Statistical Model of Color and Disparity with Application to Bayesian Stereopsis

Che-Chun Su[1], Alan C. Bovik[2], and Lawrence K. Cormack[3]

Department of Electrical and Computer Engineering[1,2], Department of Psychology[3]
The University of Texas at Austin
Austin, Texas
chechunsu@mail.utexas.edu[1], bovik@ece.utexas.edu[2], cormack@psy.utexas.edu[3]

Abstract—**Extensive research has been conducted relating the natural scene statistics of luminance and depth; however, very little work has been done on analyzing the statistical relationships between depth and chromatic information. In this paper, we examine and derive statistical models between disparity and both luminance and chrominance information by transforming natural images into the more perceptually relevant CIELAB color space. To demonstrate the effectiveness of these models, we further exploit them with application to Bayesian stereo algorithms. The simulation results show that incorporating the derived statistical models augments the performance of Bayesian stereo algorithms. In addition, these results also support psychophysical evidence that chromatic information can improve binocular visual processing.**

Keywords-**Natural scene statistics, color, disparity, Bayesian stereo algorithm**

I. INTRODUCTION

Natural scene statistics (NSS) are important factors both towards understanding the evolution of the human vision system and for designing image processing algorithms [1], [2]. Extensive research has been conducted to explore the link between NSS and neural processing of visual stimuli [3], [4]. With the increasing popularity of 3D image and video content, the statistics between 3D depth and 2D color image data in natural scenes are of high interest. However, very little work has been done due to the limited access to high quality databases of color images and associated ground-truth range maps.

Potetz et al. [5] 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. In [6], Yang et al. explored the statistical relationships between luminance and disparity in the wavelet domain, and applied the derived models to a Bayesian stereo algorithm. In addition, it has been suggested that the perception of color and depth are related [7], and chromatic information can be useful in solving stereo correspondence problems [8]. Su et al. [9] constructed a large co-registered database of high-quality 2D color images and high-resolution ground-truth range maps (1280x720), and explored the statistical relationships between the band-pass responses of luminance/chrominance and range gradients in natural scenes.

In this paper, we examine the statistical relationships between disparity and luminance/chrominance information in natural scenes, and derive statistical models for their joint distributions. We also exploit the derived statistical descriptions of disparity and luminance/chrominance by applying them to the Bayesian stereo problem. The simulation results show that the Bayesian stereo algorithm incorporating the proposed NSS models can improve upon a previous NSS-based stereo algorithm using only luminance information. The results also suggest that the statistics between color and range in natural environments could be helpful in binocular visual processing and depth perception of human vision system.

The rest of the paper is organized as follows. Section II briefly describes the acquisition and pre-processing of the image and range data. Then, the analysis and derivation of statistical models of the marginal and conditional distributions between luminance/chrominance and disparity are included in Section III. Next, Section IV explains how to apply the derived statistical models of luminance/chrominance and disparity to the Bayesian stereo algorithm, followed by simulation results in Section V. Finally, Section VI gives the conclusion.

II. DATA ACQUISITION AND PRE-PROCESSING

We obtained a dataset of 2D color images and co-registered ground-truth range maps, then converted the range maps into disparity maps. The large co-registered database of range and color images (dubbed LIVE Color+3D Database) consists of 12 sets of high-resolution color images and co-registered range maps [9], [10]. 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 [11]. Calibration was performed using the scanner operation software, RIEGL RiSCAN PRO, to compensate for inevitable translational and rotational shifts when mounting the camera onto the range scanner [12]. Then, to obtain the aligned 2D range map with the 2D color image, the 3D point clouds captured by the scanner were projected and transformed into the 2D range map by applying the pinhole camera model with lens distortion [13].

To convert ground-truth range maps into disparity maps, the parallel-viewing model was used. Figure 1 shows the geometry of the parallel-viewing model, where two scanners mounted with digital cameras were set to capture the natural scene in parallel. Based on this geometry, the disparity value is derived:

$$\frac{d_p}{f_c} = \frac{d_{io}}{R} \Rightarrow d_p = f_c \frac{d_{io}}{R} \tag{1}$$

where d_p is disparity, f_c is the focal length of the camera, d_{io} is the inter-ocular distance, and R is the ground-truth range value.

Since we want to learn and explore the statistical relationships between luminance/chrominance and disparity and how these statistics might be implicated in visual processing, some pre-processing was performed on both the 2D color images and the converted 2D disparity maps. All color images were transformed into the more perceptually relevant CIELAB color space with one luminance (L*) and two chrominance (a* and b*) components. The images were then passed through 2D Gabor filter bank with multi-scales and orientations as a model of the reception fields of the simple cells in V1 areas of human vision systems [14]. Both the luminance and chrominance components of the transformed color images and the converted disparity maps were filtered by the 2D Gabor filter banks, and the analysis was performed on these filter responses. For luminance and chrominance data, the 2D Gabor filter bank closely models image decomposition in primary visual cortex, while for disparity data, it can mimic the disparity modulation corresponding to
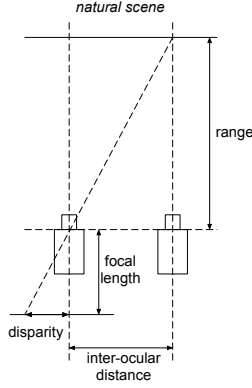
Fig. 1. Geometry of the parallel-viewing model.

Fig. 2. The marginal distribution of disparity magnitude at one sub-band.

sensitivity to range in the human vision system [15], [16]. In general, the complex 2D Gabor filters can be written in the form

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

where $R_1 = x\cos\theta + y\sin\theta$ and $R_2 = -x\sin\theta + y\cos\theta$, σ_x and σ_y are the standard deviations of an elliptical Gaussian envelope along x and y axes, ζ_x and ζ_y are the spatial center frequencies of the complex sinusoidal grating, and θ is the orientation.

III. DATA ANALYSIS AND MODELING

Since we want to derive statistical models between disparity and color image data, and further apply them to Bayesian stereo algorithms, both the marginal distribution of disparity and the conditional distribution of luminance/chrominance given disparity in natural scenes are of the most interest to us. To capture demodulated signal information, analysis was performed on the magnitude of the Gabor responses expressed as the square root of the sum of the squared complex quadrature pair of sine and cosine responses [15].

A. Marginal Distribution of Disparity

To examine the marginal distribution of disparity at different sub-bands, the disparity responses at each sub-band were first collected across all scenes in the database. Then, the marginal distribution of disparity at one particular sub-band was obtained as the histogram computed by binning the magnitude of all disparity responses at that sub-band. Figure 2 shows the marginal distribution of the magnitude of disparity responses at one sub-band. The red-dotted line represents a Weibull distribution fit of the marginal distribution of disparity. We can see that the exponential-like distribution of disparity can be well fitted by the very general Weibull distribution,

$$p(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} & , x \geq 0 \\ 0 & , x < 0 \end{cases} \quad (3)$$

where α and β represent the scale and shape parameters, respectively, which includes the exponential distribution and the Rayleigh distribution as special cases depending on the shape parameter. Note that the marginal distribution of disparity at different sub-bands share similar shapes, all well fitted by the Weibull distributions.
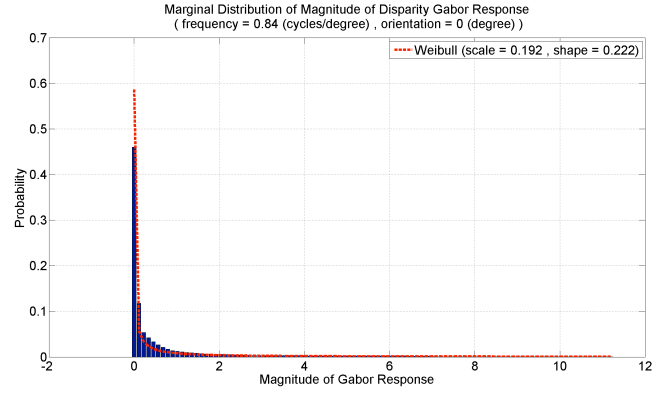
B. Conditional Distributions of Luminance/Chrominance given Disparity

Similarly, the conditional distributions of luminance and chrominance given disparity at different sub-bands are obtained by first collecting both the luminance, chrominance, and disparity responses at each sub-band across all scenes, then computing the histograms for each sub-band. At one particular sub-band, the histograms for the conditional distributions of luminance and chrominance given disparity were computed by first binning the magnitude of disparity responses, and then binning the magnitude of luminance/chrominance responses within each bin of the disparity response. Figure 3 shows the conditional distributions of all three luminance and chrominance components, as well as their corresponding Weibull-fitted distributions and parameters. Both the conditional distributions of luminance and chrominance given disparity are well-fitted by the Weibull distribution. For the conditional distribution of luminance given disparity, the scale parameter of the fitted Weibull increases monotonically and linearly as the magnitude of disparity response increases, while the shape parameters remain constant across the magnitudes of disparity responses. For the chrominance conditional distribution, both the scale and shape parameters possess a nearly linear relationship with the magnitude of disparity responses. These Weibull models and their parameters can be applied to Bayesian stereo algorithm, which is described in the next section.

IV. APPLICATION TO BAYESIAN STEREO ALGORITHMS

Given a pair of left and right images, a binocular stereo algorithm is able to compute a disparity map from one image to the other. The basic idea is to minimize some energy function which involves the differences of several binocular cues between the left and right images within an optimization framework [17]. The Bayesian stereo algorithm adopts the likelihood (conditional distribution) and the prior (marginal distribution) of natural scene statistics (NSS) within the energy function to be minimized, thus forcing the solution towards fitting the statistical relationships between luminance, chrominance, and disparity data in natural scenes derived in Section III. Given a pair of left and right images, I_l and I_r, then to estimate the disparity map, D, from the right to the left image, the canonical Bayesian stereo formulation takes the form [18]

$$D = \underset{D'}{\arg\max}\, P(D'|(I_l, I_r))$$
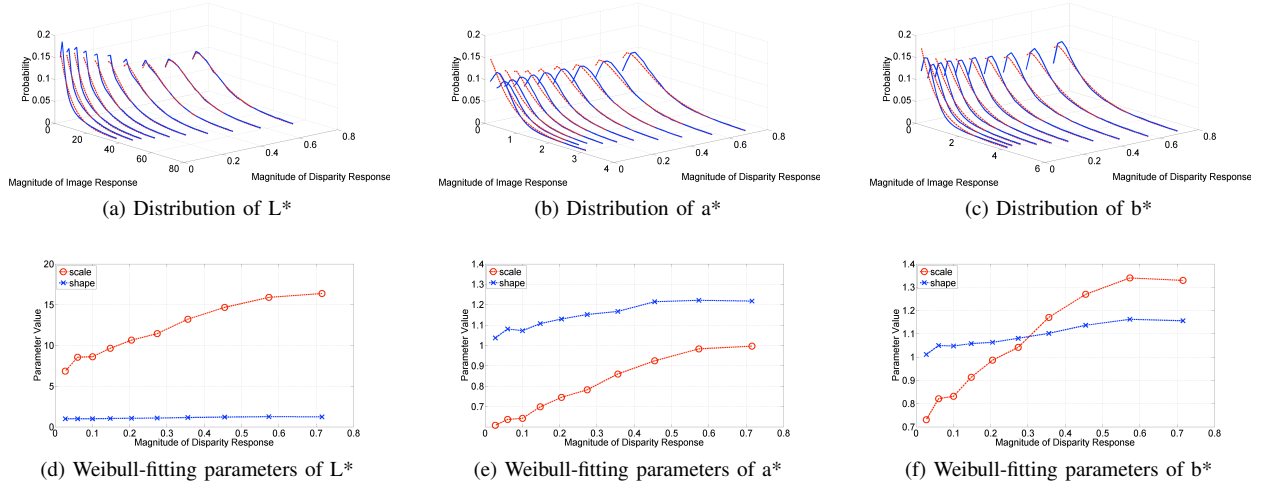$$= \underset{D'}{\arg\max}\, P((I_l, I_r)|D')P(D') \quad (4)$$

Fig. 3. The conditional distributions of luminance (L*) and chrominance (a* and b*) Gabor magnitudes given disparity Gabor magnitude at one sub-band, and their corresponding Weibull-fitting parameters.

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. By taking logarithm of the product of the likelihood and prior, the Bayesian formulation corresponds to minimization of the energy function:

$$D = \underset{D'}{\operatorname{argmin}} E_p + \lambda E_s \qquad (5)$$

where E_p is the photometric energy derived from the likelihood $P((I_l, I_r)|D')$, E_s is the smoothness term derived from the prior $P(D')$, and λ is the weighting constant. Note that E_p can take all three luminance and chrominance components, L^*, a^*, and b^*, into consideration, and can be written as

$$E_p = \sum_{i,j} \sum_{k \in \{L^*, a^*, b^*\}} |I_{l_k}(i, (j - D'(i,j))) - I_{r_k}(i,j)| \qquad (6)$$

To incorporate the derived NSS models of the marginal and conditional distributions, the Bayesian stereo formulation can be re-written as

$$D = \underset{D'}{\operatorname{argmax}} P(\tilde{D}'|(I_l, I_r), \tilde{I}_l)$$

$$= \underset{D'}{\operatorname{argmax}} P((I_l, I_r)|\tilde{D}', \tilde{I}_l) P(\tilde{I}_l|\tilde{D}') P(\tilde{D}') \qquad (7)$$

$$= \underset{D'}{\operatorname{argmin}} E_p + \lambda (E_{NSS_c} + E_{NSS_m}) \qquad (8)$$

by taking logarithm of (7), where $\tilde{I}_l$ and $\tilde{D}'$ are the magnitudes of Gabor 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 the energy terms related to the conditional and marginal distributions derived from natural scene statistics, respectively, and λ is the weighting constant.

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

$$D = \underset{D'}{\operatorname{argmin}} \sum_{i,j} \left[\sum_{k \in \{L^*, a^*, b^*\}} \left(E_{p,k} + E_{NSS_{c,k}} \right) + \lambda E_{NSS_m} \right] \qquad (9)$$

where

$$E_{p,k} = |I_{l_k}(i, (j - D'(i,j))) - I_{r_k}(i,j)| \qquad (10)$$

$$E_{NSS_{c,k}} = \left(\frac{\tilde{I}_{l_k}(i,j)}{\alpha_k} \right)^{\beta_k} \qquad (11)$$

$$E_{NSS_m} = \left(\frac{\tilde{D}'(i,j)}{\alpha_{\tilde{D}'}} \right)^{\beta_{\tilde{D}'}} \qquad (12)$$

where α_k and β_k are the scale and shape parameters, respectively, of the fitted Weibull distributions of luminance and chrominance Gabor magnitudes conditioned on disparity Gabor magnitude, $\alpha_{\tilde{D}'}$ and $\beta_{\tilde{D}'}$ are the scale and shape parameters of the fitted Weibull distribution of disparity Gabor magnitude, respectively, and λ is the weighting constant.

To solve the optimization of the proposed Bayesian stereo algorithm, we use the simulated annealing algorithm utilizing the derived energy function (9) [19].

V. SIMULATION RESULTS

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 different formulations and models, including the canonical formulation using (5), the NSS model proposed in [6], and the proposed Gabor-based NSS model using (9). In [6], the authors derived the 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. Figure 4 shows the computed disparity maps of *Tsukuba* from the Middlebury database [17]. We can see that the computed disparity map with the proposed Gabor-based NSS model is very close to the ground-truth disparity map, retaining more details than the one computed by the canonical formulation, and better adherence to smooth regions than the one computed by the previous NSS model. In addition, Table I gives a numerical comparison in terms of bad-pixel rate between the computed disparity maps and the ground-truth disparity maps for all four test image pairs available from the Middlebury database. The numerical results support the visual comparison, where the Bayesian stereo algorithm with the proposed Gabor-based NSS model outperforms the other two methods.

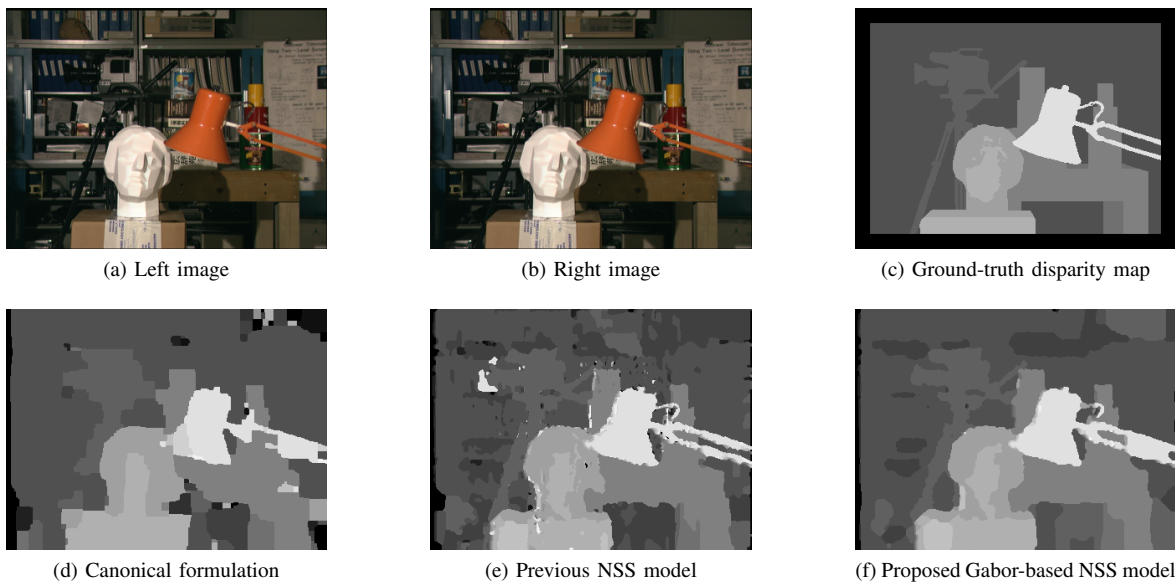| (a) Left image | (b) Right image | (c) Ground-truth disparity map |
| (d) Canonical formulation | (e) Previous NSS model | (f) Proposed Gabor-based NSS model |

Fig. 4. The simulation results of *Tsukuba* from the Middlebury database, including the original stereo image pair, the ground-truth disparity map, and the computed disparity maps using the Bayesian stereo algorithm with different formulations and models.

TABLE I
BAD-PIXEL RATE (%)

	Canonical Formulation	NSS Model in [6]	Proposed Gabor-based NSS Model
Tsukuba	4.8	4.7	**3.3**
Venus	9.8	3.7	**1.2**
Cones	29.5	8.3	**8.1**
Teddy	43.5	12.7	**12.3**

VI. CONCLUSION

We examined the marginal distribution of disparity and the conditional distributions of luminance/chrominance information given disparity in natural scenes. We modeled both the marginal and conditional distributions as Weibull probability density function. To demonstrate the efficacy of the statistical models, we deployed the probability distributions as energy priors in a Bayesian stereo algorithm. The simulation results show that the Bayesian stereo algorithm incorporating the proposed color+disparity NSS model outperforms the canonical formulation and the previous NSS model. The derived statistical models relating luminance/chrominance and disparity information in natural scenes not only improve the accuracy of the Bayesian stereo algorithm, but also yield insight into how 3D structures in the environment might be recovered from color image data. More importantly, these statistical models and simulation results bolster the psychophysical evidence that chromatic information can be useful in 3D visual processing. Future research in the stereo image and video processing fields may also benefit from statistical models relating color and disparity in natural scenes.

REFERENCES

[1] B. A. Olshausen and D. J. Field, "Vision and the coding of natural images," *American Scientist*, vol. 88, p. 238, 2000.

[2] Z. Wang and A. Bovik, "Reduced- and no-reference image quality assessment: The natural scene statistic model approach," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[3] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, no. 1, pp. 1193–1216, Mar. 2001.

[4] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annual Review of Psychology*, vol. 59, no. 1, pp. 167–192, Jan. 2008.

[5] B. Potetz and T. S. Lee, "Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes," *Journal of the Optical Society of America A*, vol. 20, no. 7, pp. 1292–1303, Jul. 2003.

[6] 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.

[7] J. R. Jordan, W. S. Geisler, and A. C. Bovik, "Color as a source of information in the stereo correspondence process," *Vision research*, vol. 30, no. 12, pp. 1955–1970, 1990.

[8] J. R. Jordan and A. C. Bovik, "Using chromatic information in dense stereo correspondence," *Pattern Recognition*, vol. 25, no. 4, pp. 367–383, Apr. 1992.

[9] C.-C. Su, A. C. Bovik, and L. K. Cormack, "Natural scene statistics of color and range," *IEEE Int. Conf. Image Processing*, Sep. 2011.

[10] C.-C. Su, L. K. Cormack, and A. C. Bovik, "LIVE Color+3D Database," http://live.ece.utexas.edu/research.

[11] RIEGL, "RIEGL VZ-400 3D Terrestrial Laser Scanner," http://rieglusa.com/products/terrestrial/vz-400/index.shtml.

[12] ——, "RIEGL RiSCAN PRO Software for 3D Terrestrial Laser Scanner," http://rieglusa.com/products/terrestrial/vz-400/software.shtml.

[13] Intel Corporation, "OpenCV: Camera Calibration and 3D Reconstruction," http://opencv.willowgarage.com/documentation/camera_calibration_and_3d_reconstruction.html.

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

[15] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[16] J. Read, "Early computational processing in binocular vision and depth perception," *Progress in Biophysics and Molecular Biology*, vol. 87, no. 1, pp. 77–108, 2005.

[17] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002.

[18] P. N. Belhumeur, "A bayesian approach to binocular steropsis," *International Journal of Computer Vision*, vol. 19, pp. 237–260, 1996.

[19] S. T. Barnard, "A stochastic approach to stereo vision," *Proc. of the 5th Nat. Conf. on Artificial Intelligence, AAAI*, vol. 1, pp. 676–680, Aug. 1986.