NEW BIVARIATE STATISTICAL MODEL OF NATURAL IMAGE CORRELATIONS

Che-Chun Su¹ Lawrence K. Cormack² Alan C. Bovik¹

¹Department of Electrical and Computer Engineering, The University of Texas at Austin ²Department of Psychology, The University of Texas at Austin

ABSTRACT

We perform bivariate statistical analysis and modeling of the joint distributions of spatially adjacent sub-band responses for both luminance/chrominance and range data in natural scenes. In particular, we introduce a multivariate generalized Gaussian distribution and an exponentiated sine function to model the underlying statistics and correlations. The experimental results show that the bivariate statistics relating spatially adjacent pixels in both 2D color images and range maps are well described by the proposed models. We validate the robustness of the proposed bivariate models using a multivariate statistical hypothesis test, and further demonstrate their effectiveness with application to a prototype depth estimation algorithm.

Index Terms— bivariate statistical modeling, 3D natural scene statistics (NSS)

1. INTRODUCTION

Natural scene statistics (NSS) have been proven to be important ingredients towards understanding both the evolution of the human vision system and the design of image processing algorithms [1]. Extensive work has been conducted towards understanding the luminance statistics of natural scenes [2, 3, 4, 5], and the link between natural scene statistics and neural processing of visual stimuli [6, 7]. The natural scene statistics and models of 2D images have been applied to various image and video applications with success, e.g., image denoising [8, 9] and image/video quality assessment [10, 11, 12, 13].

There has also been work conducted on exploring 3D natural scene statistics [14, 15, 16]. Potetz et al. [14] examined the relationships between luminance and range over multiple scales and applied their results to shape-from-shading problems. Yang et al. [15] explored the statistical relationships between luminance and disparity in the wavelet domain, and applied the derived models to a Bayesian stereo algorithm. Recently, Su et al. [16] proposed robust and reliable statistical models for both marginal and conditional distributions of luminance/chrominance and disparity in natural images, and incorporated these models in a chromatic Bayesian stereo algorithm.

However, no work has been proposed to explore the bivariate statistics and modeling of luminance/chrominance and range data in natural scenes. In this paper, we aim to fill this gap by utilizing the high-resolution, high-quality color images and corresponding ground-truth range maps from the LIVE Color+3D Database Release-2 [17]. We study the joint statistics of spatially adjacent wavelet coefficients at different sub-bands of both luminance/chrominance and range data. To model the underlying bivariate statistics, we adopt the versatile and flexible multivariate generalized Gaussian distribution (MGGD). In addition, we observe strong orientation correlation between spatially adjacent sub-band



Fig. 1. An example pair of 2D color image and corresponding ground-truth range map in the LIVE Color+3D Database Release-2.

responses, which can be well modeled by an exponentiated sine function.

The rest of this paper is organized as follows. Section 2 briefly describes the LIVE Color+3D Database Release-2 and the preprocessing steps performed on both luminance/chrominance and range data. The bivariate statistical modeling is detailed in Section 3, including a multivariate statistical hypothesis test. The application to a practical depth estimation framework is demonstrated in Section 4. Finally, Section 5 summarizes the paper.

2. DATABASE AND DATA PRE-PROCESSING

2.1. LIVE Color+3D Database Release-2

To perform the bivariate statistical modeling of natural image and range data, we use the LIVE Color+3D Database Release-2, which contains 99 pairs of stereoscopic left and right color images with accurately co-registered corresponding ground-truth range maps at a high-definition resolution of 1920×1080 [17]. The image and range data in LIVE Color+3D Database Release-2 were collected using an advanced range scanner, RIEGL VZ-400 [18], with a Nikon D700 digital camera mounted on top of it, which were operated on a solidly built robotic gentry to precisely capture stereoscopic images and range maps at an interocular distance of 65 (mm). Figure 1 shows an example pair of the color image and its ground-truth map in the LIVE Color+3D Database Release-2.

2.2. Luminance/Chrominance and Range

All color images are first transformed into the perceptually relevant CIELAB color space with one luminance (L^*) and two chrominance $(a^* \text{ and } b^*)$ components. CIELAB color space is optimized to quantify perceptual color differences and better corresponds to human color perception than does the perceptually nonuniform RGB space [19]. Each image is then transformed by the steerable pyramid decomposition, which is an over-complete wavelet transform that allows for increased orientation selectivity [20]. The utilization of

the wavelet transform is motivated by the fact that its space-scaleorientation decomposition mirrors the band-pass filtering that occurs in area V1 of the primary visual cortex [2, 21].

After applying the multi-scale, multi-orientation decomposition, we perform the perceptually significant process of divisive normalization on the image wavelet coefficients at all sub-bands [22]. The divisive normalization transform (DNT) used in this paper is implemented as follows [23].

$$u_i = \frac{x_i}{\sqrt{\alpha + \mathbf{x}_{\mathbf{g} \setminus i}^T \mathbf{x}_{\mathbf{g} \setminus i}}} = \frac{x_i}{\sqrt{\alpha + \sum_{j \setminus i} g_j x_j^2}}$$
(1)

where *i* is the current pixel location, $\{j\}$ includes the surrounding pixels, x_i represents the wavelet coefficient, u_i represents the coefficient after DNT, α is the semi-saturation constant, and $\{g_j\}$ is a Gaussian weighting function.

For range data, we perform the same multi-scale, multi-orientation wavelet decomposition and divisive normalization transform to obtain the range sub-band coefficients after DNT. In the next section, we detail the statistical modeling of joint bivariate distributions between spatially adjacent sub-band coefficients in both luminance/chrominance and range data in the LIVE Color+3D Database Release-2.

3. BIVARIATE STATISTICAL MODELING

It has been demonstrated that there are strong statistical relationships between co-located luminance/chrominance and range/disparity band-pass responses, which can be well modeled using various univariate probability distributions [15, 16]. However, there still exist higher-order dependencies between neighboring pixels in natural images and range maps, which no robust or reliable statistical model has been proposed to account for. Here we examine the statistical relationships between spatially adjacent pixels in both color images and range maps, and model the corresponding bivariate statistics using a multivariate generalized Gaussian distribution and the correlations with an exponentiated sine function.

3.1. Multivariate Generalized Gaussian Distribution

It is well known that the histogram of sub-band coefficients in natural images becomes Gaussian-like after divisive normalization. However, we found that after performing DNT, the joint statistics of spatially adjacent sub-band coefficients in natural images possess high orientation dependencies, which no longer can be modeled as independent bivariate Gaussian distributions. In order to model both univariate and bivariate statistics of sub-band coefficients in natural images, we utilized the multivariate generalized Gaussian distribution, which includes both multivariate Gaussian and Laplace distributions as special cases.

The probability density function of a multivariate generalized Gaussian distribution (MGGD) is defined as:

$$p(\mathbf{x}|\mathbf{M},\alpha,\beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} g_{\alpha,\beta}(\mathbf{x}^T \mathbf{M}^{-1} \mathbf{x})$$
(2)

where $\mathbf{x} \in \mathbb{R}^N$, **M** is an $N \times N$ symmetric scatter matrix, α and β are the scale and shape parameters, respectively, and $g_{\alpha,\beta}(\cdot)$ is a density generator defined as:

$$g_{\alpha,\beta}(y) = \frac{\beta\Gamma(\frac{N}{2})}{(2^{\frac{1}{\beta}}\pi\alpha)^{\frac{N}{2}}\Gamma(\frac{N}{2\beta})}e^{-\frac{1}{2}(\frac{y}{\alpha})^{\beta}}$$
(3)

where $y \in \mathbb{R}^+$. Note that when $\beta = 0.5$, Eq. (2) leads to the multivariate Laplacian distribution, and when $\beta = 1$, Eq. (2) corresponds to the multivariate Gaussian distribution. Moreover, when $\beta \to \infty$, the MGGD converges to a multivariate uniform distribution.

To fit MGGD to the bivariate histogram of spatially adjacent sub-band coefficients in natural images and find the corresponding model parameters, we adopt the maximum likelihood estimator (MLE) algorithm [24].

3.2. Joint Distribution and Bivariate GGD Fit

We examine the joint distribution of spatially adjacent wavelet coefficients after DNT at different sub-bands for both the color images and ground-truth range maps in the LIVE Color+3D Database Release-2. Here we utilize the steerable pyramid decomposition with five scales, from 1 (finest) to 5 (coarsest), and eight orientations, i.e., 0-deg, 22.5-deg, ..., 157.5-deg [20]. Note that the orientation is defined as the propagation direction of the sinusoidal signal.

In our analysis and modeling, we mainly focus on two ordinary cases of spatially adjacency, the horizontally and vertically adjacent pixels. Specifically, for horizontally adjacent pixels, we form an Nby-2 matrix \mathbf{W} , where N is the total number of pairs of pixels sampled from an image and its range map, and each pair of pixels is sampled from the locations (x, y) and (x + 1, y). Similarly, for vertically adjacent pixels, each pair of pixels is sampled from the locations (x, y) and (x, y + 1). Since similar statistics are observed for both horizontally and vertically adjacent pixels, we discuss the results only for the horizontal case hereafter, if not specified explicitly. In addition, the joint histogram of horizontally adjacent wavelet coefficients after DNT is computed from all the N row vectors $\in \mathbb{R}^2$ in the N-by-2 matrix \mathbf{W} . Therefore, we fit the joint histograms in both the color images and range maps with the bivariate generalized Gaussian distribution (BGGD) model by estimating the parameters **M**, α , and β in Eq. (2) with N = 2.

Figure 2 shows the joint distributions and their corresponding BGGD fits of horizontally adjacent L* sub-band coefficients at scale = 2 and two different orientations: 0-deg and 90-deg. From the threedimensional illustrations, where the blue bars represent the actual coefficient histograms and the colored meshes represent the BGGD fits, we can see that the joint distributions are well fitted by the bivariate generalized Gaussian distributions. The 2D illustrations, which represent the iso-probability contour maps of the joint distributions, also demonstrate the accurate fits of the BGGD models. The most important observation here is that the shape and height of the joint distributions both vary with the sub-band orientation. In particular, when the spatial relationship of adjacent pixels, e.g., horizontal, matches the sub-band orientation, e.g., 90-deg, the joint distribution becomes peaky and extremely elliptical. On the other hand, when the spatial relationship and the sub-band orientation are orthogonal, e.g., horizontal vs. 0-deg, the joint distribution appears to be almost circular and more Gaussian-like. These results imply that there exists much higher dependencies between spatially adjacent luminance pixels after being decomposed by band-pass filters with the matched orientation. Note that similar results are observed for both chrominance and range data, which are not shown here due to space limitations.

To further examine the orientation dependency of luminance/ chrominance and range sub-band coefficients, we plot the BGGD model parameters, i.e., α and β , as a function of orientation at scale = 2 in Figure 3(a) and (b), respectively. We can clearly see that there is strong orientation dependency for both parameters α and β , where they reach the minimum when the spatial relationship matches



Fig. 2. Joint distribution and BGGD fit of L^* at scale = 2 and two different orientations: 0-deg and 90-deg.



Fig. 3. The two fitting parameters of BGGD and correlation coefficients as a function of orientation at scale = 2.

the sub-band orientation, indicating that the horizontally adjacent sub-band coefficients possess the highest correlation at orientation = 90-deg compared to other sub-band orientations. In fact, we can examine these statistical dependencies from another, probably more intuitive, perspective in terms of correlation coefficients, which are embedded in the scatter matrix M of the BGGD model. It can been seen in Figure 3(c) that the horizontally adjacent sub-band coefficients are most correlated when the orientation of band-pass filtering aligns at 90-deg, substantiating the orientation dependencies observed in Figure 2.

3.3. Exponentiated Sine Function

As shown in Figure 3(c), the correlation coefficients between spatially adjacent sub-band responses possess strong orientation dependencies. In order to model this orientation dependency, we adopt the exponentiated sine function, which is given by:

$$y = A \left[\frac{1 + \sin\left(\frac{2\pi x}{T} + \theta\right)}{2} \right]^{\gamma} + c \tag{4}$$

where A is the amplitude, T is the period, θ is the phase, γ is the exponent, c is the offset, and x and y represent the sub-band orientation and the correlation coefficient between spatially adjacent sub-band responses, respectively.

Since the orientation-dependent correlation coefficient between horizontally adjacent sub-band responses is periodic with T = 180 (deg) and phase $\theta = -\frac{\pi}{2}$ by nature, we fix them to obtain the exponentiated sine model with the three parameters: amplitude A, ex-



Fig. 4. The orientation-dependent correlation coefficient curves and the corresponding exponentiated sine fits.

 Table 1. p-Values of Multivariate t-Test for the Exponentiated Sine

 Model Parameters at Different Scales

Scale	L*	L* a* b		Range
1	0.9996	0.9988	0.9988	0.5929
2	0.9855	0.9977	0.9956	0.5011
3	0.9425	0.9864	0.9911	0.4877
4	0.8790	0.9601	0.9394	0.8701
5	0.7780	0.7374	0.8595	0.8155

ponent γ , and offset *c*. Figure 4 shows the orientation-dependent correlation coefficient curves and their corresponding exponentiated sine fits for both luminance/chrominance and range at scale = 3. It can be seen that all of the four exponentiated sine models give extremely good fits according to the mean squared error (MSE), and each of the three fitting parameters has similar values for luminance/chrominance but slightly different values for range data.

3.4. Validation of the Exponentiated Sine Model

To further validate the robustness of the proposed exponentiated sine model, we perform a statistical hypothesis test on the three parameters of the exponentiated sine model across all natural scenes in the LIVE Color+3D Database Release-2. First, we compute the mean of the orientation-dependent correlation coefficient curves across all N images in the database, and fit the exponentiated sine function to the mean curve to obtain the three model parameters, $[A_0, \gamma_0, c_0]^T$. Then, we adopt the one-sample multivariate *t*-test to determine if the null hypothesis H_0 , i.e., the mean vector $\overline{\mathbf{x}} = \sum_{i=1}^{N} \mathbf{x}_i$ of the population $\mathbf{x}_i = [A_i, \gamma_i, c_i]^T, i \in \{1, \dots, N\}$ is equal to our model parameter vector $\boldsymbol{\mu}_0 = [A_0, \gamma_0, c_0]^T$, is supported. In particular, we compute the Hotelling's *T*-squared statistic T^2 , which is a generalization of Student's *t* statistic:

$$T^{2} = N(\overline{\mathbf{x}} - \boldsymbol{\mu}_{0})^{T} \mathbf{S}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}_{0})$$
(5)

where **S** is the sample covariance matrix of \mathbf{x}_i :

$$\mathbf{S} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})^T$$
(6)

Mathad	Pearson's ρ		Spearman's ρ	
Method	Mean	Median	Mean	Median
Previous Framework	0.39	0.38	0.42	0.41
Current Framework	0.43	0.42	0.52	0.53
Depth Transfer	0.41	0.42	0.51	0.52

Table 2. Numerical comparison of Depth Estimation Algorithms in Terms of Correlation Coefficients

Finally, since $\frac{N-P}{P(N-1)}T^2 \sim F_{P,N-P}$, where $F_{P,N-P}$ represents the *F*-distribution with parameters *P* and *N* – *P*, we are able to compute the *p*-value of our null hypothesis test, i.e., $p = 1 - C_{F_{P,N-P}}(\frac{N-P}{P(N-1)}T^2)$, with the cumulative distribution function of the *F*-distribution, $C_{F_{P,N-P}}$. Note that *P* and *N* are the dimension of \mathbf{x}_i and the total number of samples, respectively, and in our test, P = 3 and N = 99.

Table 1 lists the *p*-values of the multivariate *t*-tests for the exponentiated sine model parameters within luminance/chrominance and range data across different scales. We can see that even with a high value of the one-sided significance level $\alpha = 0.1$, the null hypotheses of both luminance/chrominance and range data across all scales are accepted. These results clearly support the validity of our exponentiated sine model fitting the orientation-dependent correlation coefficients between spatially adjacent sub-band responses in natural images. In the next section, we demonstrate a practical application of the proposed exponentiated sine model exploring the bivariate natural scene statistics to the problem of depth estimation from monocular natural images.

4. APPLICATION TO DEPTH ESTIMATION FROM MONOCULAR NATURAL IMAGES

To demonstrate the effectiveness of the proposed exponentiated sine model, we apply it to solve a practical problem of depth estimation from monocular natural images. We improve the performance of the Bayesian framework proposed in [25] by extracting new features as the three parameters, $[A, \gamma, c]^T$, of the exponentiated sine model across different scales from the correlation coefficients of horizontally adjacent sub-band responses. In particular, we append this new feature to the original one in training the regressor to learn the mean and standard deviation of range from color image patches. Since we only focus on the performance improvement due to space limitations, interested readers may refer to [25] for more details.

Table 2 shows the numerical comparison of different depth estimation algorithms, including the previous Bayesian framework, the current Bayesian framework with the exponentiated sine model, and the state-of-the-art method, Depth Transfer [26], in terms of the correlation coefficients between the estimated and ground-truth range values. We train and test these algorithms on LIVE Color+3D Database Release-2, and the results are obtained across 30 trainingtesting splits with 80% training and 20% testing. It can be seen that the new feature extracted based on the proposed exponential sine model boosts the performance of the previous framework in terms of both metrics. In addition, the current framework achieves comparable yet slightly better performance than Depth Transfer. Figure 5 gives the visual comparison of the estimated range maps from the current framework with the exponentiated sine model and Depth Transfer. We can see that besides the higher correlation with the ground-truth range values, the current framework is also able to give more details in the estimated range map.



(c) Scatter Plot of Current Framework (d) Estimated Range Map of Current (5 Scales) Framework (5 scales)



(e) Scatter Plot of Depth Transfer (f) Estimated Range Map of Depth Transfer

Fig. 5. Visual comparison between current framework with 5 scales and depth transfer.

5. CONCLUSIONS

By utilizing the large volume of high-quality, high-resolution color images and ground-truth range maps in LIVE Color+3D Database Release-2, we examined the bivariate statistical relationships and correlation coefficients between spatially adjacent sub-band responses of luminance/chrominance and range data in natural scenes. We utilized a multivariate generalized Gaussian distribution to model the bivariate histogram of spatially adjacent sub-band responses, and found that there exists strong orientation dependency along different spatial alignment of neighboring pixels, which can be well described by the model parameters. We modeled the same orientation dependency observed in the correlation coefficient between spatial adjacent sub-band responses with an exponentiated sine function, and validated its robustness using a multivariate statistical hypothesis test. We believe these bivariate statistics and models embed rich information relating luminance/chrominance and range information of natural environments. Here we briefly demonstrate their effectiveness and power by introducing the exponentiated sine model to a Bayesian framework of depth estimation from monocular images with improved performance. A variety of 3D algorithms and applications, e.g., stereoscopic quality assessment, 2D-to-3D video conversion, etc., can certainly benefit from these robust and effective bivariate statistical models.

6. REFERENCES

- Zhou Wang and Alan C. 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.
- [2] David 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.
- [3] Daniel L. Ruderman and William Bialek, "Statistics of natural images: Scaling in the woods," *Physical Review Letters*, vol. 73, pp. 814–817, Aug. 1994.
- [4] David J. Field, "Wavelets, vision and the statistics of natural scenes," *Philosophical Transactions of the Royal Society* of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 357, no. 1760, pp. 2527, 1999.
- [5] Anuj Srivastava, Ann B. Lee, Eero P. Simoncelli, and Song-Chun Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, 2003.
- [6] Eero P. Simoncelli and Bruno A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, no. 1, pp. 1193–1216, Mar. 2001.
- [7] Wilson S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annual Review of Psychology*, vol. 59, no. 1, pp. 167–192, Jan. 2008.
- [8] Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [9] Dongwook Cho and Tien D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Processing: Image Communication*, vol. 20, no. 1, pp. 77–89, Jan. 2005.
- [10] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [11] Kalpana Seshadrinathan and Alan C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 335–350, Feb. 2010.
- [12] R. Soundararajan and A.C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, Feb. 2012.
- [13] Anush K. Moorthy and Alan C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
- [14] Brian Potetz and Tai Sing Lee, "Scaling laws in natural scenes and the inference of 3D shape," Advances in Neural Information Processing Systems, vol. 18, pp. 1089–1096, 2006.
- [15] Yang Liu, Lawrence K. Cormack, and Alan C. Bovik, "Statistical modeling of 3-D natural scenes with application to bayesian stereopsis," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2515–2530, Sep. 2011.

- [16] Che-Chun Su, Lawrence K. Cormack, and Alan C. Bovik, "Color and depth priors in natural images," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2259 – 2274, June 2013.
- [17] Che-Chun Su, Lawrence K. Cormack, and Alan C. Bovik, "LIVE Color+3D Database - Release 2," http://live.ece.utexas.edu/research/ 3dnss/live_color_plus_3d.html.
- [18] RIEGL, "RIEGL VZ-400 3D Terrestrial Laser Scanner," http://rieglusa.com/products/terrestrial/ vz-400/index.shtml.
- [19] Umesh Rajashekar, Zhou Wang, and Eero P. Simoncelli, "Perceptual quality assessment of color images using adaptive signal representation," *SPIE Int. Conf. on Human Vision and Electronic Imaging*, vol. 7527, no. 1, Jan. 2010.
- [20] Eero P. Simoncelli and William T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," *IEEE International Conference on Image Processing*, vol. 3, pp. 444–447, Oct. 1995.
- [21] Bruno A. Olshausen and David J. Field, "How close are we to understanding V1?," *Neural Computation*, vol. 17, no. 8, pp. 1665–1699, Aug. 2005.
- [22] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, "Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons," *Probabilistic Models of the Brain: Perception and Neural Function*, pp. 203–222, Feb. 2002.
- [23] Siwei Lyu, "Dependency reduction with divisive normalization: justification and effectiveness," *Neural Computation*, vol. 23, pp. 2942–2973, 2011.
- [24] L. Bombrun, F. Pascal, J.-Y. Tourneret, and Y. Berthoumieu, "Performance of the maximum likelihood estimators for the parameters of multivariate generalized gaussian distributions," *IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 3525–3528, 2012.
- [25] Che-Chun Su, Lawrence K. Cormack, and Alan C. Bovik, "Depth estimation from monocular color images using natural scene statistics models," *IEEE IVMSP Workshop*, pp. 1–4, June 2013.
- [26] Kevin Karsch, Ce Liu, and SingBing Kang, "Depth extraction from video using non-parametric sampling," *European Conference on Computer Vision*, vol. 7576, pp. 775–788, Oct. 2012.