Statistical Modeling of 3-D Natural Scenes With Application to Bayesian Stereopsis

Yang Liu, Lawrence K. Cormack, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—We studied the empirical distributions of luminance, range and disparity wavelet coefficients using a coregistered database of luminance and range images. The marginal distributions of range and disparity are observed to have high peaks and heavy tails, similar to the well-known properties of luminance wavelet coefficients. However, we found that the kurtosis of range and disparity coefficients is significantly larger than that of luminance coefficients. We used generalized Gaussian models to fit the empirical marginal distributions. We found that the marginal distribution of luminance coefficients have a shape parameter p between 0.6 and 0.8, *while* range and disparity coefficients have much smaller parameters $p < 0.32$, corresponding to a much higher peak. We also examined the conditional distributions of luminance, range and disparity coefficients. The magnitudes of luminance and range (disparity) coefficients show a clear positive correlation, which means, at a location with larger luminance variation, there is a higher probability of a larger range (disparity) variation. We also used generalized Gaussians to model the conditional distributions of luminance and range (disparity) coefficients. The values of the two shape parameters (p, s) reflect the observed luminance-range (disparity) dependency. As an example of the usefulness of luminance statistics conditioned on range statistics, we modified a well-known Bayesian stereo ranging algorithm using our natural scene statistics models, which improved its performance.**

***Index Terms*—Binocular vision, disparity, natural scene statistics (NSS), wavelets.**

I. Introduction

THE human vision system (HVS) is binocular. The two frontally placed, horizontally separated eyes receive two different retinal images nearly all the time. The angular difference between the two retinal images of a scene point is defined to be the binocular disparity. Given disparity and many other visual cues, the geometry of the 3-D visual space is reconstructed so quickly, effortlessly and seamlessly that a normal person never feels how difficult and ill-posed this problem can be. Biological vision systems achieve this ability by millions of years of natural selection in visual environments. The statistical properties of natural scenes have driven the evolution of biological eyes, and have also inspired scientific and engineering studies to understand and simulate visual perception by designing artificial visual systems. Here, we explain a series of statistical studies that we have conducted on luminance, range, and disparity in natural scenes. We discuss the potential significance of these results for better understanding depth perception, as well as their relevance to engineering applications.

Although the statistical properties of natural images are not yet well understood, we do know that the high-dimensional structure of natural images obeys statistical laws. The statistical nature of natural images has certainly been exploited by biological visual systems as they have evolved. Likewise, statistical models of natural images have proven to be quite useful in many areas of image processing and machine vision. Furthering and refining these models is likely to prove quite fruitful.

Early engineering efforts on natural image modeling were made by television engineers (see [1] for a review). Their work was based on a simple observation: nearby pixels have similar intensities. Multidimensional Gaussian distributions were used to model the correlations between pixels. This simple model was a natural choice given little prior knowledge other than empirical image correlations.

However, subsequent research has shown that the distribution of natural images is not Gaussian [1]–[4]. The main discovery showed that the univariate distributions of bandpass images (wavelet coefficients) typically have a high peak at zero, with heavier tails than Gaussian distributions having the same mean and variance. The informal explanation is that the peak near zero corresponds to smooth regions in images, while the heavy tails are caused by abrupt changes such as edges and corners, which are rich in natural images. These observations motivated the search for more precise image models. Several reasonably successful models have been proposed in the last two decades. One popular approach is to model the distributions using a generalized Gaussian (or stretched exponential) distribution:

$$P(c) = \frac{e^{-|c/s|^p}}{Z(s,p)} \tag{1}$$

where $Z(s,p)$ is a normalizing constant to force the integral of $P(c)$ to be 1. The parameter p controls the shape of the function: $P(c)$ is Gaussian when $p = 2$, and Laplacian when

Manuscript received June 09, 2010; revised November 04, 2010 and January 25, 2011; accepted January 26, 2011. Date of publication February 22, 2011; date of current version August 19, 2011. This work was supported by the National Science Foundation under Grant ITR-0427372 and Grant IIS-0917175. The Associate Editor coordinating the review of this manuscript and approving it for publication was Prof. Stefan Winkler.

Y. Liu was with the Center for Perceptual Systems and the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX-78712 USA. He is now with Wellesley College, Wellesley, MA 02481 USA, and also with Harvard Medical School, Boston, MA 02115 USA (e-mail: liu.yang76@gmail.com).

L. K. Cormack is with the Center for Perceptual Systems and the Department of Psychology, University of Texas at Austin, Austin, TX-78712 USA (e-mail: cormack@psy.utexas.edu).

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

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

Digital Object Identifier 10.1109/TIP.2011.2118223

$p = 1$. In general, decreasing p causes a higher center peak and longer tails. For wavelet coefficients of natural images, p typically lies in the range [0.4, 0.8] [5], [6]. The parameter controls the scale of the density function, and hence the variance of the distribution. Maximum likelihood estimation of p and s yields very good fits between empirical histograms and the model. The model has proved to be quite useful in the design of image compression [6] and image denoising [5], [7] algorithms.

The first-order distribution models developed to date, while useful, are too simple to guide many complicated image processing applications. Although the wavelet coefficients of natural images are generally approximately decorrelated, higher order dependencies still exist. This can be visually observed by looking at the subbands of wavelet coefficients, where clusters of large coefficients concentrate on structural image features such as lines, edges, and corners. A locally adaptive statistical model called the Gaussian scale mixture (GSM) was adapted to capture the spatial dependencies between neighboring image wavelet coefficients. In this model, a neighborhood of a wavelet coefficient $\vec{x}$ is expressed as the product of a zero-mean Gaussian vector $\vec{u}$, and an independent positive scalar random variable $\sqrt{z}$:

$$\vec{x} \sim \sqrt{z}\vec{u}. \tag{2}$$

Estimating z and u is more complex than estimating the parameters of a generalized Gaussian model. However, the model (2) more specifically accesses the local structure of images and has been shown to significantly improve algorithms for image denoising [5], [7]–[10], and texture analysis/synthesis [11]. Other influential properties of natural image statistics are studied in [12] and [13].

The aforementioned statistical modeling of natural images has been limited to 2-D monocular intensity images. Although monocular images contain a certain amount of depth related information, in the absence of motion, binocular images are essential for accurate depth perception. In the last decade, several researchers have studied the prior distribution of distance in natural scenes using laser scanners [14]–[17]. Huang, Lee, and Mumford measured the first-order and second-order statistics of range maps and (Haar basis) wavelet coefficients [14]. Their results verified the intuitive assumption that range maps are more regular than luminance images, since abrupt changes of illumination, shadow, and texture are more common in luminance images, and arise from more diverse phenomena. Yang and Purves [15] acquired range images of forest scenes. They found that the 3-D geometry of the scenes in their database is quite rough, has a characteristic anisotropic topography, and is approximately scale invariant. In another paper [18], they tried to explain several phenomena in depth perception by using the distribution of physical distances from the same dataset.

Potetz and Lee [17] studied the important question of the bivariate statistical behavior of coregistered range and luminance images. The Riegl LMS-Z360 scanner they used had the ability (that the previous instruments lacked) to simultaneously measure and register range and luminance with good precision. Using 50 rural and urban image pairs (luminance and range), they found a negative range–luminance correlation, suggesting that nearer objects tend to appear brighter than far objects, on average. They also deployed a few convex range filters, selected for specific structural properties relevant to computer vision, which were used to filter both the range and luminance images. Using a canonical correlation analysis, the authors found a relatively low degree of (strictly linear) correlation between the filtered luminance and range patches. The use of shape functions, of course, presupposes that the shapes selected are relevant to the relationship between luminance and range. Moreover, only linear relationships, analyzed at a single scale, between specific shape functions were considered. In a later study on the same dataset, Potetz and Lee [19] examined the relationship between luminance and range over multiscales and applied their results to shape-from-shading problems. Roth and Black [20] studied the statistics of optical flow in natural temporal scenes, by adopting a method similar to ours. They used coregistered luminance–range maps to study the distribution of optical flow patterns.

The HVS has very fine stereo acuity; under the best conditions, the stereo acuity falls between 2 arcsec to 6 arcsec [21]. The HVS also has a large upper disparity limit. Blakemore [22] reported that the upper limit is 7° for crossed disparities, and 12° for uncrossed disparities. The broad operating range and excellent acuity of stereopsis indicate that disparity is extensively used for depth perception. In a previous effort, we studied the disparity distributions of natural scenes [23] by converting forest range maps to disparity maps. We found that the disparity distributions at eye level are centered at zero, are non-Gaussian (but well modeled as generalized Gaussian), and span about 5°. A similar study on indoor range maps showed similar results. In an effort to correlate disparity sensitivity with naturally available disparities, we cast our findings using both outdoor and indoor data against physiological data from two prior studies on the distributions of disparity tuning of disparity-tuned V1 neurons, [24]–[26]. The proportion of near- and far-tuned disparity distributions qualitatively agrees with the distribution of disparity tuning neurons in V1. But the span (about 5°) of natural disparities is larger than the operational range of the narrowly tuned V1 neurons (about 2°). This suggests that larger disparities might be encoded and represented by other visual areas, because the HVS can make disparity judgments above chance at rather large disparities (about 7°–12°) (for more details, see [21]).

According to a study on disparity tuned area MT neurons [27], the coding range of MT neurons is about 4.7°, which agrees well with natural disparities. However, far more (61%) of the MT neurons are tuned to near disparities, compared to the almost symmetric distribution of natural disparities. We also conducted a study on the elevation dependency of disparity distributions and found that the Helmholtz shear of binocular correspondence correlates with mean disparity along the vertical median plane [28]. Helmholtz shear is the property that empirically corresponding points along the vertical meridian are not geometrically corresponding points on the two retinas of a human binocular vision system. Helmholtz proposed that the shear of empirically corresponding points may help us in walking.

Depth perception involves multiple visual cues. The cues can be classified into monocular cues and binocular cues. Motion

parallax, familiar size, relative size, shading, aerial perspective, and occlusion are all monocular cues. Binocular cues include disparity and convergence. All these cues are likely combined in the brain in a probabilistic manner to infer depth. Several methods have been proposed to explain the weighting of different visual cues [29]–[33], but oversimplified stimuli were used to test these models, owing to a lack of knowledge regarding accurate distributions of these cues in natural scenes. Accurate scene statistics obtained from coregistered luminance, range, and disparity images of natural environments have the potential to improve these algorithms and to deepen our understanding of depth perception. Likewise, to enrich our understanding of visual neurophysiology, accurate statistical models of the disparity processing of simple and complex neurons are also needed [34], [35].

Our work utilized the coregistered luminance–range dataset from [17]. We derived disparity maps from the range maps by assuming that a human observer is viewing a scene with two eyes (cameras) that are parallel, and that fixate at infinity. We then decomposed the luminance, range, and disparity maps into multiresolution and multioriented subbands of wavelet coefficients. We discuss the nature of these in detail in Section II. We describe marginal, joint and conditional distributions of luminance, range, and disparity in Sections III–VI. In order to show the potential applications of these results both in vision science and computer vision, we applied our results to a classical problem in both areas: stereo correspondence. The stereo correspondence problem is formalized into a Bayesian framework, and the statistical relationships we learned from the coregistered luminance–range guide our algorithm. The details on our implementation and the results and analysis of the algorithm are in Section VII. Finally, Section VIII summarizes the results and discusses potential applications both in vision science and in engineering.

II. Dataset

We used the coregistered range maps and color images from [17]. The authors used a Riegl LMS-Z360 terrestrial scanner, which has a maximum range of 200 m and an accuracy of 12 mm. The original color images were converted to 8 bit gray-scale light intensity images. There are several regions that do not yield valid distances. Usually, these regions include highly reflective surfaces such as water and glass, or places that are out of range, like the sky. In [17], these patches containing invalid pixels were excluded from analysis (It is easy because the sky pixels are labeled as 0 s in the range maps). Because we are interested in the statistics of wavelet coefficients, we first assigned all pixels of sky 1000 m distance in order to be able to compute valid wavelet transforms on range maps. Later, we carefully excluded all coefficients contaminated by the sky pixels in all subbands in our analysis.

We adopted the same method as we used in [23] to convert distance to disparity. A pixel of a range map that describes the distance of a visual direction (α, β) in a scene is denoted by $[x(\alpha,\beta), y(\alpha,\beta), z(\alpha,\beta)]$, assuming that the 3-D coordinate system is centered at the scanner's location. The azimuth and elevation of the visual direction from the scanner are represented by α and β, respectively. As Fig. 1 shows, we assumed an observer located at the coordinate center O, with his two eyes (interocular distance 6.5 cm) at $O_l[-0.0325, 0, 0]$ and $O_r[0.0325, 0, 0]$. With the x, y, z values of a physical point P specified, we can derive the geometrical relationship between the two retinal projections P_l and P_r. In Fig. 1, O_l and O_r are the nodal points of the left eye and right eye, respectively. C_l and C_r are the foveal centers of the two eyes. We assumed that the two visual axes C_rO_r and C_lO_l are parallel. The four points O_l, O_r, C_l, and C_r are located on the horizontal median plane, which cut the two eyeballs in the middle and parallel to the ground. The perpendicular projections of P_l and P_r on the horizontal median plane are Q_l and Q_r. The two angles $\angle Q_rO_rC_r$ and $\angle Q_lO_lC_l$ can be easily calculated. We then defined the horizontal disparity to be:

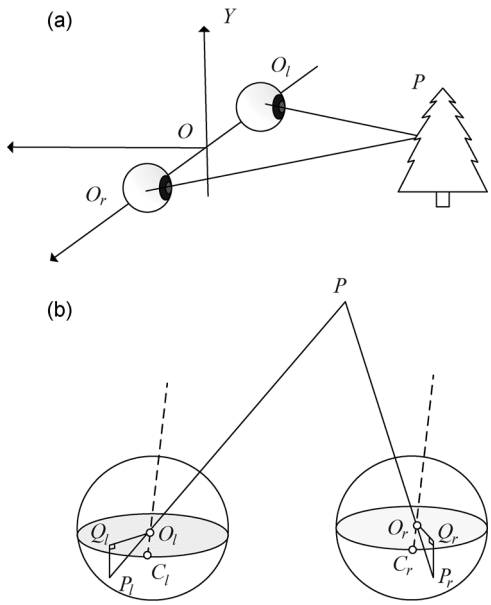

Fig. 1. Binocular vision system in natural scenes. Assuming two eyes are fixating at infinity, the disparity of a physical point P can be calculated easily from the coordinates of P and a simplified eye model.

$$d = \angle Q_rO_rC_r - \angle Q_lO_lC_l. \tag{3}$$

Thus, we obtained a disparity map $D(i,j)$ once every distance in the range map $R(i,j)$ was converted. There are several points that need to be emphasized regarding range to disparity conversion:

1) We treated a range map with coregistered luminance image as a cloud of particles. Each particle has its (x, y, z) position, and its luminance specified. While this is not a complete description of the 3-D scene, these data simply and accurately represent the local 3-D scene at the position of the terrestrial scanner.
2) We assumed that all points are visible for both eyes. Naturally, there could be some half occluded points that are only visible to one eye, but occluded in the other eye. However, given the large scene distance and small interocular distance (or the baseline distance in the language of computational stereo matching), half occlusions may be regarded as generally neglectable.

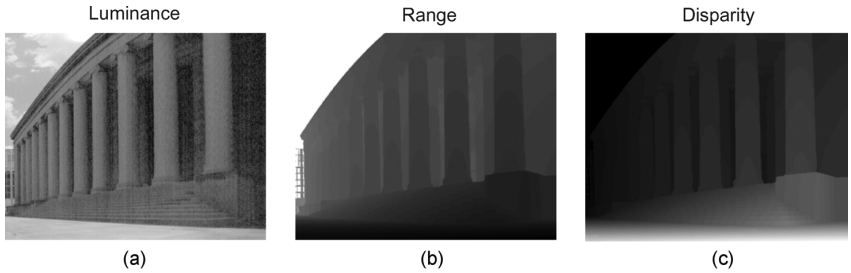

Fig. 2. Luminance, range, and disparity of a natural scene are shown in (a), (b), and (c). Brighter pixels represent larger values. Notice the sky (the white region) pixels in the range map are assigned 1000 m uniformly. Interestingly, the large, prominent range discontinuity between the sky and the building are not represented in the disparity map. Instead, the nearer steps in the range map are more observable in the disparity map than the farther steps. This illustrates the effect wherein relative disparity scales with distances.

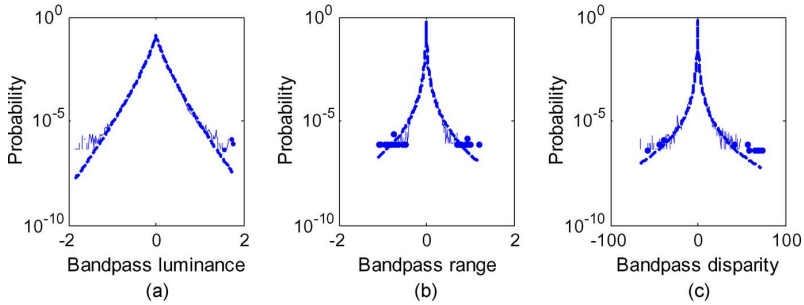

Fig. 3. Bandpass luminance, range, and disparity histograms and their generalized Gaussian fits of one subband. The solid lines are the empirical histograms. The dashed lines are the best fitting generalized Gaussian. The isolated dots occur where the corresponding histogram bins contain some elements, but the neighboring bins contain zero elements.

3) The unit of disparity used is degrees, which is equivalent to disparity in pixels.

For each natural scene, we acquired three kinds of information: $I(i,j)$, $R(i,j)$, and $D(i,j)$. All of the following analysis is based on multiresolution wavelet decompositions of the three scene properties. Fig. 2 shows the three images of luminance, range, and disparity of the same scene. In the range and disparity maps of Fig. 2, the magnitudes increase from dark to bright. One obvious observation is that most of the large disparities are below eye level, which is caused by the nearer ground plane. Several images in our database contain ground planes in the lower half and the sky in the upper half. This dependency of disparity distribution on elevations caused a shearing of the corresponding positions of retinal points along the vertical meridian. This is known as the Helmholtz shear, which Helmholtz first noticed [36]. He explained it by assuming that the HVS utilizes the shear for walking.

III. Marginal Distributions

We decomposed $I(i,j)$, $R(i,j)$, and $D(i,j)$ into a wavelet pyramid of three scales and six orientations using the steerable filterbank designed by Simoncelli [37]. To be consistent with luminance perception of the HVS and the convention of natural scene analysis [17], [38], the wavelet decomposition was performed on the logarithm of luminance and the logarithm of distance.

The distributions of the wavelet coefficients of natural photographic images are generally known to be non-Gaussian, to have a high peak and heavy tails. Many researchers have used a generalized Gaussian to fit the distribution. The parameters of the generalized Gaussian can be derived easily either through maximum likelihood methods, or by simply fitting the mean and variance of the distribution. Huang and Mumford [14] modeled Haar coefficients of range maps. However, the wavelet coefficients of disparity have not been deeply modeled. In our previous study on disparity statistics, which is based on the range dataset from [15], we addressed the non-Gaussianity of disparity distributions and simply pointed out that a Laplacian model (a special case of the generalized Gaussian model with $p = 1$) is a better fit than Gaussian distribution. Mumford *et al.* studied and modeled bandpass luminances and ranges in combination [12], [14], [39]. However, no serious model has been proposed for the bandpass disparities of natural scenes.

Without loss of generality, we chose the second scale, horizontal subbands of luminance, range, and disparity coefficients to be fitted using the generalized Gaussian model (see Fig. 3). The generalized Gaussian distribution fitting of other subbands with different scales and orientations are very similar to Fig. 3. There have been many methods proposed to estimate the parameters of the generalized Gaussian distribution [4], [40]–[45]. We used a simple one that only involves estimating the first and second moments using the sample mean and sample variance [4], [45]. The solid plots in Fig. 3 show the normalized histogram of wavelet coefficients using 200 bins evenly spaced between the maximum value and the minimum value of all coefficients, and the dashed plots show the generalized Gaussians with fitted parameters. The left, center, and right panels of Fig. 3 represent bandpass luminance, bandpass range, and bandpass disparity. We found that range and disparity coefficients have much larger kurtosis (94.3 for bandpass range and 849 for bandpass disparity) than those of luminance coefficients (about 11).

TABLE I
SHAPE PARAMETER p OF LUMINANCE WAVELET MARGINALS

Scale	*horizontal*	30°	60°	*Vertical*	120°	150°
1	0.70	0.72	0.76	0.72	0.70	0.70
2	0.72	0.78	0.79	0.77	0.80	0.77
3	0.66	0.78	0.69	0.64	0.69	0.74

TABLE II
SHAPE PARAMETER p OF RANGE WAVELET MARGINALS

Scale	*horizontal*	30°	60°	*Vertical*	120°	150°
1	0.23	0.24	0.23	0.24	0.23	0.24
2	0.25	0.26	0.26	0.25	0.26	0.27
3	0.29	0.32	0.32	0.32	0.30	0.32

TABLE III
SHAPE PARAMETER p OF DISPARITY WAVELET MARGINALS

Scale	*horizontal*	30°	60°	*Vertical*	120°	150°
1	0.18	0.18	0.19	0.19	0.19	0.19
2	0.22	0.23	0.22	0.22	0.21	0.22
3	0.25	0.28	0.29	0.29	0.29	0.27

This indicates that the distributions of range and disparity coefficients have much sharper peaks and heavier tails than those of luminance coefficients. This corresponds well to a simple fact: significant, sharp luminance variations are more common than similar range variations. The disparity distributions have even larger kurtosis, because the distance-to-disparity conversion greatly attenuates larger distances, which means the disparity map contains even more small values. The scale invariance can be observed by the uniform goodness of the fits and the similar shapes observed at different scales. We listed the shape parameter p of all subbands of three scales and six orientations of luminance, range, and disparity in Tables I–III, respectively. Bandpass luminance generally has a large p from 0.66 to 0.8. It is larger than the shape parameter p of bandpass range and bandpass disparity, which is typically between 0.2 and 0.3.

IV. JOINT DISTRIBUTIONS

Julesz [46] showed that humans can perceive depth from random dot stereograms, largely separating the perception of depth structure from that of luminance structure. However, we do not live in a random world but a world obeying physical laws. Many visual cues may contribute to depth perception. The statistical relationships between these visual cues in natural scenes are essential for probing the mechanism of depth perception.

We use a simple method to show that even the coarsest features of luminance and range are not independent.

Figs. 4(a) and (b) show the contour plots of the joint histograms of coregistered luminance–range, and luminance–disparity (similar to $P(x,y)$) of the wavelet coefficients in a single subband. The isoprobability contours are roughly symmetric about the horizontal and vertical axes. Fig. 4(c) and (d) show the contour plots of the joint histograms of independent luminance–range, and luminance–disparity by independently drawing samples from the luminance, range, and disparity coefficients (similar to $P(x)P(y)$). There are subtle differences between the histograms of the coregistered data and that of the

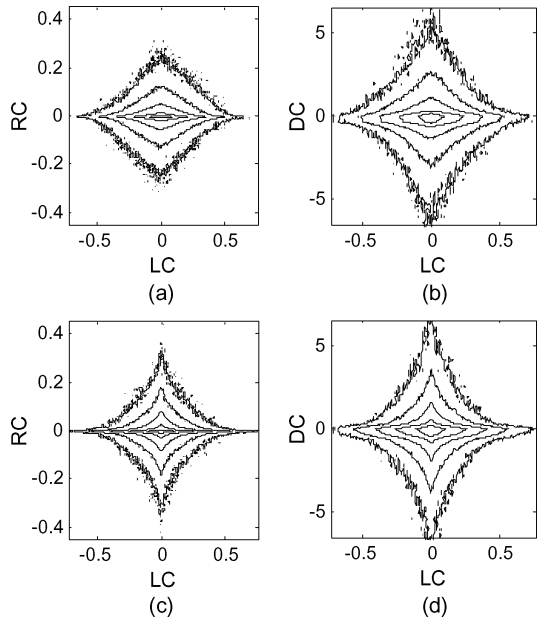

Fig. 4. (a), (b) show the isoprobability (logarithm) contours of the joint histograms of luminance coefficients (LC) and range coefficients (RC), the joint histograms of luminance coefficients (LC) and disparity coefficients (DC). (c), (d) show the joint histograms of independent LC-RC, LC-DC by breaking up the coregistered relationship in (a), (b), respectively. A simple interpretation is that the top row shows the joint probability $P(x,y)$, while the bottom row shows the product of two marginals $P(x)P(y)$.

independent data. For example, if luminance coefficients were independent of range coefficients, the isoprobability contours would have the star-like shape that characterizes the product of two highly peaked, generalized Gaussian probability density functions [as in Fig. 4(c)], rather than a leaf-like shape seen in Fig. 4(a). These differences indicate that there exists higher order (>2), nonlinear statistical dependencies between luminance and range. As we will show in later sections (see Fig. 7), the correlations between magnitude of bandpass luminance and that of bandpass range/disparity are weakly positive across all scales and all orientations. This suggests that bandpass luminance and bandpass range/disparity are not independent.

V. CONDITIONAL DISTRIBUTIONS

In this section, we explain our studies of the conditional distributions of range, luminance, and disparity coefficients. Previous work on wavelet coefficient distributions of luminance images has indicated that there exist interesting dependencies between neighboring coefficients at different scales, orientations, and spatial positions [5]. If a wavelet coefficient has a large magnitude, then the distribution of its neighbors will often have a large variance. This observation strongly motivates a multiscale analysis of the relationship between luminance and range.

Human depth perception is not totally damaged if we lose one eye. That means that stereopsis is not the sole source of depth information in the brain. There are many monocular visual cues contributing to depth perception. Since the brain is well modeled as performing multichannel decompositions of natural scenes, we are interested in the following question: At a location in 3-D natural scene, if we know the magnitude of its bandpass luminance from one monocular image, what can we expect of

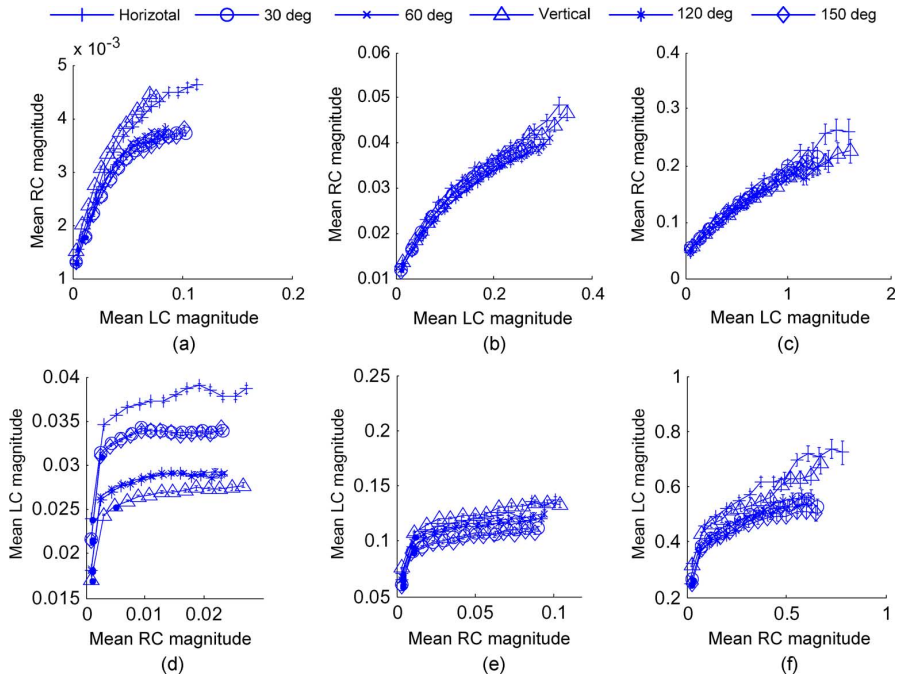

Fig. 5. (a), (b), (c) are the mean conditional magnitudes of range coefficients given the magnitudes of luminance coefficients in all six subbands at three scales ((a)–(c): from the finest to the coarsest scale). (d), (e), (f) show the mean conditional magnitudes of luminance coefficients given the magnitudes of range coefficients ((d)–(f): from the finest to the coarsest scale). All these plots are monotonic.

the bandpass range (disparity), or vice versa? We tried to answer this question by analyzing the dependency between luminance wavelet coefficient and range (disparity) coefficient in the same spatial location (coordinate) of the same subband (orientation and scale).

Specifically, we studied the conditional histograms of luminance coefficients that are conditioned on the magnitudes of range/disparity coefficients, and the conditional histograms of range/disparity coefficients conditioned on the magnitudes of luminance coefficients. Overall, we found that the magnitudes of luminance coefficients and the magnitudes of range (disparity) coefficients have a significant positive correlation, and the positive correlations are stronger at coarser scales.

Denote luminance coefficient, range coefficient, and disparity coefficient as $\tilde{L}$, $\tilde{R}$, and $\tilde{D}$. The conditional expectation of a range coefficient given a luminance coefficient can be expressed as $E(\|\tilde{R}(x,y)\|\|\|\tilde{L}(x,y)\| = \tilde{l})$, and the conditional expectation of a luminance coefficient given a range coefficient is $E(\|\tilde{L}(x,y)\|\|\|\tilde{R}(x,y)\| = \tilde{r})$, where $\|\cdot\|$ represents wavelet magnitude.

To estimate $E(\|\tilde{R}(x,y)\|\|\|\tilde{L}(x,y)\| = \tilde{l})$, we set n evenly spaced bins between the maximum and the minimum of all luminance coefficients, then we computed the mean magnitude and 95% confidence interval of the range coefficients whose coregistered luminance coefficients' magnitudes fall in each luminance bin:

$$\hat{E}\left(\|\tilde{R}\|\|\|\tilde{L}\| \in [\tilde{l}_k, \tilde{l}_{k+1}]\right) = \frac{1}{N_l}\left(\sum_{1}^{N_l} \tilde{R}(x,y)\right) \left\{(x,y)|\tilde{l}_k \le \left\|\tilde{L}(x,y)\right\| \le \tilde{l}_{k+1}\right\}. \quad (4)$$

Similarly, we can also bin the range coefficients according to their magnitudes and compute the mean magnitude of luminance coefficients as:

$$\hat{E}\left(\|\tilde{L}\|\|\|\tilde{R}\| \in [\tilde{r}_k, \tilde{r}_{k+1}]\right) = \frac{1}{N_r}\left(\sum_{1}^{N_r} \tilde{L}(x,y)\right) \left\{(x,y)|\tilde{r}_k \le \left\|\tilde{R}(x,y)\right\| \le \tilde{r}_{k+1}\right\}. \quad (5)$$

The same analysis also goes for the conditional expectation of luminance coefficients conditioned on disparity coefficients, and vice versa.

In Fig. 5(a)–(c), we plotted the conditional mean magnitude of bandpass range given the mean magnitudes of bandpass luminance, in 15 bins of luminance coefficients in all subbands of three scales, with (a)–(c) representing the finest scale to the coarsest scale, respectively. It is obvious that the mean magnitude of bandpass range increases with the magnitude of bandpass luminance. Fig. 5(d)–(f) show the conditional mean magnitudes of bandpass luminance given bandpass range. There exists a notable difference between the curves of conditional mean bandpass range $\hat{E}(\|\tilde{R}\|\|\|\tilde{L}\|)$ in Fig. 5(a)–(c) and the curves of conditional mean bandpass luminance $\hat{E}(\|\tilde{L}\|\|\|\tilde{R}\|)$ in Fig. 5(d)–(f). The conditional bandpass luminance $\hat{E}(\|\tilde{L}\|\|\|\tilde{R}\|)$ is also monotonic, but saturates faster than conditional bandpass range $\hat{E}(\|\tilde{R}\|\|\|\tilde{L}\|)$, as shown in Fig. 5(a)–(c). The monotonic trend suggests that smooth surfaces generally create small luminance variations, while rough surfaces contain large luminance variations. Furthermore, the saturation suggests that the magnitudes of luminance changes at different range changes stabilize at a certain level when the range changes are large enough.

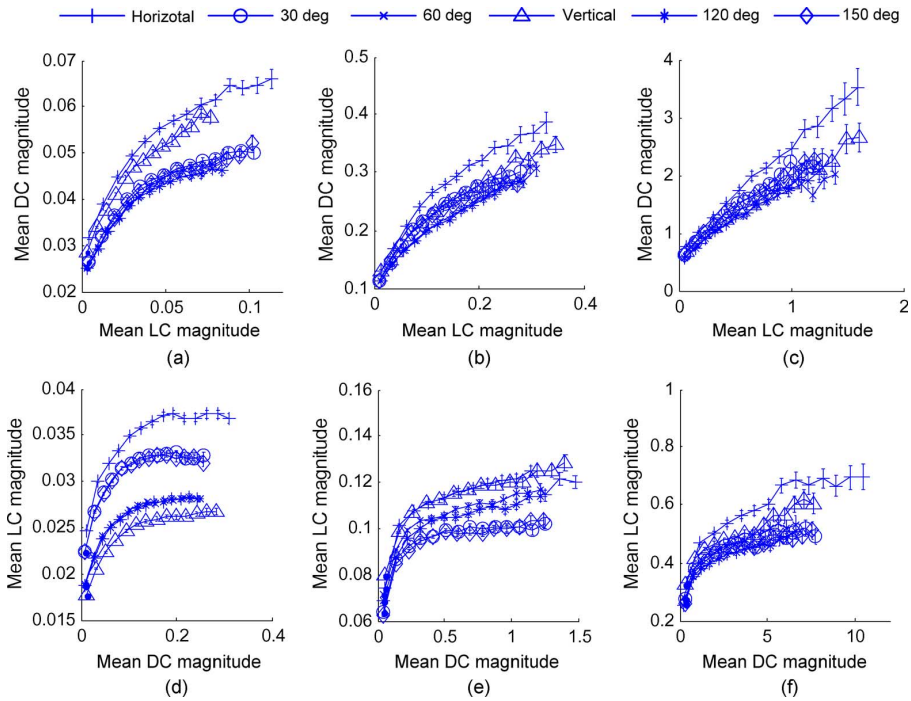

Fig. 6. (a), (b), (c) are the mean conditional magnitudes of disparity coefficients given the magnitudes of luminance coefficients in all six subbands at three scales ((a)–(c): from the finest to the coarsest scale). (d), (e), (f) show the mean conditional magnitudes of luminance coefficients given the magnitudes of disparity coefficients ((d)–(f): from the finest to the coarsest scale). Similar to Fig. 5, all these plots are monotonic.

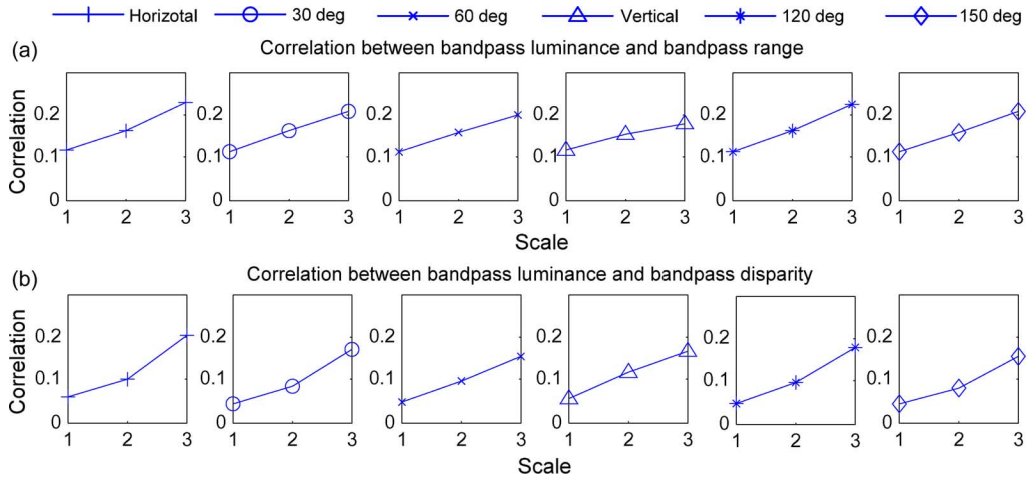

Fig. 7. (a) Shows the correlation coefficients between the magnitudes of coregistered bandpass luminance and those of bandpass range at all six orientations, and all three scales of subbands; (b) shows the correlation between coregistered bandpass luminance and bandpass disparity. All these plots suggest that: (1) the magnitudes of bandpass luminance and of bandpass range (disparity) have a positive correlation; and (2) the correlation gets stronger at coarser scales. Applying the t-test of significance of these positive correlations, we found that all the p values to be $< 10^{-3}$. The weakly positive correlations represented by these plots are statistically significant.

Likewise, Fig. 6 shows the conditional mean magnitudes of bandpass disparity $\hat{E}(||\tilde{D}|||||\tilde{L}||)$, given the magnitudes of bandpass luminance in (a)–(c), and the mean magnitudes of bandpass luminance $\hat{E}(||\tilde{L}|||||\tilde{D}||)$ conditioned on bandpass disparity in (d)–(f). The curves show a clear positive correlation between $||\tilde{L}||$ and $||\tilde{D}||$. We also notice the saturation of bandpass luminance $||\tilde{L}||$ conditioned on bandpass disparity $||\tilde{D}||$.

We plotted the correlation coefficients between and bandpass range $||\tilde{R}||$ (disparity $||\tilde{D}||$) in Fig. 7. As expected from Figs. 5 and 6, $||\tilde{L}||$ and $||\tilde{R}||$ ($||\tilde{D}||$) display positive correlations across all scales and orientations. Most of the correlation coefficients fall between 0.05 and 0.3. The positive correlation is significant given the large number of coefficients. For example, there are about 330,000 coefficients at the coarsest scale. The t-test of the significance of these correlations yielded a very small p values ($p < 10^{-3}$) because of the large degree of freedom.

Most importantly, we observed a consistent tendency: $||\tilde{L}||$ and $||\tilde{R}||$ ($||\tilde{D}||$) have stronger positive correlations at coarser scales. This monotonic relationship between correlation and scale implies that luminance variations over a larger scale are more strongly correlated with large scale range variat ions. We attempt to explain this observation as follows.

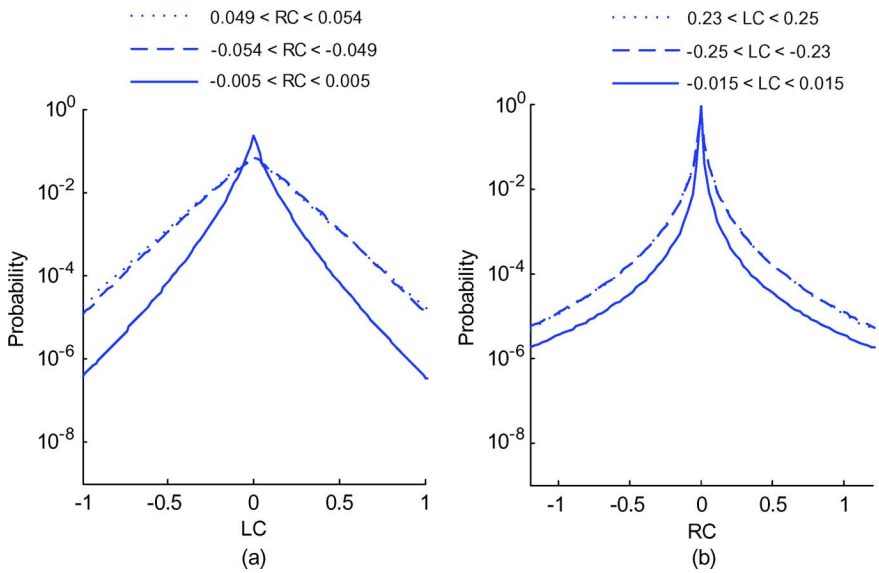

Fig. 8. Generalized Gaussian fits of (a) histograms of luminance coefficients at different range coefficients, and (b) range coefficients, given specific luminance coefficients. Larger magnitudes of range coefficients indicate a larger variance of the luminance coefficient distribution, and vice versa.

Generally, a 3-D natural scene contains piecewise continuous surfaces and range discontinuities. Luminance changes occur almost surely at range discontinuities but depend on many factors (surface reflectance, shadow, illumination, etc.) on continuous surfaces. The range boundaries (discontinuities) and their associated luminance boundaries tend to be larger in scale, and more regular in shape, which are well captured by coarser scale subbands, and yield stronger positive correlations. Conversely, the luminance variations within object surfaces tend to be smaller in scale, and more irregular in shape, which are more likely to be captured by finer scale subbands. In this case, the correlation between luminance variation and range variation is weak, because range variations are mostly small within a smooth surface. Consider a frequent experience in daily life: one can clearly see the outline (luminance change coinciding with a large scale range change) of a bill board 500 meters away when driving on a highway but can only read the print (luminance change not coinciding with a small scale range change) on the bill board at 30 m.

The findings in Figs. 5–7 have many important implications. In binocular stereopsis, two monocular luminance images are the starting point for all subsequent cortical computations. Likewise, many computational stereo algorithms use luminance edges or luminance wavelet coefficients as primitives for stereo correspondence. From the earlier analysis, it seems that monocular bandpass luminance contains strong information regarding the behavior of range and disparity. Since the ultimate purpose of depth perception is to correctly measure distances using every available piece of sensory information, it is interesting to speculate how the brain may utilize the notable relationship in Figs. 5–7 combined with stereopsis, to reach its goal. Correspondingly, it is of great interest to wonder if the relationship might be used to improve the state-of-the-art of computational stereo ranging algorithms. For example, luminance cues might be used to predict range, or at least serve as a consistency check in stochastic approaches to stereo disambiguation.

VI. Modeling Conditional Distributions

We studied the histograms of mutual luminance and range (disparity) coefficient conditioning, and used generalized Gaussians to model the conditional histograms.

First, we investigated the histogram of range coefficients of a particular orientation and scale subband, conditioning on luminance coefficients at the same location of the same subband. We found the maximum value of luminance coefficients $\tilde{l}_{\max}$, and the minimum value of luminance coefficients $\tilde{l}_{\min}$. We set n evenly spaced bins between $\tilde{l}_{\max}$ and $\tilde{l}_{\min}$, where $\tilde{l}_{\min} = \tilde{l}_0 < \tilde{l}_1 < \ldots < \tilde{l}_k \ldots < \tilde{l}_n = \tilde{l}_{\max}$, where the kth bin is $[\tilde{l}_{k-1}, \tilde{l}_k]$, and $\tilde{l}_k - \tilde{l}_{k-1} = (\tilde{l}_{\max} - \tilde{l}_{\min})/n$. The histograms of the range coefficients in each bin are conditional histograms. We found that all conditional histograms are well modeled as generalized Gaussian with different parameters. We did a thorough analysis over all orientations and all scales and found that generalized Gaussian fitting is robust over all scales and orientations. The shape parameters of the fitted generalized Gaussians of scene features from one domain, conditioned on features from another domain (luminance, range, or disparity), exhibits certain dependencies. In short, it is consistent with the observed magnitude dependencies depicted in Figs. 5 and 6.

Fig. 8(a) shows the conditional distributions (generalized Gaussian fits) of luminance coefficients from one subband, given different magnitudes of range coefficients. For visual clarity, we only display the conditionals in three range coefficient bins, corresponding to second, fifth, and eighth decentiles, respectively. The curves represent the generalized Gaussian fits to the normalized conditional histograms (200 bins) of luminance coefficients. The solid, dashed, and dotted curves are the fitted generalized Gaussian probabilities $P(\tilde{L}|\tilde{R})$, when all of the following: $-0.005 < \tilde{R}(x,y) < 0.005$, $-0.054 < \tilde{R}(x,y) < -0.49$ and $0.049 < \tilde{R}(x,y) < 0.054$.

It is visually obvious that larger range coefficient magnitudes have lower peaks at zeros and higher probabilities for

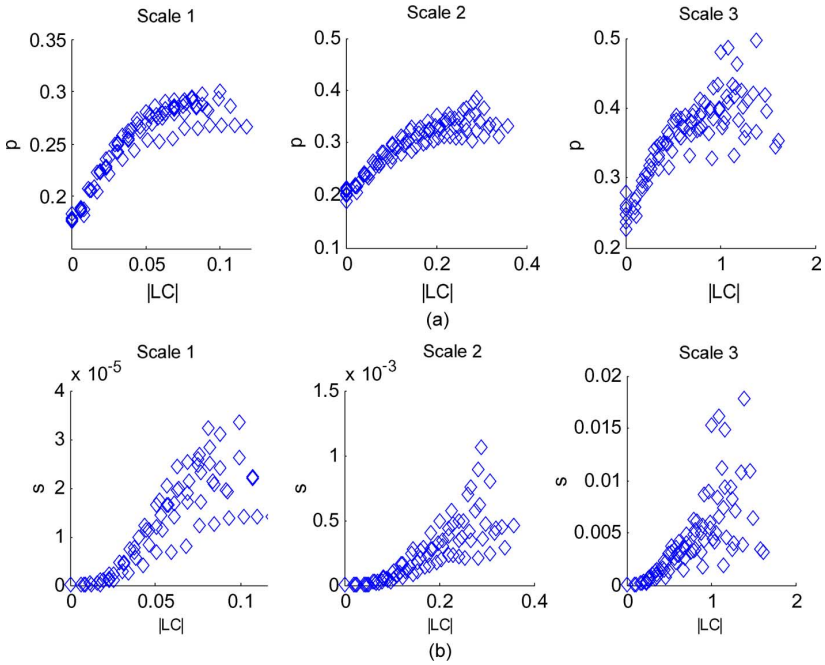

Fig. 9. We used generalized Gaussians to model $P(\bar{R}|\bar{L})$, the conditional distribution of range coefficients given luminance coefficients. As shown in Fig. 8, the dependency between the shapes of the fitted generalized Gaussians are kept in the shape parameters p and s. (a) is the scatter plot of p and the magnitudes of luminance coefficients, and (b) shows the scatter plots of s and magnitudes of luminance coefficients. Each panel shows p or s and their associated luminance coefficients from the 6 oriented subbands within a scale. From left to right, the scale goes from the finest to the coarsest.

TABLE IV
CORRELATION COEFFICIENTS BETWEEN SHAPE PARAMETER (p, s) AND $\|\bar{L}\|$ IN $P(\bar{R}|\|\bar{L}\|)$

Scale	$corr(p, \|\bar{l}\|)$	$corr(s, \|\bar{l}\|)$
1	0.87	0.86
2	0.88	0.82
3	0.78	0.32

larger luminance coefficient magnitudes. The differences in the curves fit by generalized Gaussians are expressed by the shape parameters.

The conditionals of range coefficients $P(\tilde{R}|\tilde{L})$ given luminance coefficients are plotted in Fig. 8(b). We observe similar behavior as in Fig. 8(a): The shapes of $P(\tilde{R}|\tilde{L})$ show a strong dependence on bandpass luminance $\tilde{L}$. Larger luminance coefficients yield a lower peak and heavier tails. We also observed very similar behavior of $P(\tilde{D}|\tilde{L})$ and $P(\tilde{L}|\tilde{D})$, which are not plotted for brevity.

We visualized the p and s of $P(\tilde{R}|\tilde{L})$ in Fig. 9(a) and (b), respectively. Each panel in Fig. 9 displays the scatter plots of the shape parameters of $P(\tilde{R}|\tilde{L})$ against $||\tilde{L}||$ in all six orientation subbands of one scale. It is obvious that both p and s have a very strong positive correlation with $||\tilde{L}||$. Table IV lists the correlation coefficients between p and $||\tilde{L}||$, and those between s and $||\tilde{L}||$.

Similarly, the conditional distribution of disparity coefficients given luminance coefficients, $P(\tilde{D}|\tilde{L})$ is also modeled as generalized Gaussian, with shape parameters p and s that depend on the magnitudes of the disparity coefficients $||\tilde{L}||$. Fig. 10 shows the scatter plots of p and s against the magnitudes of bandpass luminance $||\tilde{L}||$. Again, there are strong positive correlations between p and $||\tilde{L}||$, and between s and $||\tilde{L}||$. We list these correlation coefficients in Table V.

We also performed the same analysis on the conditional distribution of luminance coefficient $P(\tilde{L}|\tilde{R})$ and $P(\tilde{L}|\tilde{D})$. Fig. 11 shows the scatter plots of the shape parameters of $P(\tilde{L}|\tilde{R})$ against $||\tilde{R}||$, and Fig. 12 shows the scatter plots of the shape parameters of $P(\tilde{L}|\tilde{D})$ against $||\tilde{D}||$. All these plots clearly show the positive correlation between the shape parameters and bandpass range (disparity). These positive correlations are consistent over all orientations and all scales. We list the correlation coefficients in Tables VI and VII. We performed the t-test of significance of the correlations in Tables IV–VII. We found that all of the p values satisfied <0.001, indicating strong positive correlations.

With this newly gained knowledge on the statistical dependencies between 2-D and 3-D scene features in natural scenes, we believe that the human vision system uses these existing statistical regularities to solve problems in depth perception. Toward this end, we use these conditional natural scene statistics (NSS) models to explore one of most fundamental and extensively studied problems to computer vision, visual psychophysics, and visual neuroscience: stereo correspondence.

VII. APPLICATION TO COMPUTED STEREOPSIS

A. Motivation and Scope

Stereo correspondence has been one of the most important questions in computational vision for more than three decades. The history of research on the neuronal mechanisms of biological stereopsis is much longer. Great progress has been made

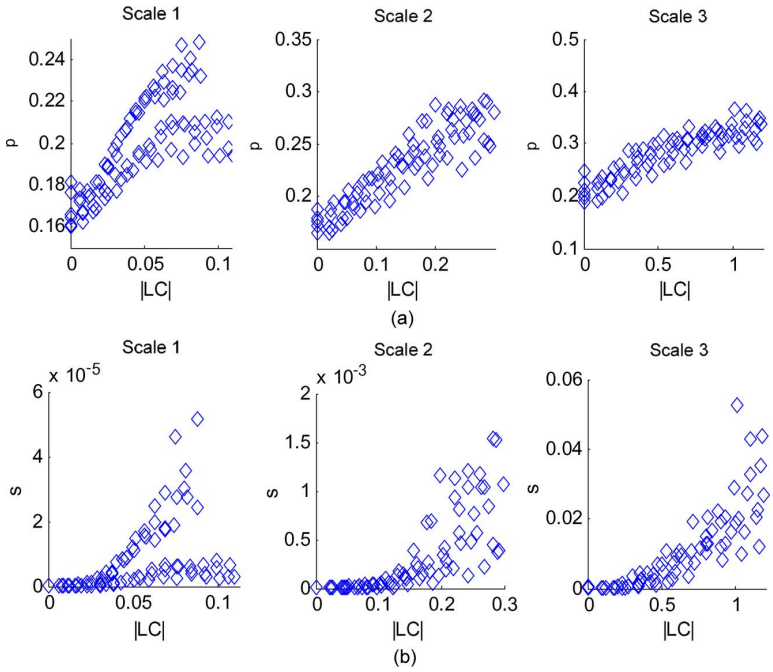

Fig. 10. Shape parameters of generalized Gaussian fits of $P(\bar{D}|\bar{L})$, the conditional distribution of disparity coefficients, are plotted against the magnitudes of luminance coefficients. (a) Shows the scatter plots of p and the magnitudes of luminance coefficients and (b) shows the scatter plots of s and magnitudes of luminance coefficients. Each panel shows p or s and their associated luminance coefficients from 6 oriented subbands within a scale. From left to right, the scale goes from the finest to the coarsest.

TABLE V
CORRELATION COEFFICIENTS BETWEEN SHAPE PARAMETERS (p, s) AND $\|\bar{L}\|$ IN $P(\bar{D}|\|\bar{L}\|)$

Scale	$corr(p, \|\bar{L}\|)$	$corr(s, \|\bar{L}\|)$
1	0.69	0.47
2	0.89	0.76
3	0.85	0.81

in understanding disparity tuning of simple and complex neurons in cortical area V1, but our current knowledge is still local and isolated and does not explain the underlying organization of larger populations of binocular neurons located over many levels of cortical areas and layers [47]. Regarding stereo correspondence, visual neuroscientists and computer scientists deploy different methodologies to approach this question. Neuroscientists won't sway from known properties of binocular neurons. Their interest is to explain their experimental findings on single unit recording, or fMRI, usually given simple and unambiguous binocular stimuli. The building blocks of their stereo correspondence algorithm are models of binocular neurons [34], [48]. Because we have only limited knowledge on the encoding and representation of complicated 3-D natural scenes in the brain, most of the stereo correspondence algorithms designed by visual neuroscientists are not comparable in performance with state-of-the-art stereo algorithms designed by computer scientists. Computer scientists enjoy the freedom to fine-tune their algorithms, regardless of biological compatibility, which is an advantage when dealing with complex, cluttered, and sometimes ambiguous natural scenes. It is beyond our scope here to review a lot of stereo matching algorithms (for this purpose, see [49], [50]). Our implementation is closely related to methods that define and optimize a global energy (cost) function to solve the stereo problem. The energy functions can often be given a Bayesian interpretation of the luminance and depth interaction and depth priors. Generally, these models are based on Markov random field models [51]–[53]. Some methods also provide specific handlings of difficult regions such as occlusions and depth discontinuities [52], [54], or by varying smoothness terms according to color gradients [55]. These methods are indeed close to our approach as they share a Bayesian framework with similarly defined probabilities. However, our approach is rooted in direct and accurate measurements of luminance and range from natural scenes without making *a priori* assumptions on the probabilistic definitions.

In our study, we try to bridge the gap between biologically compatible, neuroscience-based algorithms and heuristic, performance-tuned, artificial algorithms. We assume that the human brain has internal models of visual space. Furthermore, the internal models should be consistent with the statistical properties of natural environments. We hypothesize that relevant and optimal solutions can be achieved by formulating the stereo correspondence problem within a Bayesian framework. Since we have developed quantitative models of luminance and disparity in the previous sections, our aim is to apply our probability models of 3-D natural scenes to Bayesian stereo correspondence. Our goal is to demonstrate the utility of the newly found statistical regularities of 3-D natural scenes using a simple formulation, rather than to compare our results with the most complex and sophisticated stereo matching algorithms. However, we did upload our results to the Middlebury Stereo online evaluation Web site, which can be found at (see http://vision.middlebury.edu/stereo/submit/). Our results ranked 71st among 94 algorithms on the evaluation page.

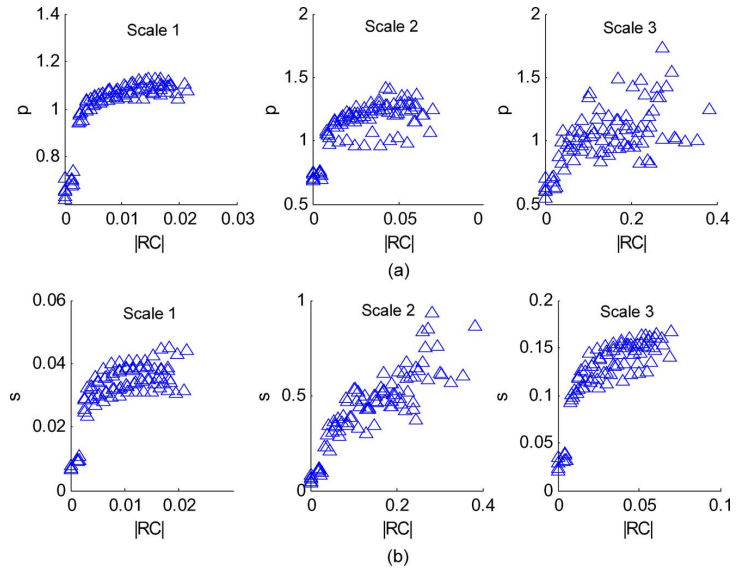

Fig. 11. Shape parameters of generalized Gaussian fits of $P(\tilde{L}|\tilde{R})$, the conditional distribution of luminance coefficients, are plotted against the magnitudes of range coefficients. (a) Shows the scatter plots of p and the magnitudes of $\tilde{R}$ and (b) shows the scatter plots of s and magnitudes of $\tilde{R}$. Each panel shows p or s and their associated $\tilde{R}$ from 6 oriented subbands within a scale. From left to right, the scale goes from the finest to the coarsest.

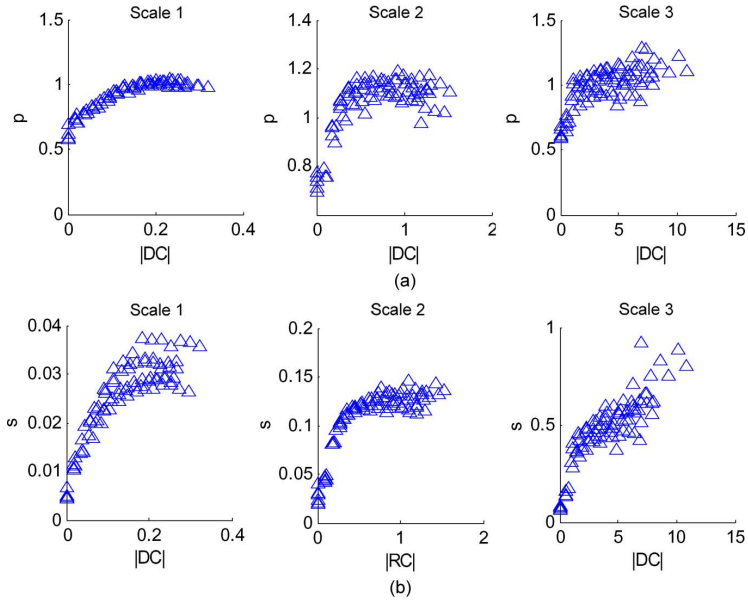

Fig. 12. Shape parameters of generalized Gaussian fittings of $P(\tilde{L}|\tilde{D})$, the conditional distribution of luminance coefficients are plotted against the magnitudes of disparity coefficients. (a) is the scatter plot of p and the magnitudes of $\tilde{D}$ and (b) shows the scatter plots of s and the magnitudes of $\tilde{D}$ from 6 oriented subbands within a scale. From left to right, the scale goes from the finest to the coarsest.

TABLE VI
CORRELATION COEFFICIENTS BETWEEN SHAPE PARAMETERS (p, s) AND $\|\tilde{R}\|$ IN $P(\tilde{L}|\|\tilde{R}\|)$

Scale	$corr(p, \|\tilde{R}\|)$	$corr(s, \|\tilde{R}\|)$
1	0.72	0.71
2	0.73	0.83
3	0.63	0.85

TABLE VII
CORRELATION COEFFICIENTS BETWEEN SHAPE PARAMETERS (p, s) AND $\|\tilde{D}\|$ IN $P(\tilde{L}|\|\tilde{D}\|)$

Scale	$corr(p, \|\tilde{D}\|)$	$corr(s, \|\tilde{D}\|)$
1	0.86	0.85
2	0.66	0.78
3	0.70	0.86

B. Method

We denoted the left image I_l, the right image I_r, and the disparity map D. Many global stereo correspondence algorithms define an energy function [49], [56]. The solution in disparity space that minimizes the energy function is the final disparity map. For example, Barnard [57] defined the energy function as:

$$\begin{aligned} E &= E_{\text{data}} + \lambda E_{\text{smooth}} \\ E_{\text{data}} &= \sum_{(i,j)} |I_l(i,j) - I_r(i, j + D(i,j))| \\ E_{\text{smooth}} &= \sum_{(i,j)} |\nabla D(i,j)| \end{aligned} \tag{6}$$

where E_{data} is a photometric constraint that assumes two correctly matched pixels are similar. E_{smooth} is a smoothness con-

straint that penalizes large changes in local surface neighborhood, and λ balances the contributions from the two terms. These heuristic definitions are generally valid, since they implicitly match the properties of natural scenes: mutually visible points have similar luminance in the left and right image, and the 3-D space is piecewise continuous.

In a Bayesian formulation, Belhumeur [58] showed that E_{data} equals the likelihood of interocular luminance difference, and E_{smooth} equals disparity prior. We adopt Belhumeur's Bayesian framework and modify the likelihood and prior to accommodate our natural scene models.

Suppose the reference image is I_l, then we may assume from what we have learned from our coregistered range-luminance study that I_l contains information about its disparity map D. Denote a subband of wavelet decomposition of I_l as $\tilde{I}_l$. We know that $P(\tilde{D}|\tilde{I}_l)$ can be modeled as generalized Gaussian. In order to try to solve the problem in the wavelet domain, we write the posterior probability of finding the disparity map given the two images as

$$P\left(\tilde{D}|F(\tilde{I}_l,\tilde{I}_r),\tilde{I}_l\right) \sim P\left(F(\tilde{I}_l,\tilde{I}_r)|\tilde{D},\tilde{I}_l\right)P(\tilde{D}|\tilde{I}_l) \quad (7)$$

We assume

$$P\left(F(\tilde{I}_l,\tilde{I}_r)|\tilde{D},\tilde{I}_l\right) = P\left(F(\tilde{I}_l,\tilde{I}_r)|\tilde{D}\right).$$

We make the same Gaussian assumption as in Belhumeur [58].

$$P\left(F(\tilde{I}_l,\tilde{I}_r)|\tilde{D},\tilde{I}_l\right) = \frac{1}{\sqrt{2\pi}\sigma_I}\left(\exp\left(-\frac{1}{2\sigma_I^2}\sum_{(i,j)}(\tilde{I}_l-\tilde{I}_r)^2\right)\right). \quad (8)$$

To further simplify the problem, we assume that bandpass disparities are spatially independent. This is a rather strong assumption, which is not necessarily supported by real distributions of natural scenes. We make this assumption to make this problem tractable. Under the independence assumption.

$$\begin{aligned} P(\tilde{D}|\tilde{I}_l) &= \prod_{(i,j)} P\left(\tilde{D}(i,j)|\tilde{I}_l(i,j)\right) \\ &= \prod_{(i,j)} \frac{1}{z\left(s(i,j),p(i,j)\right)} \\ &\quad \times \exp\left(-\sum_{(i,j)}\left|\tilde{d}(i,j)/s(i,j)\right|^{p(i,j)}\right). \end{aligned} \quad (9)$$

The stereo correspondence problem is formulated as an optimization problem.

$$\arg\max_D\left(P\left(\tilde{D}|F(\tilde{I}_l,\tilde{I}_r),\tilde{I}_l\right)\right) = \arg\max_D(E_{\text{data}}+\lambda E_{\text{smooth}}) \quad (10)$$

where

$$\begin{aligned} E_{\text{data}} &= \sum_{(i,j)}(\tilde{I}_l-\tilde{I}_r)^2 \\ E_{\text{smooth}} &= \sum_{(i,j)}\left|\tilde{D}/s(i,j)\right|^{p(i,j)} \\ \lambda &= 2\sigma_I^2. \end{aligned} \quad (11)$$

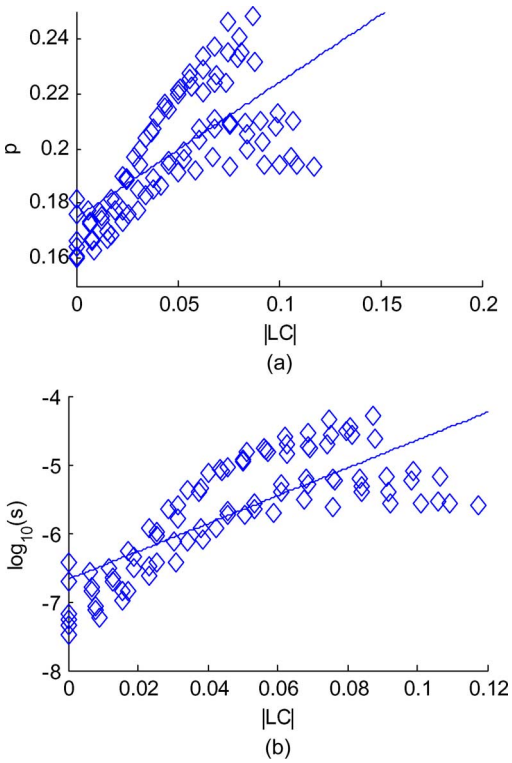

Fig. 13. We used a linear model to estimate the relationship between p and bandpass luminance $\|\tilde{L}\|$. (a) Shows the linear fit of p against $\|\tilde{L}\|$ and (b) shows the linear fit of $\log_{10}(s)$ against $\|\tilde{L}\|$.

Fig. 14. Left to right: one sample image, the map of shape parameter p, the map of shape parameter s. Both p and s are estimated from one orientation subband (diagonal). It is apparent that at large luminance changes along diagonal directions, p and s also tend to be large. In our implementation, we use a four orientation decomposition.

Given (10), we deploy simulated annealing as in Barnard [57] to solve the optimization. The major difference between our method and Barnard's method is in the definition of the term E_{smooth}. In our approach, we modify the penalty on disparity changes adaptively by changing the shape parameters p and s. Given the strong positive correlation between (p,s) and bandpass luminance $\|\tilde{L}\|$, at regions with large $\|\tilde{L}\|$, the parameters p and s are set such that the penalty on large disparity changes are attenuated. But at regions having small luminance variations, the values of p and s intensify the penalty on large disparity changes.

By observing the plots in Fig. 11, we use a linear model to estimate $p(i,j)$ from bandpass luminance $\|\tilde{L}(i,j)\|$ at a location (i,j). We also model $log_{10}(s)$ and $\|\tilde{L}\|$ as linear by observing the scatter plot. Fig. 13 shows the linear fits and the scatter plots at one scale. The spatially variant shape parameters (p,s) are displayed as images in Fig. 14. The values of p and s are estimated from the finest scale.

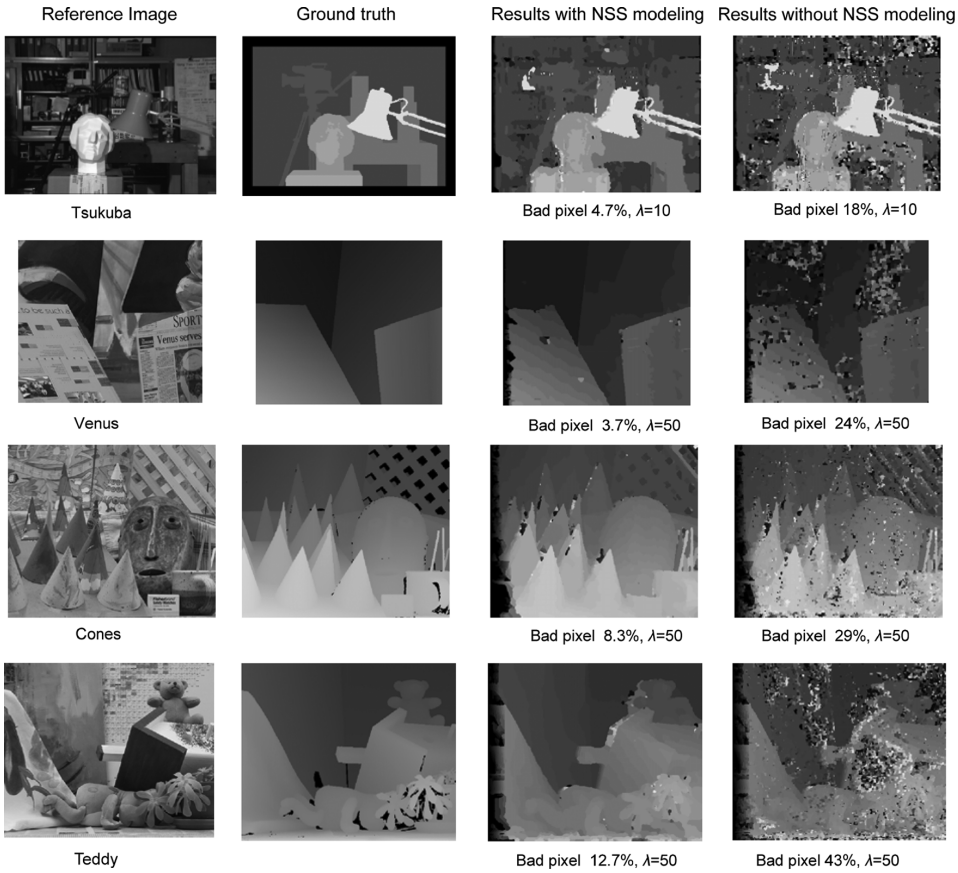

Fig. 15. Left to right: the reference images, the ground truth disparity maps, the computed disparity maps using NSS modeling and the computed disparity maps with energy function defined as in (10).

C. Implementation and Results

The objective function (10) is defined on one subband. To optimize over all subbands, we sum (10) over all bands and minimize:

$$\arg\min_{D}\left(\sum_{m}\sum_{l}\left(E_{\text{data}}(l,m)+\lambda E_{\text{smooth}}(l,m)\right)\right) \quad (12)$$

where l is the subband orientation, and m is the subband scale. We adapted the simulated annealing implementation algorithm from the Middlebury stereo dataset [50] to implement our energy function defined in (12) based on 3-D natural scene statistics.

We set the initial temperature at 200, and the maximum iteration to 5000. We tested our algorithm on four stereo pairs with ground truth from the Middlebury stereo dataset [50]. For comparison, we also ran the simulation using the classical energy definition [as in (10)] developed by Barnard [57].

Another important parameter is λ: the relative weight between E_{data} and E_{smooth}. We know that λ is actually the variance of the interocular difference between matched pixels from (20). We want to emphasize that it is "*matched*" pixels, rather than "*mutually visible*" pixels. In Belhumeur's *World I* definition [58], all the "*matched*" pixels are actually "*mutually visible*," which means there are no occlusions or disparity discontinuities. In Belhumeur's case, the interocular difference between two mutually visible pixels is very small, and depends on the differences between two optical imaging processes and the orientation differences of the reflectance of a spatial point in space. However, there are many occluded areas in the four test stereo pairs (also in natural environments). Because our algorithm does not have occlusion detection, all of the pixels in the reference images are matched with pixels in the other images without distinction. The variance of interocular difference between two matched pixels depends on the scene content: more occluded areas yield larger variance, and more mutually visible areas yield smaller variances. Since there is no easy way to find the best λ theoretically, we tested a few values of λ ($\lambda = 1$, 5, 10, 50, 100, 500), to generally illustrate the results.

Fig. 15 shows the reference (left) images of each stereo pair, the ground truth disparity maps, the computed disparity maps with generalized Gaussian modeling of $P(\tilde{D}|\tilde{I}_l)$, and the computed disparity maps using Barnard's basic energy definition [57]. It is visually obvious that our algorithm yields much better disparity maps than the original energy function definition in (6). Using the evaluation criteria included with the Middlebury stereo software [50], our approach significantly improved the Belhumeur algorithm in terms of "bad pixel rate" $(\text{disparity errors} > \pm 1 \text{ pixels})$. On the stereo pairs Tsukuba, Venus, Cones, and Teddy using the energy definition (10) the error rates were 18%, 24%, 29%, and 43%, whereas our

algorithm improved the bad pixel rates to 4.7%, 3.7%, 8.3%, and 12.7%, respectively.

We believe that this experiment demonstrates the validity of utilizing monocular luminance–disparity dependencies that occur in the natural environment for helping to resolve stereo correspondences. Other related and complementary work has yielded similar conclusions.

For example, in an elegant framework, Scharstein and Pal [59] use machine learning techniques to learn conditional random fields from a different set of calibrated ground truth disparity maps computed via structured-light methods. The authors report good results. Their work differences from ours in that we deploy closed-form parametric models of conditional distributions in wavelet space to describe luminance–range relationships, which, if accurate, leads to simpler development and algorithm design. Our results suggest that our conditional models are quite accurate.

The development of new ground truth stereo databases is timely, especially in view of the ascendant popularity of 3-D displays and presentations. Two factors are relevant here: larger databases (more stereo pairs with ground truth), thus reducing the possibility of "training" ones algorithm (inadvertently or otherwise) to achieve optimal performance specific to a particular data set, and better ground truth, arrived at by a modern sensor, such as a lidar-based terrestrial scanner.

VIII. Conclusion

Using a public coregistered database of luminance and range natural images, we examined the priors and conditional distributions of low level, fundamental image features such as luminance, disparity, and distance. Similar to the well-known properties of wavelet subband histograms of luminance images, the distributions of range and disparity subband coefficients tend to have much higher kurtosis than luminance coefficients, owing to a greater degree of regularity in range and disparity maps as compared to luminance images. Generalized Gaussians generally give good fits to marginal distributions of luminance, range, and disparity wavelet coefficients, but the shape parameter p associated with marginal luminance distributions ($\sim$[0.6, 0.8]) tends to be significantly larger than those for marginal range and disparity distributions ($\sim$[0.2, 0.3]).

We found that the conditional magnitudes of luminance and range (disparity) coefficients mutually depend on each others' magnitudes. Generally, regions with larger luminance variation tend to have larger range (disparity) variation and vice versa. Our analysis also shows that the correlations between bandpass luminance and bandpass range (disparity) are stronger in coarser scales. The shape parameters (p, s) of the conditionals display a clear dependency on the scene features that are conditioned on.

We developed a stereo correspondence algorithm based on our statistical models of 3-D natural scenes. Using a Bayesian framework, we showed that adaptively changing the smoothness cost at different luminance variation can improve the quality of the computed disparity maps. We believe that statistics of this type will also prove useful in other 3-D vision and image processing applications, such as shape-from-X (shading, texture, etc), 3-D recognition, and 3-D (stereo) image quality assessment.

Acknowledgment

The authors would like to thank B. Potetz and T. S. Lee for the kind use of their range map database.

References

[1] Ruderman and Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, no. 6, pp. 814–817, Aug. 1994.

[2] P. J. Burt and E. H. Adelson, "The Laplacian Pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.

[3] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 12, pp. 2379–2394, Dec. 1987.

[4] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[5] 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.: Publicat. IEEE Signal Process. Soc.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

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

[7] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Time-Freq. Time-Scale Anal., 1998. Proc. IEEE-SP Int. Symp.*, 1998, pp. 633–636.

[8] P. Scheunders and S. De Backer, "Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1865–1872, Jul. 2007.

[9] D. Hammond and E. Simoncelli, "Image modeling and denoising with orientation-adapted gaussian scale mixtures," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2089–2101, Nov. 2008.

[10] J. A. Guerrero-Colón, L. Mancera, and J. Portilla, "Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids," *IEEE Trans. Image Process.: Publicat. IEEE Signal Process. Soc.*, vol. 17, no. 1, pp. 27–41, Jan. 2008.

[11] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[12] D. Mumford and B. Gidas, "Stochastic models for generic images," *Quart. Appl. Math*, vol. 59, pp. 85–111, 1999.

[13] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale invariant dead leaves model," *Int. J. Comput. Vis.*, vol. 19, pp. 35–59, 2001.

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

[15] Z. Yang and D. Purves, "Image/source statistics of surfaces in natural scenes," *Network (Bristol, England)*, vol. 14, no. 3, pp. 371–390, Aug. 2003.

[16] C. Q. Howe and D. Purves, "Range image statistics can explain the anomalous perception of length," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 20, pp. 13184–13188, Oct. 2002.

[17] 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.

[18] Z. Yang and D. Purves, "A statistical explanation of visual space," *Nat. Neurosci.*, vol. 6, no. 6, pp. 632–640, Jun. 2003.

[19] B. Potetz and T. S. Lee, "Scaling Laws in Natural Scenes and the Inference of 3D Shape," in *Advances in Neural Information Processing Systems 18*. Cambridge, MA: MIT Press, 2006.

[20] S. Roth and M. J. Black, "On the spatial statistics of optical flow," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 33–50, 2007.

[21] I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis*, 1st ed. New York, USA: Oxford University Press, 1995.

[22] C. Blakemore, "The range and scope of binocular depth discrimination in man," *J. Physiol.*, vol. 211, no. 3, pp. 599–622, Dec. 1970.
[23] 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.
[24] 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.
[25] 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.
[26] 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.
[27] 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.
[28] Y. Liu, L. K. Cormack, and A. C. Bovik, "Relationship between the Helmholtz shear of vertical meridians and disparity statistics in natural scenes [Abstract]," *J. Vis.*, vol. 8, no. 6, p. 846, May 2008.
[29] E. B. Johnston, B. G. Cumming, and A. J. Parker, "Integration of depth modules: Stereopsis and texture," *Vis. Res.*, vol. 33, no. 5, pp. 813–826, Apr. 1993.
[30] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: In defense of weak fusion," *Vis. Res.*, vol. 35, pp. 389–412, Feb. 1995.
[31] D. C. Knill, "Mixture models and the probabilistic structure of depth cues," *Vis. Res.*, vol. 43, no. 7, pp. 831–854, Mar. 2003.
[32] D. C. Knill, "Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception," *J. Vis.*, vol. 5, no. 2, pp. 103–115, Feb. 2005.
[33] A. E. Welchman, A. Deubelius, V. Conrad, H. H. Bülthoff, and Z. Kourtzi, "3-D shape perception from combined depth cues in human visual cortex," *Nature Neurosci.*, vol. 8, no. 6, pp. 820–827, Jun. 2005.
[34] J. C. A. Read and B. G. Cumming, "Testing quantitative models of binocular disparity selectivity in primary visual cortex," *J. Neurophysiol.*, vol. 90, no. 5, pp. 2795–2817, Nov. 2003.
[35] S. J. Prince and R. A. Eagle, "Weighted directional energy model of human stereo correspondence," *Vis. Res.*, vol. 40, no. 9, pp. 1143–1155, 2000.
[36] H. V. Helmholtz, *Helmholtz's Treatise on Physiological Optics*, 1924th ed. London, U.K.: Thoemmes Continuum, 2000.
[37] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. 1995 Int. Conf. Image Process. (Vol. 3)-Volume 3—Volume 3*, 1995, p. 3444.
[38] D. J. Field, "What is the goal of sensory coding?," *Neural Comput.*, vol. 6, no. 4, pp. 559–601, 1994.
[39] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *Int. J. Comput. Vis.*, vol. 41, pp. 35–59, 2001.
[40] Y. Zhang, M. Loew, and R. Pickholtz, "A generalized Gaussian autoregressive model for chest X-ray images," in *Eng. Med. Biol. Soc., 1989. Images 21st Century., Proc. Annu. Int. Conf. IEEE Eng.*, 1989, vol. 2, pp. 361–362.
[41] S. S. Saquib, C. A. Bouman, and K. Sauer, "Efficient ML estimation of the shape parameter for generalized Gaussian MRFs," in *Proc. Acoust., Speech, Signal Process., 1996. Conf. Proc., 1996 IEEE Int. Conf.—Volume 04*, 1996, pp. 2227–2230.
[42] D. Gonzalez-Jimenez, F. Perez-Gonzalez, P. Comesana-Alfaro, L. Perez-Freire, and J. Alba-Castro, "Modeling gabor coefficients via generalized Gaussian distributions for face recognition," in *IEEE Int. Conf. Image Processing*, 2007, pp. IV-485–IV-488.
[43] M. Z. Coban and R. M. Mersereau, "Adaptive subband video coding using bivariate generalized Gaussian distribution model," in *Proc. Acoust., Speech, Signal Process., 1996 Conf. Proc., 1996 IEEE Int. Conf.—Volume 04*, 1996, pp. 1990–1993.
[44] K. Kokkinakis and A. K. Nandi, "Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling," *Signal Process.*, vol. 85, no. 9, pp. 1852–1858, 2005.
[45] 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.
[46] B. Julesz, *Foundations of Cyclopean Perception*. Chicago, IL: Univ. Chicago Press, 1971.
[47] B. G. Cumming and G. C. DeAngelis, "The physiology of stereopsis," *Annu. Rev. Neurosci.*, vol. 24, pp. 203–238, 2001.
[48] N. Qian, "Computing stereo disparity and motion with known binocular cell properties," *Neural Comput.*, vol. 6, no. 3, pp. 390–404, May 1994.
[49] M. Brown, D. Burschka, and G. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, Aug. 2003.
[50] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7–42, 2002.
[51] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
[52] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 787–800, Jul. 2003.
[53] D. Scharstein, "Learning conditional random fields for stereo," in *CVPR*, 2007, pp. 1–8.
[54] R. Ben-Ari and N. Sochen, "Stereo matching with Mumford-Shah regularization and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2071–2084, Nov. 2010.
[55] J. J. Weinman, L. Tran, and C. J. Pal, "Efficiently learning random fields for stereo vision with sparse message passing," in *Proc. 10th Eur. Conf. Comput. Vis.: Part I*, 2008, pp. 617–630.
[56] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, no. 1, pp. 7–42, 2002.
[57] S. T. Barnard, "A stochastic approach to stereo vision," in *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*. San Mateo, CA: Morgan Kaufmann, 1987, pp. 21–25.
[58] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
[59] P. N. Belhumeur, "A Bayesian approach to binocular stereopsis," *Int. J. Comput. Vis.*, vol. 19, no. 3, pp. 237–260, 1996.

Yang Liu received the B. Eng. degree from Zhejiang University, Hangzhou, China, in 1998, the M.S. degree in computer science from the University of Regina, Regina, Canada, in 2002, and the Ph.D. degree from University of Texas at Austin, TX, in Summer 2010.

He joined the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin in the Fall of 2002. He is currently a Postdoctoral Fellow, learning color perception under Prof. Bevil Conway, in Wellesley College, Wellesley, MA, and Harvard Medical School, Boston, MA. His research interests include 2D/3-D natural scene statistics, computational aspects of biological stereo vision, and visual neuroscience for color perception.

Lawrence K. Cormack is from Socorro, N.M. He received the B.S. degree (with Highest Honors) in psychology from the University of Florida, Gainesville, FL, in 1986 and the Ph.D. degree in physiological optics (now vision science) from the University of California at Berkeley, CA, in 1992.

He joined the Psychology Department at the University of Texas (U.T.) at Austin, in the Fall of 1992 and has been there ever since. At U.T., he is also currently an active member of the Center for Perceptual Systems, the Institute for Neuroscience, and the Imaging Research Center. His research interests include primarily in the brain mechanisms underlying vision, including the basic questions of 1) the way depth and motion are perceived (i.e., dynamic 3-D perception) and 2) the role that the structure of the natural environment has played in the evolution of the visual system.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1980, 1982, and 1984, respectively.

He is currently the Curry/Cullen Trust Endowed Chair Professor at the University of Texas at Austin, where he is also the Director of the Laboratory for Image and Video Engineering (LIVE), and also a faculty member in the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute for Neuroscience. He is the author or coauthor of more than 500 technical papers. He holds two U.S. patents. He is the author of The Handbook of Image and Video Processing (New York, Academic, 2005), Modern Image Quality Assessment (Morgan & Claypool, 2006), and two new books, The Essential Guide to Image Processing and The Essential Guide to Video Processing (New York, Academic). His research interests include image and video processing, computational vision, and visual perception.

Dr. Bovik is a Fellow of the Optical Society of America, and the Society of Photo-Optical and Instrumentation Engineers. He has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996-1998; Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996-2002; Editorial Board, THE PROCEEDINGS OF THE IEEE, 1998-2004; Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003-present; and Founding General Chairman, First IEEE International Conference on Image Processing, Austin, Texas, in November, 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions. 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 has also received 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 was recently named the IS&T/SPIE Imaging Scientist of the Year for 2011. He is also a recipient of the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin.