ROBUST ANISOTROPIC DISPARITY ESTIMATION
WITH PERCEPTUAL MAXIMUM VARIATION MODELING

Jangheon Kim and Thomas Sikora

Department of Communication System, Technical University of Berlin, Germany
E-mail: {j.kim, sikora}@nue.tu-berlin.de

ABSTRACT

We present a robust anisotropic dense disparity estimation algorithm which employs perceptual maximum variation modeling. Edge-preserving dense disparity vectors are estimated using a coarse-to-fine diffusive method on iteratively filtered images, i.e. the scale-space. While an energy-minimization framework adjusts local disparity, the edges are efficiently preserved by anisotropic disparity-field diffusion. However, the localization at weak image edges which have small brightness variations is fundamentally difficult. In this paper, perceptual maximum variation modeling prevents the delocalization flow over edges, e.g. over-diffusion and back-diffusion, computed by evaluating small variations. We perform disparity-field diffusion on a perceptually optimized color space, which combines the small differences in both brightness and chromaticity. Additionally a consistency constraint is employed in the modeling to avoid the influence of global color distributions and to enhance important edges as the human vision system does. The experimental results show the excellent localization performance preserving the disparity discontinuity of each object.

***Index Terms*—** stereo vision, image matching, diffusion processes, image color analysis

1. INTRODUCTION

Dense disparity estimation is important for many 3D applications of image-based modeling and rendering e.g. depth scanning, photogrammetry, light-fields, arbitrary viewpoint synthesis, etc. Given two images taken simultaneously with a pair of cameras, the goal of disparity estimation is to locate for each point in one image its corresponding point in the other image. The main difficulty of disparity estimation is the ambiguity of local image structure due to image noise, unbalanced brightness, similar texture and occlusion. If pels in an image look alike, it is difficult to find corresponding pels from another viewpoint image.

To obtain more reliable estimation performance, recent algorithms employ local appearance matching with boundary constraints between features, edges and disparity discontinuity etc. The constraints are utilized as a landmark of the coherency of objects [1, 2]. However, their performances are not satisfactory for producing a dense field with full resolution because of the ill-posed localization problem. The latest researches incorporate regularization that attempts to filter off delocalized errors. Thus, an energy-based formulation is used to iteratively minimize an energy function employing a regularization term. In isotropic regularization [3], convolution carries out filtering based on variance. However, the scale of linear transformations in the convolution leads to undesired smoothing of the important discontinuity. In contrast, anisotropic diffusion methods [4, 5] are applied to prevent important structure from blurring. The basic idea is to modify the scale of diffusion at the discontinuity edges with steep intensity gradients. This method can be applied for disparity estimation to yield smoothed but detailed results in some images. However ill-posed local minima solving the PDE (partial differential equation) are the serious drawback.

In our previous work [9], Gaussian scale-space disparity estimation with anisotropic disparity-field diffusion efficiently solved the problems. The multiple-resolutions of the scale-space provide the best trade-off between the detected features and the localization performance. Global disparity is estimated with the constraint of strong and meaningful boundaries in coarse resolution and then it is iteratively refined into finer resolution using PDE. While the disparity-field is diffusively smoothed with isotropic propagation following the streamline of Gaussian scale-space, anisotropic diffusion uses different weighting to suppress the length of the propagation for only the orthogonal direction to the edges. This method gave us good dense disparity maps preserving importantly strong variation structures. However, the regularization often results in over-diffusion when evaluating small brightness variations in images. If important edges e.g. object boundaries are formed by small brightness variations, delocalized flow causes over-diffusion smoothing the discontinuity as shown in Figure 1.

In this paper, we solve this small variation problem by perceptual maximum variation modeling. Perceptually-motivated color spaces can be utilized to evaluate mutual coherency and geometrical continuation. Perceptual maximum variation modeling is employed to combine the small differences in both brightness and chromaticity, using a least squares optimization applying PCA (principal component analysis). A consistency constraint for the modeling is considered to avoid influence from global color distributions and to enhance homogeneous color regions. We apply this new approach for anisotropic disparity estimation.

In the next section, we describe the perceptual maximum variation modeling in detail. Its application to anisotropic dense disparity estimation is shown in section 3. Section 4 represents the simulation results.

2. PERCEPTUAL MAXIMUM VARIATION MODELING

2.1. Perceptual maxima estimation in color space

Our approach is based on the fact that the human visual system detects the boundary of objects by considering chromaticity and brightness simultaneously. If the brightness difference of neighboring object is very small, our eyes can detect the important
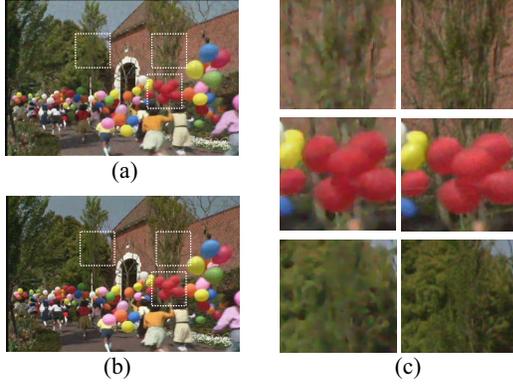
Figure 1. (a) anisotropic diffusion by evaluating brightness variation (b) anisotropic diffusion with perceptual maximum variation modeling (c) details with over-diffusion on small variations (left: 1a, right: 1b).
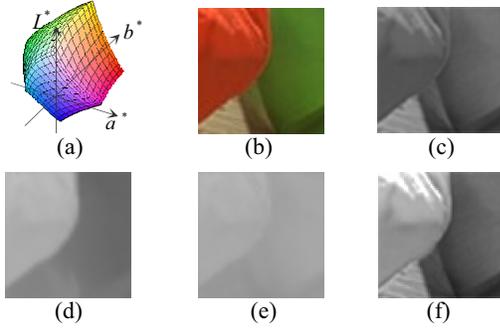
Figure 2. (a) CIE-$L^*a^*b^*$ color space, (b) original color image (c) L^* (d) a^* and (e) b^* (f) $\mathcal{P}$-space from our modeling (here, the size of $\mathcal{P}$-space is 255 for visualization purpose).

edges using the chromaticity difference. We estimate a higher dimensional perceptual $\mathcal{P}$-space (note that the size of this space is not always the same with the size of the common color space) by projecting the maximum variations of both brightness and chromaticity. First, a color image is converted into the perceptually uniform color space of $CIE\text{-}L^*a^*b^*$, as depicted in Figure 2a. Perceptual difference between two neighboring color pels in $CIE\text{-}L^*a^*b^*$ can be simply measured by the Euclidean distance between the two vectorial values $c_p(\mathrm{x})=[L^*_p,a^*_p,b^*_p]$ and $c_q(\mathrm{x})=[L^*_q,a^*_q,b^*_q]$ with the perceptual color metric δc^*_{pq} [6] as

$$\delta c^*_{pq} = \sqrt{\left(L^*_p - L^*_q\right)^2 + \left(a^*_p - a^*_q\right)^2 + \left(b^*_p - b^*_q\right)^2} \qquad (1)$$

where x denotes a pel x=(x,y) on the $CIE\text{-}L^*a^*b^*$ color domain $I(\mathrm{x}):\mathbb{R}^3_+=\{c(L^*_p,a^*_p,b^*_p)\geq 0\}$. Color pels with the same δc^*_{pq} are perceptually equal. The brightness $\omega(\mathrm{x})$ is obtained from the luminance domain L^* and the chromaticity $\varphi(\mathrm{x})$ from a^* and b^* respectively. $\omega(\mathrm{x})$ represents the length of the color vectors and $\varphi(\mathrm{x})$ the normalized color components.

$$\omega(\mathrm{x}) = \sum_{i=1}^{n}\left\|I_n(\mathrm{x})\right\| \ and \ \varphi(\mathrm{x}) = I(\mathrm{x})/\left\|I(\mathrm{x})\right\| = I(\mathrm{x})/\omega(\mathrm{x}) \quad (2)$$

We estimate the visually maximum variations using the perceptual difference combining $\omega(\mathrm{x})$ and $\varphi(\mathrm{x})$ in the color metric δc^*_{pq}. A linear transform which projects original luminance values L^*_p and L^*_q into a perceptual intensity $\mathcal{P}_p$ and $\mathcal{P}_q$ (which remains

proportional to the perceptual color difference) is obtained by minimizing the following quadratic function.

$$\mathcal{P}\left(c_{0,\cdots,p,q,\cdots,n}(\mathrm{x})\right) = \min\sum_{p=0}^{n-2}\sum_{q=p+1}^{n-1}\left(w(\mathcal{P}_p,\mathcal{P}_q)\big/\delta c^*_{pq} - K\right)^2 \qquad (3)$$

$w(\mathcal{P}_p,\mathcal{P}_q)=|\mathcal{P}_p-\mathcal{P}_q|$ is a positive, symmetric dissimilarity function. The constant K is chosen for the proportional weighting e.g. when $K=1$, L^*_p and L^*_q are equally modulated with respect to the perceptual differences. $\mathcal{P}_p$ and $\mathcal{P}_q$ generally have higher dynamic range than L^*_p and L^*_q, since they include total difference of both brightness and chromaticity. The $\mathcal{P}$-space has maximum values in human visual range and $\nabla\mathcal{P}$ is the perceptual maximum variation. Applying PCA to (3), the entire distribution of color values can be projected onto the primary L^*-axis of the ellipsoid using a linear transform. The principal components are the eigenvectors of its covariance matrix. By choosing the projection base with the largest eigenvalue, the perceptual maximum variation can be estimated as shown in Figure 2f.

2.2. Consistency constraint

Although PCA performs very well in aligning the colors which have locally-compact or globally-smooth spatial distributions, the global color distribution of natural image may not be suitable to fit local properties. (3) is convex, but may converge into multiple minima. We solve the problem by considering a consistency constraint based on chromaticity. The chromaticity difference is calculated in the equation (1) between two color pels as

$$\delta\varphi_{pq} = \sqrt{(a^*_p - a^*_q)^2 + (b^*_p - b^*_q)^2} \qquad (4)$$

As shown in Figure 2d and 2e, the chromaticity domain can be used as a good consistency measure because it inherently has piecewise smoothness over the image - while the brightness in L^* is affected by saturation, lightness e.g. shadow, reflection and noise. The consistency measure is decreased when the spatial distance to the pel under consideration increases due to the coherency of objects. In this paper, consistency Λ is defined as a similarity group based on probability, using the chromaticity difference $\delta\varphi_{pq}$ in (4) and spatial distance χ_{pq} with proportional constant k. γ_ξ and γ_d determine the scale of the spatial proximity of the pels.

$$\Lambda\left(c_{0,\cdots,p,q,\cdots,n}(\mathrm{x})\right) = \sum_{i=1}^{n}k\cdot exp\left(-(\delta\varphi_{pq}/\gamma_\xi + \chi_{pq}/\gamma_d)\right) \qquad (5)$$

During the optimization in (1), the consistency constraint enhances the localization of homogeneity between color regions by removing the diversity of global color distribution from the image. Accordingly, anisotropic diffusion evaluating the maximum variations allows a robust convergence performance as Figure 3 illustrates.

3. ANISOTROPIC DISPARITY ESTIMATION WITH PERCEPTUAL MAXIMUM VARIATIONS

3.1. Local appearance matching using scale-space

As described in the introduction, a good measure for homogeneous regions is needed to restrict matching within coherent objects. A supporting region for matching can be established by analyzing a smooth varying orientation structure in Gaussian scale-space as

$$\eta = \theta_+ = \nabla\mathcal{P}_\sigma / \|\nabla\mathcal{P}_\sigma\| \quad \text{and} \quad \xi = \theta_- = \nabla\mathcal{P}_\sigma^\perp / \|\nabla\mathcal{P}_\sigma\| \qquad (6)$$

where $\mathcal{P}_\sigma = (2\pi\sigma^2)^{-1} e^{-x2/2\sigma2} * I(x)$ is pre-filtered $\mathcal{P}$-space map by Gaussian kernel and the unit vectors η and ξ are defined by gradient direction of the image Gaussian and its orthogonal direction (i.e. isophote direction), respectively. The two varying orientation θ_+ and θ_- correspond to the vector edges of gradients and the isophotes. The scale parameter σ of the Gaussian-filter kernel is used to control the boundary strength. The coarse-to-fine structure of the scale-space provides the best trade-off between detection and localization performance. In *Helmholtz* theory, any vector field $\vec{F}$ can be represented as a sum of a conservative and solenoidal vector field with vector potential $\vec{A}$.

$$\vec{F} = \vec{F}_{con} + \vec{F}_{sol} = -\nabla V + \nabla \times \vec{A} \qquad (7)$$

Taking the divergence of both side as

$$\vec{\nabla} \cdot \vec{F} = -\Delta V + \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) \qquad (8)$$

where Δ is the Laplacian. Since the second term $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A})$ in (4) is zero, the boundary function V [7] can be solved by the Poisson equation. Let the $\mathcal{P}$-space be a continuous function that is only divided by perceptual edges into $n+1$ regions $\{R_0, ..., R_i, R_j, ..., R_n\}$. The supporting region $\mathcal{W}_\sigma$ is calculated by combining similar pels enclosed by boundary.

$$\mathcal{W}_\sigma(V) = \int_{R_i} w_\sigma\big(\mathcal{P}_p(x), \mathcal{P}_q(x)\big)dx + \int_{R_i}\int_{R_j} w_\sigma\big(\mathcal{P}_p(x), \mathcal{P}_q(x)\big)dx \quad (9)$$

where $w_\sigma(\mathcal{P}_p, \mathcal{P}_q)$ is a dissimilarity function with a scale between the neighboring pels, $\{\mathcal{P}_p[(\theta_+, \theta_-),\sigma], \mathcal{P}_q[(\theta_+, \theta_-),\sigma]\} \in R_N$ on the smoothly varying oriented structure of (6). The first and the second terms respectively purpose the grouping of similar pels in a region e.g. R_i, and between two regions e.g. R_i and R_j. Dense disparity vectors d are locally estimated in the supporting region $\mathcal{W}_\sigma$ by matching the perceptual maximum intensity. They are refined in a coarse-to-fine scheme. The energy function is

$$E_\Omega(d) = \int_\Omega \rho[\mathcal{P}_l(x,y) - \mathcal{P}_r(x+d_{l\to r}(x) \in \mathcal{W}_\sigma(V), y)]^2 dx + \lambda \int_\Omega e_\sigma(d)dx \qquad (10)$$

Ω is the image domain, the subscripts denote the matching direction, e.g. $l\to r$ for left-to-right direction and e_σ is a regularization term with Lagrange multiplier λ. This method results in a more accurate local disparity estimation solution due to the restriction of matching errors within $\mathcal{W}_\sigma$. ρ is the determinant of a potential function that derives boundary flow to restrict the outlier.

### 3.2. Anisotropic outlier removal in disparity-field with perceptual maximum variation

We regularize the locally estimated disparity vector field by globally removing the outlier in (10) using an edge-preserving anisotropic diffusion as shown in Figure 4:

$$e_\sigma = G(\|\nabla\mathcal{P}_{\sigma,l}\|)\nabla d_{l\to r} \qquad (11)$$

$\nabla\mathcal{P}_\sigma$ is the perceptual maximum variation in a scale in (6) and $G(\nabla)$ is an anisotropic diffusion function which is called "*edge-stopping function*" [9]– used to modify the diffusion coefficient at
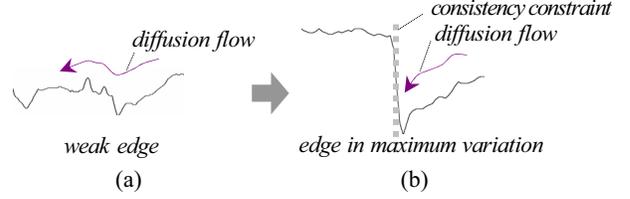
Figure 3. (a) delocalized flow on a weak edge which causes over-diffusion and back-diffusion artifacts, (b) the proposed method.
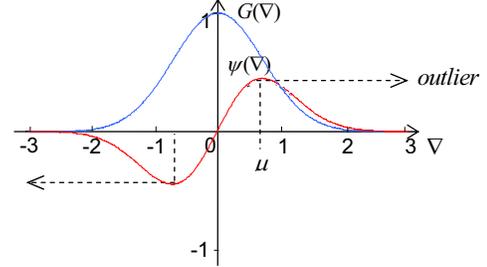
Figure 4. anisotropic diffusion function G and flux function ψ.

edges and to derive discontinuities in the flux function $\Psi(\nabla) = G(\nabla)\nabla'$ as

$$G(\nabla) = e^{-(\nabla^2/\mu^2)} \qquad (12)$$

where a positive constant μ controls the level of contrast of edges affecting the diffusion process as Figure 4 represents. We solve the energy-minimization problem of (10) by discretizing the following numerical equation using finite differences.

$$\begin{aligned}\frac{d_{l\to r}^{t+1} - d_{l\to r}^t}{\tau_\sigma} &= \lambda div\left[\left(\frac{G\big(\|\nabla\mathcal{P}_{\sigma,l}\|\big)}{\|\nabla\mathcal{P}_{\sigma,l}\|}\right)\nabla d_{l\to r}^t\right] + \\ \frac{\partial\mathcal{P}_{\sigma,r}(x+d_{l\to r}^t, y)}{\partial x}&\Big[\big(\mathcal{P}_{\sigma,l}(x,y) - \mathcal{P}_{\sigma,r}(x+d_{l\to r}, y) \in \mathcal{W}_\sigma(V)\cdot\vec{N}\big) \\ &+ \frac{\partial\mathcal{P}_{\sigma,r}(x+d_{l\to r}^t(x), y)}{\partial x}(d_{l\to r}^{t+1} - d_{l\to r}^t)\Big]\end{aligned} \qquad (13)$$

Inhomogeneous time diffusion process with discrete sampling solves the problem. Different pels diffuse at a different time-scale related to the pel confidence. By increasing the time step τ_σ while refining resolution in scale-space, higher confidence pels diffuse much slower than low confidence pels. In the coarse scales, disparity estimation with diffusion is performed for a wider range between strong edges following the smooth field. Occlusion can be handled in the range. With further iterations, this gradually occurs in a narrower range between weaker edges as well. To remove outliers of matching, a local matching region with the ranges is fitted using normal vector $\vec{N}$ of the edge following streamlines of scale-space. The divergence term div performs global diffusion of the local disparity map to remove outlier in each scale.

4. SIMULATION RESULTS

In Figure 5, the improvement of our previous method [9] is exemplified for the "Balloon" (shown in Figure 1, 720×480) and "Wagon" stereo natural images (720×576). Edge-preserving disparity estimation can be robustly achieved in image regions containing important structure with strong and weak brightness
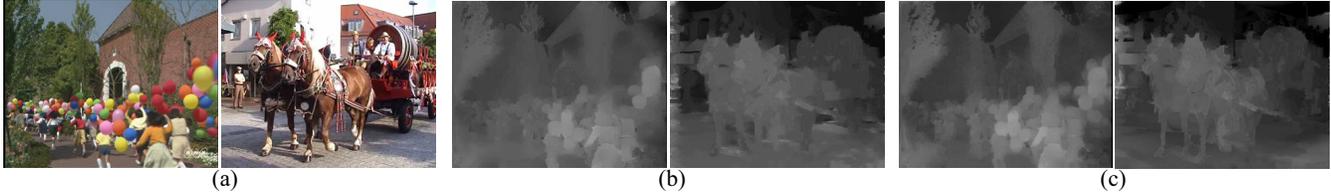
Figure 5. (a) base images of "Balloon" and "Wagon" (b) dense disparity maps using our previous work [9] (c) improvements in this paper.

| (a) | (b) | (c) | (d) | (e) | (f) |
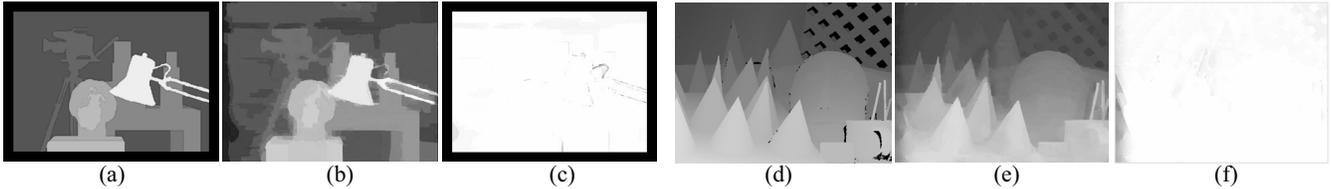
Figure 6. (a) ground-truth of "Tsukuba" (b) result of "Tsukuba" (c) error pels (i.e. darker pels have larger error) of 6b that have absolute difference from ground-truth (6a) > 1.0 (d) ground-truth of "Cones" (e) result of "Cones" (d) error pels between 6d and 6e (diff. >1.0).

Methods	"Tsukuba"			"Cones"		
	noocc.	*all*	*disc.*	*noocc.*	*all*	*disc.*
SSD + min-filter	5.23	7.07	24.1	10.6	19.8	26.3
layered stereo [10]	1.30	1.57	6.92	6.59	14.7	14.4
graph-cut [11]	1.94	4.12	9.39	7.70	18.2	15.3
proposed method	1.06	1.21	4.47	3.41	8.55	5.02

| *noocc.* | *all* | *disc.* | *noocc.* | *all* | *disc.* |
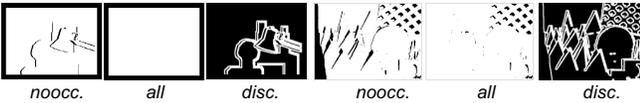
Table 1. performance comparison using percent of error pels [12] in the white regions of alpha maps; *noocc.*: non-occluded regions; *disc.*: areas of depth discontinuities.

variations because the maximum variation modeling improves the convergences into the global minima of (10). Figure 6 shows the absolute disparity error compared to the ground-truth of images "Tsukuba" and "Cones" of [12] using the percentage of error pels. We ignore 18 pels of the border for the "Tsukuba" image when computing the statistics due to the absence of data in the ground-truth. Figure 6b and 6e show the robust performance of the proposed method for discontinuities and homogeneous regions. Table 1 depicts the performance comparison between well-known reliable optimization methods (e.g. layered stereo [10] and graph-cut [11]) requiring heavy computation costs. The errors are calculated using alpha maps which decide evaluating regions. The proposed method achieves high accuracy performance in texture and discontinuity over these techniques.

5. SUMMARY AND CONCLUSION

In this paper, we proposed a novel reliable anisotropic disparity estimation algorithm employing perceptual maximum variation modeling. Falling into local minima and over-splitting in texture gradient areas can efficiently be avoided by using the properties of the scale-space. Over-diffusion and back-diffusion are prevented applying perceptual maximum variations. Thus, anisotropic disparity estimation can be accurately performed in geometrical continuation coherency of a scene. Simulation results using several stereo natural images prove the excellent performance.

7. REFERENCES

[1] M. Agrawal and L. Davis, "Window based, discontinuity preserving stereo," in Proc. of *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 66-73, 2003.

[2] S. Birchfield and C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo," *Int. Jour. of Comp. Vision*, 35, pp. 269-293, 1999.

[3] Y. Yang, A. Yuille, and J. Lu, "Local, global, and multilevel stereo matching," in Proc. of *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 274–279, 1993.

[4] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert, "Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach," *Journal of Visual Communication and Image Representation*, vol. 13, pp. 96-114, 2002.

[5] C. Strecha and L. Van Gool, "PDES-based multi-view depth estimation," in Proc. of *Int. Symp. on 3D Data Processing Visualization and Transmission*, pp. 416-425, 2002.

[6] K. Bauml, X. Zhang, and B. Wandell, "Color Spaces and Metrics," in *Vision Models and Applications to Image and Video Processing*, Kluwer Academic Publishers, 2001.

[7] B. Sumengen, B. Manjunath and C. Kenney, "Image Segmentation using Curve Evolution and Flow Fields," in Proc. of *IEEE Int. Conf. on Image Processing*, pp. 105-108, 2002.

[8] M. Black, G. Sapiro, D. Marimont and D. Heeger, "Robust Anisotropic Diffusion," *IEEE Transactions on Image Processing*, vol. 7, pp. 421-432, 1998.

[9] J. Kim and T. Sikora, "Gaussian Scale-Space Dense Disparity Estimation with Anisotropic Disparity-Field Diffusion," in Proc. of *IEEE Int. Conf. on 3-D Digital Imaging and Modeling*, 2005.

[10] M. Bleyer and M. Gelautz, "A layered stereo algorithm using image segmentation and global visibility constraints," in Proc. of *Int. Conf. on Image Processing*, pp. 2997-3000, 2004.

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222-1239, 2001.

[12] *http://www.middlebury.edu/stereo*.