, all three parameters exhibit a nearly linear relationship with the disparity magnitude

responses. In general, as the disparity magnitude response increases, the conditional distributions of both luminance and chrominance become more heavy-tailed, which implies that if there is a large disparity variation, i.e. a large range/depth discontinuity, large luminance and chrominance variations are highly likely to be co-located in the corresponding color images. These monotonic correlations between the Gabor responses of disparity and luminance/chrominance in natural images confirm the observations as well as the computed conditional statistics between range and luminance/chrominance variations discussed in Section IV-B. Moreover, the linear relationships between the parameters of the generalized log-normal model and the magnitude of disparity Gabor responses nicely captures the heavy-tailed conditional distributions of both filtered luminance and chrominance channels. Next, we will leverage these new joint NSS models to solve an exemplar 3D image processing problem: binocular correspondence.

## VI. APPLICATION TO A CHROMATIC BAYESIAN STEREO ALGORITHM

Given a pair of left and right images, a binocular stereo algorithm computes a disparity map from one image to the other. The basic idea is to minimize an energy functional which captures differential binocular cues between left and right images within an optimization framework [56]. A Bayesian stereo algorithm is able to adapt a likelihood (conditional distribution) and a prior (marginal distribution) of natural scene statistics (NSS) within the energy functional to be minimized, thus forcing the solution to be consistent with

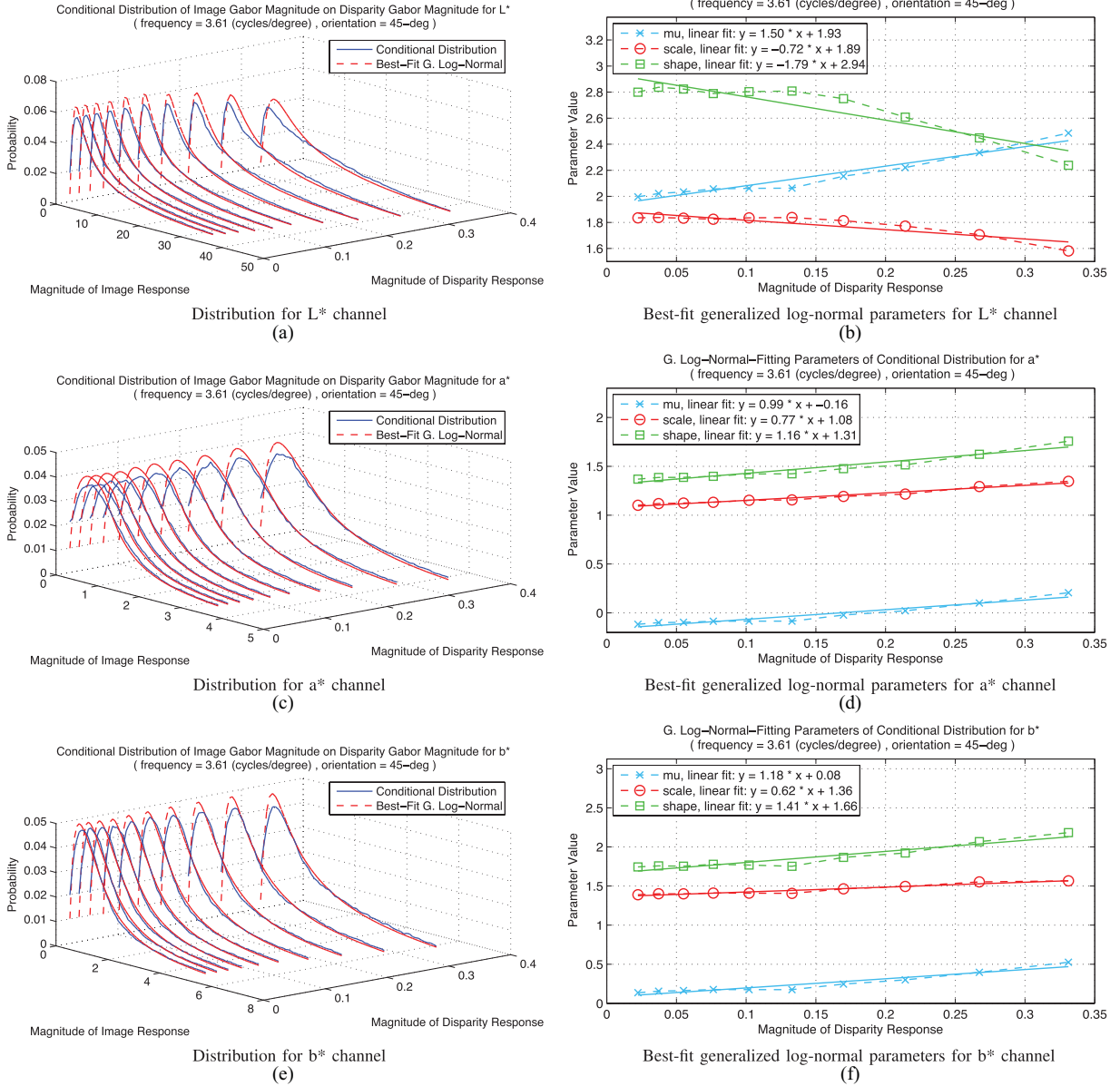


Fig. 7. Conditional distributions of filtered luminance ( $L^*$ ) and chrominance ( $a^*$  and  $b^*$ ) magnitudes given filtered disparity magnitudes from one sub-band (a), (c), and (e) conditional distributions (solid lines) along with the best-fit generalized log-normal models (dotted lines), (b), (d), and (f) their corresponding best-fit generalized log-normal parameters.

the observed statistical relationships that occur between luminance, chrominance, and disparity data in natural scenes, as derived in Section V. Given a pair of left and right images,  $I_l$  and  $I_r$ , then to estimate the disparity map,  $D$ , from the right (matching) to the left (reference) image, the canonical Bayesian stereo formulation takes the form [57]

$$\begin{aligned} D &= \underset{D'}{\operatorname{argmax}} P(D'|(I_l, I_r)) \\ &= \underset{D'}{\operatorname{argmax}} P((I_l, I_r)|D')P(D') \end{aligned} \quad (9)$$

where  $P(D'|(I_l, I_r))$  is the posterior probability to be maximized, and  $P((I_l, I_r)|D')$  and  $P(D')$  are the likelihood and prior probabilities, respectively. Taking the logarithm of the product of the likelihood and prior, the Bayesian formulation

corresponds to minimization of the energy function:

$$D = \underset{D'}{\operatorname{argmin}} E_p + \lambda E_s \quad (10)$$

where  $E_p$  is the photometric energy expressed by the likelihood  $P((I_l, I_r)|D')$ ,  $E_s$  is a smoothness term derived from the prior  $P(D')$ , and  $\lambda$  is a weight. Note that  $E_p$  can encapsulate all three luminance and chrominance components,  $L^*$ ,  $a^*$ , and  $b^*$ , and be written

$$E_p = \sum_{i,j} \sum_{k \in \{L^*, a^*, b^*\}} |I_{l_k}(i, (j - D'(i, j))) - I_{r_k}(i, j)|. \quad (11)$$

To incorporate the marginal and conditional NSS distributions that we have measured and modeled, the Bayesian stereo

formulation can be re-written as

$$\begin{aligned} D &= \operatorname{argmax}_{D'} P(\tilde{D}' | (I_l, I_r), \tilde{I}_l) \\ &= \operatorname{argmax}_{D'} P((I_l, I_r) | \tilde{D}', \tilde{I}_l) P(\tilde{I}_l | \tilde{D}') P(\tilde{D}') \quad (12) \\ &= \operatorname{argmin}_{D'} E_p + \lambda (E_{NSS_c} + E_{NSS_m}) \quad (13) \end{aligned}$$

(taking logarithm of (12)), where  $\tilde{I}_l$  and  $\tilde{D}'$  are the magnitudes of the Gabor filtered responses of  $I_l$  and  $D'$ , respectively,  $E_p$  is the photometric energy derived from  $P((I_l, I_r) | \tilde{D}', \tilde{I}_l)$ ,  $E_{NSS_c}$  and  $E_{NSS_m}$  are energy terms related to the conditional and marginal NSS distributions, respectively, and  $\lambda$  is the constant weight.

Finally, since both the marginal distribution of disparity and the conditional distributions of luminance and chrominance given disparity can be modeled as generalized log-normal, the complete formulation of the proposed Bayesian stereo algorithm incorporating the NSS models can be written

$$D = \operatorname{argmin}_{D'} \sum_{i,j} \left[ \sum_{k \in \{L^*, a^*, b^*\}} (E_{p,k} + \lambda_k E_{NSS_{c,k}}) + \lambda_m E_{NSS_m} \right] \quad (14)$$

where by introducing Eqs. (7) and (11)

$$E_{p,k} = |I_k(i, (j - D'(i, j))) - I_{rk}(i, j)| \quad (15)$$

$$\begin{aligned} E_{NSS_{c,k}} &= \ln(\tilde{I}_k(i, j)) + \ln\left(\frac{2\alpha_k \Gamma(\frac{1}{\beta_k})}{\beta_k}\right) \\ &\quad + \left(\frac{|\ln(\tilde{I}_k(i, j)) - \mu_k|}{\alpha_k}\right)^{\beta_k} \quad (16) \end{aligned}$$

$$\begin{aligned} E_{NSS_m} &= \ln(\tilde{D}'(i, j)) + \ln\left(\frac{2\alpha_{\tilde{D}'} \Gamma(\frac{1}{\beta_{\tilde{D}'}})}{\beta_{\tilde{D}'}}\right) \\ &\quad + \left(\frac{|\ln(\tilde{D}'(i, j)) - \mu_{\tilde{D}'}|}{\alpha_{\tilde{D}'}}\right)^{\beta_{\tilde{D}'}} \quad (17) \end{aligned}$$

where  $\mu_k$ ,  $\alpha_k$ , and  $\beta_k$  are the location, scale, and shape parameters, respectively, of the best-fit generalized log-normal distributions of filtered luminance and chrominance conditioned on filtered disparity,  $\mu_{\tilde{D}'}$ ,  $\alpha_{\tilde{D}'}$ , and  $\beta_{\tilde{D}'}$  are the location, scale, and shape parameters of the best-fit generalized log-normal distribution of filtered disparity, respectively, and  $\lambda_k$  and  $\lambda_m$  are their corresponding constant weights. Note that the three parameters,  $\mu_k$ ,  $\alpha_k$  and  $\beta_k$ , can be further linearly modeled with the disparity Gabor magnitudes, as illustrated in Fig. 7:

$$\mu_k = m_{\mu,k} \tilde{D}' + b_{\mu,k} \quad (18)$$

$$\alpha_k = m_{\alpha,k} \tilde{D}' + b_{\alpha,k} \quad (19)$$

$$\beta_k = m_{\beta,k} \tilde{D}' + b_{\beta,k} \quad (20)$$

where  $m_{\mu,k}$ ,  $m_{\alpha,k}$ , and  $m_{\beta,k}$  are the slope parameters for  $\mu_k$ ,  $\alpha_k$ , and  $\beta_k$ , respectively, and  $b_{\mu,k}$ ,  $b_{\alpha,k}$ , and  $b_{\beta,k}$  are the corresponding offset parameters. To solve the optimization of the proposed Bayesian stereo algorithm, we apply simulated annealing on the derived energy function (14) [58].

## VII. SIMULATION RESULTS

We simulate and evaluate the proposed Bayesian stereo algorithm utilizing the derived NSS models on stereo image pairs from the widely used Middlebury database [56]. To demonstrate the effectiveness of the derived statistical models relating luminance/chrominance and disparity in natural scenes, we compared the computed disparity maps using the Bayesian stereo algorithm with related formulations and models, including the canonical formulation using (10), the NSS model proposed in [26], and the proposed luminance-chrominance-range NSS model (14). In [26], the authors derived an NSS model using only luminance information in the wavelet domain, and incorporated only the conditional distribution of luminance given disparity into the Bayesian stereo algorithm. Using the proposed luminance-chrominance-range NSS model, we implement the Bayesian stereo algorithm using two formulations: one includes only the luminance component (L\*), while the other includes both luminance and chrominance (a\* and b\*) components.

### A. Comparison With Previous Models

Figs. 8–11 show simulation results of the four stereo image pairs, *Tsukuba*, *Venus*, *Cones*, and *Teddy*, from the Middlebury database, including the original left and right images, the ground-truth disparity map, and the computed disparity maps obtained by the three different Bayesian formulations. Generally, computed disparity maps delivered by the stereo model embodying both luminance (L\*) and chrominance (a\* and b\*) NSS priors are very close to the corresponding ground-truth disparity maps, retaining more details than the canonical formulation, and better adherence to smooth regions than the ones computed by the previous NSS model. On *Tsukuba*, the canonical formulation scrubs regions, e.g., around the camera, while the algorithm using the previous luminance-only NSS model tends to “over-segment.” The proposed algorithm using luminance-chrominance-range NSS priors is better able to find a balance between disparity smoothness and 3D detail with the aid of the additional regularity supplied by modeling the disparity and luminance/chrominance channels. For *Venus* and *Teddy*, it can be seen that the canonical formulation fails to find binocular correspondences in some smooth regions without edges, while the previous luminance-only NSS model is able to solve those disparity ambiguities using luminance and disparity priors. By introducing conditional priors of chrominance given depth, the luminance-chrominance-range NSS model further improves the accuracy of the computed disparity map by cleaning up the smooth 3D surfaces. On *Cones*, both the canonical formulation and the previous luminance-only NSS model do a good job matching image details while maintaining disparity smoothness; yet, they are not able to find binocular correspondences around some of the occluded regions on the cones. However, the luminance-chrominance-range NSS model allows most of the binocular correspondences around those occluded regions to be successfully resolved.

In addition to visual comparison, we also conducted a quantitative evaluation to compare the performance of the stereo

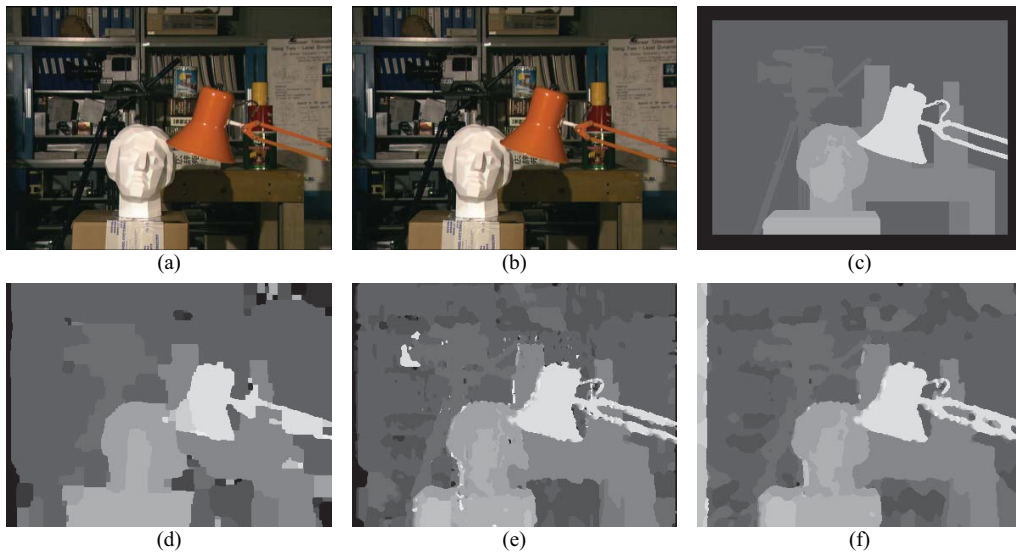


Fig. 8. Simulation results on *Tsukuba* from the Middlebury database, including the original stereo image pair, the ground-truth disparity map, and disparity maps using computed Bayesian stereopsis under different NSS models. (a) Left image. (b) Right image. (c) Ground-truth disparity map. (d) Canonical formulation. (e) Previous NSS model. (f) Proposed luminance-chrominance-range NSS model with  $L^*$ ,  $a^*$ , and  $b^*$ .

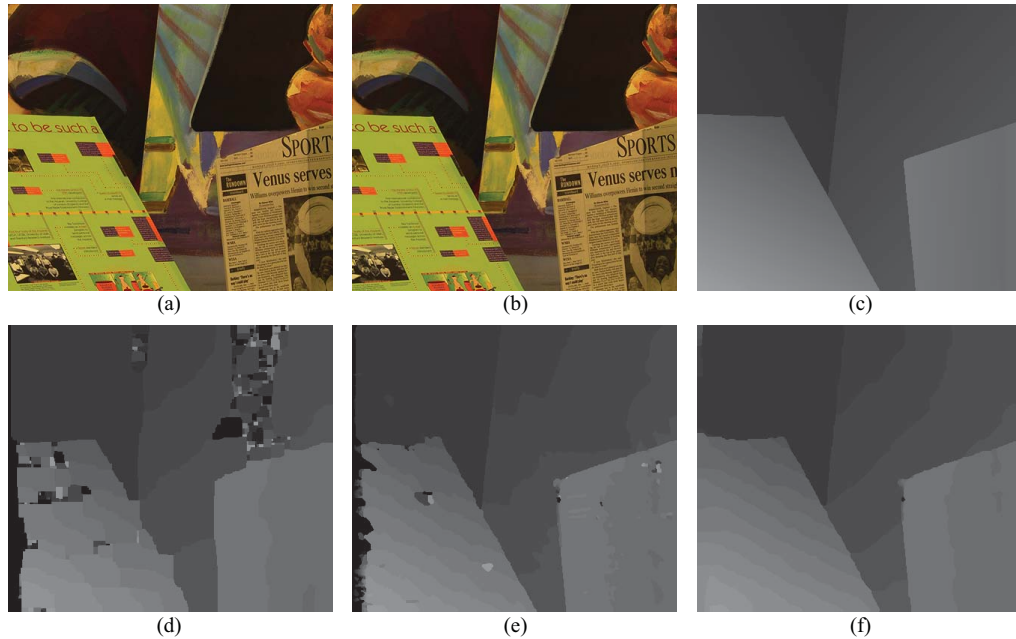


Fig. 9. Simulation results on *Venus* from the Middlebury database, including the original stereo image pair, the ground-truth disparity map, and disparity maps using computed Bayesian stereopsis under different NSS models. (a) Left image. (b) Right image. (c) Ground-truth disparity map. (d) Canonical formulation. (e) Previous NSS model. (f) Proposed luminance-chrominance-range NSS model with  $L^*$ ,  $a^*$ , and  $b^*$ .

algorithm embodying the derived luminance-chrominance-range NSS model with the canonical formulation and the previous luminance-only NSS model. Tables III through V give numerical comparisons between the proposed model and the other two formulations using three different metrics: bad-pixel percentage, including overall, non-occluded, and textured, for all four test image pairs.

Bad-pixel percentage,  $P_{bp}$ , is a commonly used error metric to measure pixel-wise differences between computed and ground-truth depth maps [56], [59]. It takes the form:

$$P_{bp} = \frac{1}{N_S} \sum_{i,j \in S} (|D_C(i, j) - D_G(i, j)| > \delta_D) \quad (21)$$

where  $D_C$  and  $D_G$  are the computed and ground-truth disparity maps, respectively,  $S$  is the image region over which  $P_{bp}$  is calculated,  $N_S$  is the number of pixels in  $S$ , and  $\delta_D$  is a threshold expressing disparity error tolerance. Here we use  $\delta_D = 1.0$ , which coincides with previously published work comparing stereo algorithms [60], [61]. The three metrics of bad-pixel percentage used here are distinguished by the different regions  $S$ . The overall bad-pixel percentage in Table III is calculated over the entire disparity map, i.e.  $S = \{(i, j) \mid 1 \leq i \leq h, 1 \leq j \leq w\}$  where  $h$  and  $w$  are the height and width of the disparity map, respectively. For the non-occluded bad-pixel percentage,  $S$  is defined as the region that is not occluded in the matching image, i.e. pixels appearing in the



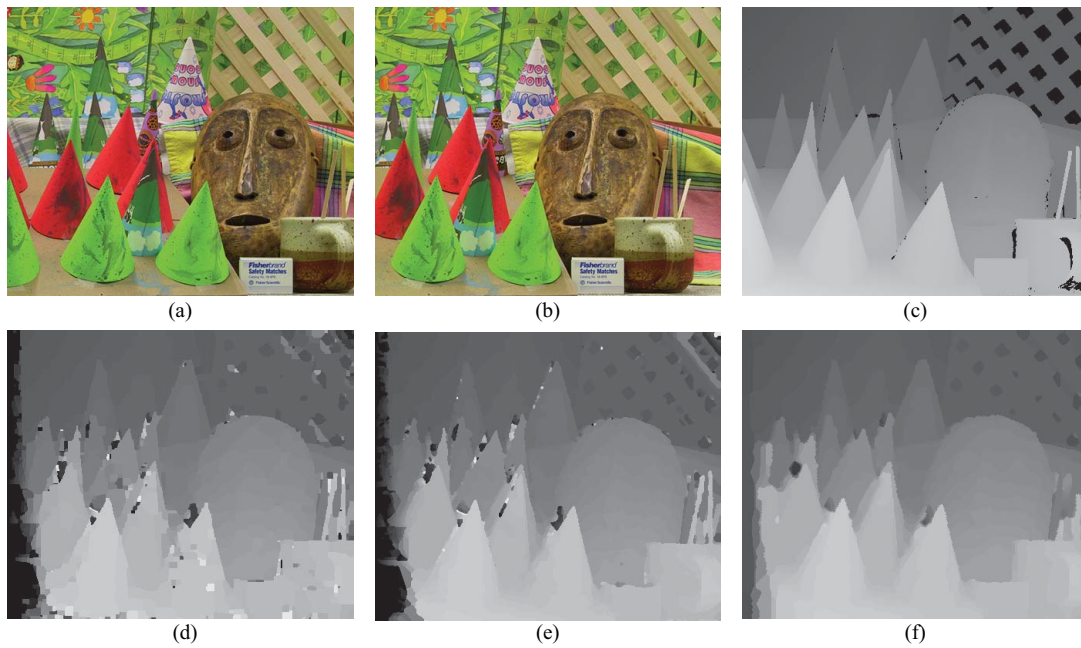


Fig. 10. Simulation results on *Cones* from the Middlebury database, including the original stereo image pair, the ground-truth disparity map, and disparity maps using computed Bayesian stereopsis under different NSS models. (a) Left image. (b) Right image. (c) Ground-truth disparity map. (d) Canonical formulation. (e) Previous NSS model. (f) Proposed luminance-chrominance-range NSS model with  $L^*$ ,  $a^*$ , and  $b^*$ .

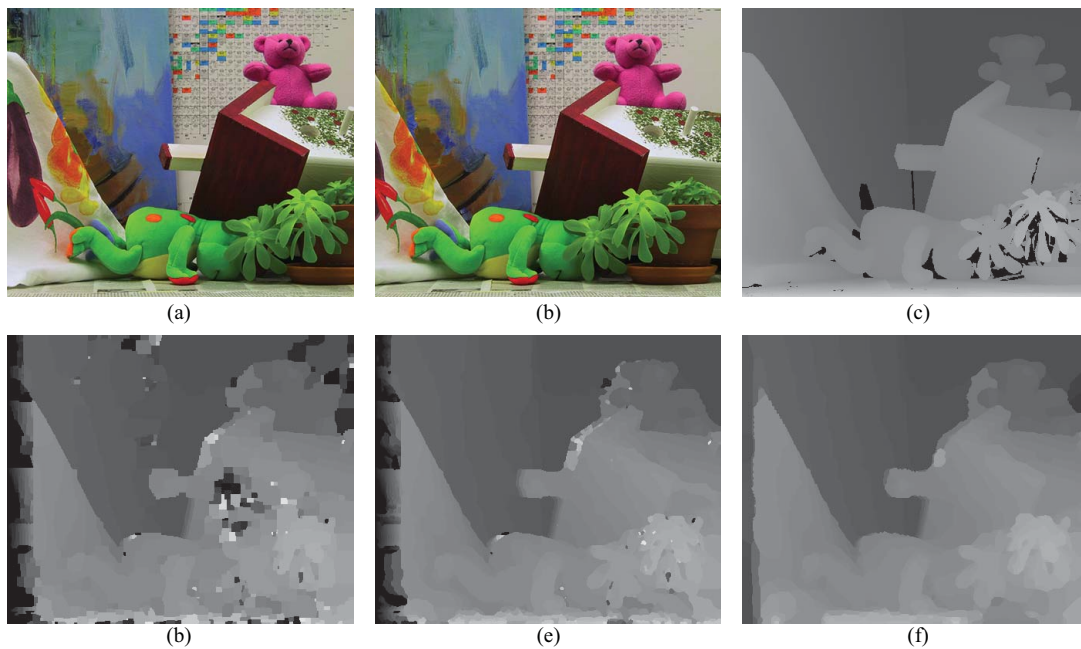


Fig. 11. Simulation results on *Teddy* from the Middlebury database, including the original stereo image pair, the ground-truth disparity map, and disparity maps using computed Bayesian stereopsis under different NSS models. (a) Left image. (b) Right image. (c) Ground-truth disparity map. (d) Canonical formulation. (e) Previous NSS model. (f) Proposed luminance-chrominance-range NSS model with  $L^*$ ,  $a^*$ , and  $b^*$ .

reference image have correspondences in the matching image. Finally, the textured bad-pixel percentage is calculated only over regions where the intensity of image horizontal gradients is beyond some threshold, i.e. pixels belonging to prominent image details, edges, and texture in the reference image.

From Tables III through V, it is apparent that the numerical results support the visual comparisons: the Bayesian stereo

algorithm using the luminance-chrominance-range NSS model outperforms the other two methods in terms of all three different metrics of bad-pixel percentage. Taking *Venus* for example, with respect to all three bad-pixel percentage metrics, the proposed luminance-chrominance-range NSS model achieves more than 100% improvement over the previous NSS model, which, in turn, significantly improves on the canonical



TABLE III  
COMPARISON OF BAYESIAN STEREO ALGORITHM UNDER  
DIFFERENT NATURAL SCENE MODELS USING OVERALL  
BAD-PIXEL PERCENTAGE (%)

	<i>Tsukuba</i>	<i>Venus</i>	<i>Cones</i>	<i>Teddy</i>
Canonical formulation	9.71	11.21	24.31	32.18
Previous NSS model in [26]	6.45	5.34	20.78	23.35
Proposed NSS model using only $L^*$	5.19	2.55	18.86	20.58
Proposed NSS model using $L^*$ , $a^*$ , and $b^*$	<b>4.91</b>	<b>2.21</b>	<b>18.57</b>	<b>20.37</b>

TABLE IV  
COMPARISON OF BAYESIAN STEREO ALGORITHM UNDER DIFFERENT  
NATURAL SCENE MODELS USING NONOCCLUDED  
BAD-PIXEL PERCENTAGE (%)

	<i>Tsukuba</i>	<i>Venus</i>	<i>Cones</i>	<i>Teddy</i>
Canonical formulation	7.78	9.79	12.72	23.34
Previous NSS model in [26]	4.26	3.69	8.54	13.20
Proposed NSS model using only $L^*$	2.92	1.44	7.58	12.37
Proposed NSS model using $L^*$ , $a^*$ , and $b^*$	<b>2.64</b>	<b>1.18</b>	<b>7.35</b>	<b>12.15</b>

TABLE V  
COMPARISON OF BAYESIAN STEREO ALGORITHM UNDER  
DIFFERENT NATURAL SCENE MODELS USING TEXTURED  
BAD-PIXEL PERCENTAGE (%)

	<i>Tsukuba</i>	<i>Venus</i>	<i>Cones</i>	<i>Teddy</i>
Canonical formulation	4.77	3.85	12.49	19.82
Previous NSS model in [26]	4.70	2.96	8.39	12.20
Proposed NSS model using only $L^*$	3.34	1.62	7.61	11.10
Proposed NSS model using $L^*$ , $a^*$ , and $b^*$	<b>3.30</b>	<b>1.41</b>	<b>7.39</b>	<b>10.93</b>

formulation. By observing the textured bad-pixel percentage on *Tsukuba* in Table V, it is apparent that the Bayesian stereo algorithms using the canonical formulation and the previous NSS model deliver similar performance, while the proposed luminance-chrominance-range NSS model yields a bolstered Bayesian stereo algorithm that delivers a significantly more accurate disparity map. Moreover, on the complicated image pairs *Cones* and *Teddy*, while the previous NSS model generates fairly good disparity maps on non-occluded and textured regions, the new NSS models further improve the results with fewer pixel errors, demonstrating the utility of the derived marginal and conditional models that serve to regularize the range/depth and luminance/chrominance statistics of the stereo solution on natural images.

### B. Augmentation by Chrominance

In Tables III–V, we also list numerical results from the proposed luminance-chrominance-range NSS model using only

the luminance channel ( $L^*$ ). It can be seen that for all four image pairs, the results using both the luminance ( $L^*$ ) and chrominance ( $a^*$  and  $b^*$ ) channels yields better performance than using only the luminance channel with respect to all three different performance metrics. For example, the vivid and diverse colored objects in *Teddy* increase the difficulty of finding binocular correspondences; however, the proposed luminance-chrominance-range NSS model is better able to solve the problem by exploiting the derived conditional model between the natural depth and chrominance channels, resulting in more accurate disparity maps with lower non-occluded and textured bad-pixel percentages. Based on this quantitative comparison, we may conclude that chromatic information not only augments the performance of Bayesian stereo algorithms, but could also play a useful role in human binocular visual perception. For example, stereo processing in human vision systems could possibly leverage the statistical relationship between chrominance and range/depth cues in natural images to augment a variety of 3D visual tasks [30]–[32].

## VIII. CONCLUSION

By utilizing high-resolution, high-quality color images and co-registered range maps in the LIVE Color+3D Database, we examined the statistical relationships between multi-scale, multi-orientation Gabor decompositions of luminance/chrominance and range/depth data in natural scenes. We showed that the marginal statistics of both image and range magnitude responses follows the well-known  $1/f^2$  power law, and the conditional statistics of range gradients given image magnitude responses provide evidences supporting the co-occurrence of natural image and range variations. We further derived marginal and conditional priors relating natural luminance/chrominance and disparity, and demonstrated their efficacy with application to the Bayesian stereo algorithm. We also demonstrated that including the chrominance-range models augments the performance of the Bayesian stereo algorithm over using only the luminance information. More importantly, the superior performance incorporating color and range priors to previous luminance-only models bolsters the psychophysical evidence that not only image intensity, but also chromatic information is useful in 3D visual processing.

The statistical analysis we have performed and the color-range priors we have derived in this paper yield insight into how 3D structures in the environment might be recovered from color image data. We believe that these fundamental regularities between luminance/chrominance and range/depth information in natural images can be further utilized in a variety of 3D image and video applications. For example, shape-from-X (shading, texture, etc.) algorithms can generate more accurate three-dimensional structures using additional color-range statistics, and 3D (stereo) quality assessment can better judge distortions from irregular chrominance and range correspondences. Future work will involve utilizing more detailed psychophysical models of human color and depth processing, and discovering more complete statistical models describing the interactions between color and range data in natural images.

## REFERENCES

- [1] L. Meesters, W. Ijsselstein, and P. Seuntjens, "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.
- [2] H. Kalva, L. Christodoulou, L. Mayron, O. Marques, and B. Furht, "Challenges and opportunities in video coding for 3D TV," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 1689–1692.
- [3] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, pp. 99–108, Aug. 2004.
- [4] M. Lambooi, W. Ijsselstein, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 3, pp. 030201-1–030201-14, 2009.
- [5] S. Reichelt, R. Haussler, G. Futterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," *Proc. SPIE*, vol. 7690, no. 1, p. 76900B, Apr. 2010.
- [6] B. A. Olshausen and D. J. Field, "Vision and the coding of natural images," *Amer. Sci.*, vol. 88, no. 3, pp. 238–245, 2000.
- [7] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment: The natural scene statistic model approach," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.
- [8] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [9] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, pp. 814–817, Aug. 1994.
- [10] D. J. Field, "Wavelets, vision and the statistics of natural scenes," *Phil. Trans. Royal Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 357, no. 1760, pp. 2527–2542, 1999.
- [11] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, no. 1, pp. 17–33, 2003.
- [12] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, Mar. 2001.
- [13] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, no. 1, pp. 167–192, Jan. 2008.
- [14] A. van der Schaaf and J. H. van Hateren, "Modelling the power spectra of natural images: Statistics and information," *Vis. Res.*, vol. 36, no. 17, pp. 2759–2770, 1996.
- [15] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," *Proc. SPIE*, vol. 3813, no. 1, pp. 188–195, 1999.
- [16] S. Lyu and E. P. Simoncelli, "Statistical modeling of images with fields of Gaussian scale mixtures," in *Advances in Neural Information Processing Systems*, vol. 19. Cambridge, MA, USA: MIT Press, 2007, p. 945.
- [17] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [18] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [19] 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.
- [20] 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.
- [21] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, Jun. 1996.
- [22] B. Olshausen and D. Field, "Natural image statistics and efficient coding," *Comput. Neural Syst.*, vol. 7, no. 2, pp. 333–339, 1996.
- [23] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [24] B. Potetz and T. S. Lee, "Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes," *J. Opt. Soc. Amer. A*, vol. 20, no. 7, pp. 1292–1303, Jul. 2003.
- [25] B. Potetz and T. S. Lee, "Scaling laws in natural scenes and the inference of 3D shape," in *Advances in Neural Information Processing Systems*, vol. 18. Cambridge, MA, USA: MIT Press, 2006, pp. 1089–1096.
- [26] 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.
- [27] C.-C. Su, A. C. Bovik, and L. K. Cormack, "Natural scene statistics of color and range," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 257–260.
- [28] C.-C. Su, L. K. Cormack, and A. C. Bovik, *LIVE Color+3D Database* [Online]. Available: [http://live.ece.utexas.edu/research/3dnss/live\\_color\\_plus\\_3d.html](http://live.ece.utexas.edu/research/3dnss/live_color_plus_3d.html)
- [29] I. Fine, D. I. A. MacLeod, and G. M. Boynton, "Surface segmentation based on the luminance and color statistics of natural scenes," *J. Opt. Soc. Amer. A*, vol. 20, no. 7, pp. 1283–1291, 2003.
- [30] J. R. Jordan, W. S. Geisler, and A. C. Bovik, "Color as a source of information in the stereo correspondence process," *Vis. Res.*, vol. 30, no. 12, pp. 1955–1970, 1990.
- [31] J. R. Jordan and A. C. Bovik, "Using chromatic information in edge-based stereo correspondence," *Comput. Vis., Graph., Image Process., Image Understand.*, vol. 54, no. 1, pp. 98–118, Jul. 1991.
- [32] J. R. Jordan and A. C. Bovik, "Using chromatic information in dense stereo correspondence," *Pattern Recognit.*, vol. 25, no. 4, pp. 367–383, Apr. 1992.
- [33] U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Perceptual quality assessment of color images using adaptive signal representation," *Proc. SPIE*, vol. 7527, no. 1, p. 75271, Jan. 2010.
- [34] D. J. Fleet, H. Wagner, and D. J. Heeger, "Neural encoding of binocular disparity: Energy models, position shifts and phase shifts," *Vis. Res.*, vol. 36, no. 12, pp. 1839–1857, 1996.
- [35] B. G. Cumming and G. C. DeAngelis, "The physiology of stereopsis," *Neuroscience*, vol. 24, no. 1, pp. 203–238, 2001.
- [36] I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis*. New York, USA: Oxford Univ. Press, 1995.
- [37] C. Blakemore, "The range and scope of binocular depth discrimination in man," *J. Physiol.*, vol. 211, no. 3, pp. 599–622, Dec. 1970.
- [38] Y. Liu, A. C. Bovik, and L. K. Cormack, "Disparity statistics in natural scenes," *J. Vis.*, vol. 8, no. 11, pp. 1–14, Aug. 2008.
- [39] S. 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, 2002.
- [40] RIEGL. *RIEGL VZ-400 3D Terrestrial Laser Scanner*, Orlando, FL, USA [Online]. Available: <http://rieglusa.com/products/terrestrial/vz-400/index.shtml>
- [41] RIEGL. *RIEGL RiSCAN PRO Software for 3D Terrestrial Laser Scanner*, Orlando, FL, USA [Online]. Available: <http://products.rieglusa.com/item/software-packages/friscan-pro-operating-and-processing-software/item-1008>
- [42] Intel Corporation. *OpenCV: Camera Calibration and 3D Reconstruction*, Santa Clara, CA, USA [Online]. Available: [http://opencv.willowgarage.com/documentation/camera\\_calibration\\_and\\_3d\\_reconstruction.html](http://opencv.willowgarage.com/documentation/camera_calibration_and_3d_reconstruction.html)
- [43] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.
- [44] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer. A*, vol. 70, no. 11, pp. 1297–1300, 1980.
- [45] 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.
- [46] D. H. Hubel, "The visual cortex of the brain," *Sci. Amer.*, vol. 209, no. 5, pp. 54–63, Nov. 1963.
- [47] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A*, vol. 2, no. 2, pp. 1160–1169, Jul. 1985.
- [48] 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.
- [49] R. L. D. Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vis. Res.*, vol. 22, no. 5, pp. 545–559, 1982.
- [50] R. L. D. Valois, E. W. Yund, and N. Hepler, "The orientation and direction selectivity of cells in macaque visual cortex," *Vis. Res.*, vol. 22, no. 5, pp. 531–544, 1982.
- [51] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, no. 6, pp. 1233–1258, 1987.
- [52] G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes," *Appl. Opt.*, vol. 26, no. 1, pp. 157–170, 1987.

- [53] W. Bialek, D. L. Ruderman, and A. Zee, "Optimal sampling of natural images: A design principle for the visual system?" in *Advances in Neural Information Processing Systems*, vol. 3. Cambridge, MA, USA: MIT Press, 1991, pp. 363–369.
- [54] J. H. van Hateren, "Real and optimal neural images in early vision," *Nature*, vol. 360, pp. 68–70, Nov. 1992.
- [55] J. H. van Hateren, "Spatiotemporal contrast sensitivity of early vision," *Vis. Res.*, vol. 33, no. 2, pp. 257–267, 1993.
- [56] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, 2002.
- [57] P. N. Belhumeur, "A Bayesian approach to binocular stereopsis," *Int. J. Comput. Vis.*, vol. 19, no. 3, pp. 237–260, 1996.
- [58] S. T. Barnard, "A stochastic approach to stereo vision," in *Proc. 5th Nat. Conf. Artif. Intell.*, vol. 1. Aug. 1986, pp. 676–680.
- [59] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.
- [60] R. Szeliski and R. Zabih, "An experimental comparison of stereo algorithms," in *Vision Algorithms: Theory and Practice*, vol. 1883. New York, USA: Springer-Verlag, 2000, pp. 1–19.
- [61] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Jul. 2001, pp. 508–515.



**Che-Chun Su** received the B.S. degree in electrical engineering, and the M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 2004 and 2006, respectively.

He joined the Laboratory for Image and Video Engineering, University of Texas at Austin, Austin, TX, USA, in 2009. His current research interests include image/video processing and compression, human vision perception, and 3D natural scene statistics.



**Lawrence K. Cormack** received the B.S. with Highest Honors in Psychology from the University of Florida Gainesville, FL, USA, in 1986 and the Ph.D. degree in physiological optics from the University of California, Berkeley, CA, USA, in 1992.

He is a Professor of psychology and neuroscience with the University of Texas at Austin, Austin, TX, USA, and is an Active Member of the Center for Perceptual Systems. His current research interests include the brain processing of the motion of objects through 3-D environments. He is also interested in

natural scene statistics and their relation to the evolution of visual systems, particularly with regard to depth and motion processing. He teaches mostly graduate statistics.



**Alan C. Bovik** (F'96) is 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 with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions. He has authored over 650 technical articles in his areas of expertise and holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009). His current research interests include image and video processing, computational vision, and visual perception.

He was named the SPIE/IS&T Imaging Scientist of 2011. He was a recipient of a number of major awards from the IEEE Signal Processing Society, including, the Best Paper Award in 2009, the Education Award 2007, the Technical Achievement Award in 2005, and the Meritorious Service Award in 1998. He was a recipient of 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 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 (SPIE), and the American Institute of Medical and Biomedical Engineering. He has been involved in numerous professional society activities, including the Board of Governors, 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, Editorial Board, the PROCEEDINGS OF THE IEEE, from 1998 to 2004, Series Editor for *Image, Video, and Multimedia Processing*, (Morgan and Claypool Publishing Company), since 2003, and the Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, TX, USA, in November, 1994.