DISPARITY AND NORMAL ESTIMATION THROUGH ALTERNATING MAXIMIZATION

Ramya Narasimha [1], *Elise Arnaud* [1,2], *Florence Forbes* [1], *Radu Horaud* [1]

[1] INRIA Rhône-Alpes [2] Université Joseph Fourier, Laboratoire Jean Kuntzmann

ABSTRACT

In this paper, we propose an algorithm that recovers binocular disparities in accordance with the surface properties of the scene under consideration. To do so, we estimate the disparity as well as the normals in the disparity space, by setting the two tasks in a unified framework. A novel joint probabilistic model is defined through two random fields to favor both intra field (within neighboring disparities and neighboring normals) and inter field (between disparities and normals) consistency. Geometric contextual information is introduced in the models for both normals and disparities, which is optimized using an appropriate alternating maximization procedure. We illustrate the performance of our approach on synthetic and real data.

***Index Terms*—** Stereo Vision, MRF, CRF, Alternating Maximization

1. INTRODUCTION

Most of the recent algorithms in stereo disparity estimation make an inherent *fronto-parallel assumption* in their modeling, thus biasing the results towards piecewise- constant “staircase solutions”. In an attempt to move beyond this assumption, Devernay et. al. [1] proposed to extend the classical correlation method to compute both the disparity and its derivatives, which relate to the differential properties of the surface. In [2], the authors estimate the scene structure as a set of smooth surface patches, while performing segmentation and correspondence iteratively. However, they do not consider the geometrical properties of the surface itself. Most of the recent methods cast the problem as an energy minimization in Markovian framework and use approximate inference algorithms like Belief Propagation, Mean Field or Graph-cuts to find the local minima. A good taxonomy of these approaches can be found in [3]. In such a setting, slanted surfaces are usually recovered by post-processing the disparities using plane-fitting on segmented regions [4]. These approaches implicitly assume fronto-parallel planes in the definition of their objective function and cannot handle curved surfaces. In order to overcome this limitation, [5, 6] present a framework incorporating higher-order priors to encode the surface properties. While [5] uses a new quadratic pseudo-boolean optimization, [6] suggests a non-parametric approach casting the pixels and disparity together as networks using sparse graphs which are matched then using graph cuts.

Our work is inspired by Li et. al. ’s work [7], which explicitly takes into account the differential geometric contextual information in a Markov Random Field (MRF) based disparity estimation framework. Li et. al. measure the consistency of the normals by transporting them along the surface and guide the disparity estimation towards a geometrically consistent map. In order to overcome the numerical instability issues encountered by [1], Li et. al. perform all the derivative computations in the depth space. As well as requiring the knowledge of the internal camera parameters, this algorithm precomputes the local surface normals.

In our work, we propose to carry out cooperatively both disparity and normal estimations using two Random Fields (RFs) that are linked to encode consistency between disparities and surface properties. This idea of using multiple RFs to estimate one or more variables has been previously used in contexts such as: estimation of motion discontinuities and optical flow in [8], estimation of disparity by integrating it with line process and occlusion in [9], estimation of disparity and image boundaries in [10]. In our case, the disparity is modeled as a MRF including geometric contextual information in the pair-wise regularizing term, thus favoring a disparity solution consistent with the scene surfaces – possibly slanted and/or curved. The normal field, modeled as Conditional Random Field (CRF), is built under the assumption that the scene in question is made of piecewise smooth surfaces and disparity is used as observed data. The proposed joint model results in a posterior distribution, for both the disparity and normal fields, which is used for their estimation according to a Maximum A Posteriori (MAP) principle. While the Mean Field algorithm ([11]) is used to estimate the disparities, the normals are estimated using Iterated Conditional Modes (ICM).

2. JOINT DISPARITY AND NORMAL MODEL

We consider a finite set $\mathcal{S}$ of $p \times q$ pixels on a regular 2D-grid. The observed data are made of left and right images, $\mathbf{I}_L$ and $\mathbf{I}_R$, which are together referred to as $\mathbf{I}$. In our setting, the left image is taken as the reference image. We denote by $\mathbf{D} = \{D_\mathbf{x}, \mathbf{x} \in \mathcal{S}\}$ the unknown disparity values at each pixel $\mathbf{x} = (u, v)$. The $D_\mathbf{x}$’s are considered as random variables that take their values in a finite discrete set of L disparity labels $\mathcal{L}$. $\mathbf{D}$ is referred to as the disparity field or disparity map and takes its values in $\mathcal{D} = \mathcal{L}^{p \times q}$. Similarly, we consider a surface normal field $\mathbf{N} = \{\mathbf{N}_\mathbf{x}, \mathbf{x} \in \mathcal{S}\}$. We use small letters $\mathbf{d}$ and $\mathbf{n}$ to denote specific realizations of the random fields $\mathbf{D}$ and $\mathbf{N}$. Ideally, we are interested in finding the MAP estimates of $\mathbf{D}$ and $\mathbf{N}$, $(\mathbf{d}^{MAP}, \mathbf{n}^{MAP}) = \arg\max_{\mathbf{d},\mathbf{n}} p(\mathbf{d}, \mathbf{n}|\mathbf{I})$. However, this global optimization problem has in general no straightforward solution. Thus, we consider instead an iterative approach consisting in maximizing the posterior probability alternately in the first and second variable. At a given iteration t, this alternation between the two variables can be done as follows:

$$\mathbf{d}^{(t+1)} = \arg\max_{\mathbf{d} \in \mathcal{D}} p(\mathbf{d}|\mathbf{n}^{(t)}, \mathbf{I}) \tag{1}$$

$$\mathbf{n}^{(t+1)} = \arg\max_{\mathbf{n} \in \mathcal{N}} p(\mathbf{n}|\mathbf{d}^{(t+1)}, \mathbf{I}) \, . \tag{2}$$

It is, therefore, sufficient to define these two conditionals $p(\mathbf{d}|\mathbf{n}, \mathbf{I})$ and $p(\mathbf{n}|\mathbf{d}, \mathbf{I})$ to account for cooperation mechanisms between $\mathbf{D}$ and $\mathbf{N}$.

2.1. Disparity Model given the Normals

We first specify the disparity distribution conditionally to the normal field and the observed data as $p(\mathbf{d}|\mathbf{n}, \mathbf{I})$. We model this distribution as a MRF on $\mathcal{D}$ with an energy function consisting of two terms, a data dependent term and an interaction term, as follows:

$$p(\mathbf{d}|\mathbf{n}, \mathbf{I}) \propto \Phi_d(\mathbf{d}, \mathbf{I}) \, \Psi(\mathbf{d}, \mathbf{n}) \, . \tag{3}$$

Our **data term** $\Phi_d(\mathbf{d}, \mathbf{I})$ is similar to the one described in Yang et al [4]. A cost is assigned at location $\mathbf{x}$ based on weighted window matching metric that takes into account both the color and the proximity of the pixels within the window. We formulate this cost as a robust function $\Phi_d(\mathbf{d}, \mathbf{I}) = \exp\big(-\sum_{\mathbf{x} \in \mathcal{S}} \lambda \min\left(\phi(\mathbf{I}_L, \mathbf{I}_R, \mathbf{d}), 2T\right)\big)$, depending on two parameters λ and T, where,

$$\phi(\mathbf{I}_L, \mathbf{I}_R, \mathbf{d}) = \frac{\displaystyle\sum_{\mathbf{y} \in W_\mathbf{x}, \bar{\mathbf{y}} \in W_{\bar{\mathbf{x}}}} w(\mathbf{x}, \mathbf{y}) w(\bar{\mathbf{x}}, \bar{\mathbf{y}}) e\big(\mathbf{I}_L(\mathbf{y}), \mathbf{I}_R(\bar{\mathbf{y}})\big)}{\displaystyle\sum_{\mathbf{y} \in W_\mathbf{x}, \bar{\mathbf{y}} \in W_{\bar{\mathbf{x}}}} w(\mathbf{x}, \mathbf{y}) w(\bar{\mathbf{x}}, \bar{\mathbf{y}})} . \tag{4}$$

In order to understand the above equation, consider a candidate correspondence $\bar{\mathbf{x}}$ in the right image for the point $\mathbf{x}$ in the left, i.e $\bar{\mathbf{x}} = \mathbf{x} - (0, d_\mathbf{x})$. To compute the cost of $d_\mathbf{x}$, the pixel-wise cost $e\big(\mathbf{I}_L(\mathbf{y}), \mathbf{I}_R(\bar{\mathbf{y}})\big) = |\mathbf{I}_L(\mathbf{y}) - \mathbf{I}_R(\bar{\mathbf{y}})|$, within two windows $W_\mathbf{x}$ and $W_{\bar{\mathbf{x}}}$ centered at $\mathbf{x}$ and $\bar{\mathbf{x}}$ are weighted and summed. Each pixel within the window $\mathbf{y} \in W_\mathbf{x}$ is weighted according its color difference $\nabla c_{\mathbf{xy}}$ and its spatial proximity $\nabla g_{\mathbf{xy}}$ to $\mathbf{x}$, as follows: $w(\mathbf{x}, \mathbf{y}) = \exp\left(-\nabla c_{\mathbf{xy}}/\gamma_c - \nabla g_{\mathbf{xy}}/\gamma_g\right)$. A similar weight is computed for $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in W_{\bar{\mathbf{x}}}$. The window size parameter[1] W in our case is set to 5×5 and the parameters γ_c and γ_g are set to 10 and 21 respectively. The parameters of the robust function are set to $\lambda = 1$ and T is fixed to the average pixel cost computed over all pixels and disparity labels.

Our **interaction term** is a symmetric modified version of the one presented in Li et al [7]. Although the interpretation is similar, we propose to include geometric information via surface normals considered as a separate random field. Expressing compatibility between the disparity and normal fields enables us to encode geometric constraints without computing disparity derivatives directly from the disparity field. Furthermore, we use only first order differential information obtained from the normal field, which avoids the numerical instabilities. This term has then a standard pair-wise interaction form $\Psi(\mathbf{d}, \mathbf{n}) = \prod_{(\mathbf{x},\mathbf{y})} \psi(d_\mathbf{x}, d_\mathbf{y}, \mathbf{n})$ where $(\mathbf{x}, \mathbf{y})$ denotes neighboring pixels on the image grid. It is to be noted that we consider 8-neighborhood system. The term $\psi(d_\mathbf{x}, d_\mathbf{y}, \mathbf{n})$ incorporates a general surface model (the terms in exponential of (5)) to ensure that the neighboring disparities lie on the same planar surface. This term, thus, encodes geometric constraints via consistency with the surface normal field $\mathbf{n}$ and is expressed as follows:

$$\begin{aligned} \psi(d_\mathbf{x}, d_\mathbf{y}, \mathbf{n}) &= \\ \exp\Bigg(&- \frac{|d_\mathbf{y} - d_\mathbf{x} - \frac{\partial d_\mathbf{x}}{\partial u}(u_\mathbf{y} - u_\mathbf{x}) - \frac{\partial d_\mathbf{x}}{\partial v}(v_\mathbf{y} - v_\mathbf{x})|}{\sigma_D} \\ &- \frac{|d_\mathbf{x} - d_\mathbf{y} - \frac{\partial d_\mathbf{y}}{\partial u}(u_\mathbf{x} - u_\mathbf{y}) - \frac{\partial d_\mathbf{y}}{\partial v}(v_\mathbf{x} - v_\mathbf{y})|}{\sigma_D} \Bigg), \end{aligned} \tag{5}$$

where $\mathbf{x} = (u_\mathbf{x}, v_\mathbf{x})$, $\mathbf{y} = (u_\mathbf{y}, v_\mathbf{y})$ and σ_D are scalar parameters for robustness. σ_D is set to 3.0 in our experiments. With $\mathbf{n}_\mathbf{x} = (n_u, n_v, n_d)$, the disparity partial derivatives are computed from the normals as $\frac{\partial d_\mathbf{x}}{\partial u} = -\frac{n_u}{n_d}$ and $\frac{\partial d_\mathbf{x}}{\partial v} = -\frac{n_v}{n_d}$.

The optimization for disparity is done using Mean field approximation. At each iteration of the Mean Field procedure, we compute an interpolated disparity map using plane fitting. This enables us to follow [7] by considering so-called *floating disparity* labels. Starting from a discrete set of L integer disparity labels $\mathcal{L} = \{d_1, \ldots, d_L\}$, we allow them to move to another set of L possibly non-integer labels. The idea is to capture finer geometric features by adapting the initial disparity discretized grid to the image scene. Importantly, this can be done while keeping the discrete pair-wise MRF formulation. Considering at iteration t a current continuous value d at pixel $\mathbf{x}$, we find l such that $d \in [d_l; d_{l+1})$ and then change the disparity label d_l to d. This provides an efficient alternative to the quickly intractable increase of L.

2.2. Normal Model given Disparity

We express the disparity conditional normal model as a Conditional Random Field with a Gaussian distribution:

$$p(\mathbf{n}|\mathbf{d}, \mathbf{I}) \propto \prod_{\mathbf{x} \in \mathcal{S}} \prod_{\mathbf{y} \in \mathcal{N}_\mathbf{x}} \exp\left(-\frac{\|\mathbf{n}_\mathbf{x} - \vec{E}_{\mathbf{xy}}(\mathbf{d}, \mathbf{I}, \mathbf{n}_\mathbf{y})\|^2}{2\sigma^2}\right) \tag{6}$$

where $\mathcal{N}_\mathbf{x}$ is the 8-neighborhood set. The above equation represents a pairwise relationship between the normal at $\mathbf{x}$ and its neighbors $\mathbf{y} \in \mathcal{N}_\mathbf{x}$. Instead of just computing an Euclidean distance between the two normals at positions $\mathbf{x}$ and $\mathbf{y}$, we compute the distance between $\mathbf{n}_\mathbf{x}$ and vector $\vec{E}_{\mathbf{xy}}$, which is the influence of the neighboring normal $\mathbf{n}_\mathbf{y}$ taking into account disparity $\mathbf{d}$ and the image $\mathbf{I}$ information. The vector

[1] The parameters are set manually and are the same for all experiments shown in this paper.

$\vec{E}_{\mathbf{xy}}$ is determined using a method inspired by Page et. al. [12]. We express $\vec{E}_{\mathbf{xy}}$ as:

$$\vec{E}_{\mathbf{xy}} = w_{\mathbf{xy}}(\mathbf{d},\mathbf{I})\,\vec{N}_{\mathbf{xy}}, \tag{7}$$

where,

$$\vec{N}_{\mathbf{xy}} = \mathbf{n}_{\mathbf{y}} + 2\cos\big(\theta_{\mathbf{xy}}(\mathcal{S},\mathbf{d})\big)(\vec{xy}/\|\vec{xy}\|) \tag{8}$$

The first term in the above equation is the current normal estimate at site $\mathbf{y}$. The $\theta_{\mathbf{xy}}$ in the second term, is the angle between the normal at $\mathbf{y}$ and the vector $\vec{xy}$, from x to y, where $x = (\mathbf{x}, d_{\mathbf{x}})$ and $(\mathbf{y}, d_{\mathbf{y}})$. If the two points $(\mathbf{x}, d_{\mathbf{x}})$ and $(\mathbf{y}, d_{\mathbf{y}})$ are on a plane consistent with $\mathbf{n}_{\mathbf{y}}$ then this value is zero. The second term, thus, gives the error between $\vec{xy}$ and the plane described by $\mathbf{n}_{\mathbf{y}}$. The weight $w_{\mathbf{xy}}(\mathbf{d},\mathbf{I})$ is described as $w_{\mathbf{xy}}(\mathbf{d},\mathbf{I}) = \exp\Big(-\frac{|d_{\mathbf{x}}-d_{\mathbf{y}}|+|\nabla\mathbf{I}(\mathbf{x})|}{\sigma_N}\Big)$ where $|\nabla\mathbf{I}(\mathbf{x})|$ represents the gradient magnitude at $\mathbf{x}$ of the reference image[2]. The image gradient is used to weight the influence of the neighbors so as to prevent normals from being smoothed across boundaries. As it can be seen, σ_N is the only parameter to be controlled in the estimation of normals which is set to 1.9 in our experiments.

While Page et. al. describe a deterministic voting procedure that uses eigen decomposition to determine the normals, we maximize $p(\mathbf{n}|\mathbf{d},\mathbf{I})$ using an approximate the MAP estimate of $\mathbf{n}$. This is done by using an ICM procedure, in which $\mathbf{n}$ is set iteratively $\forall\mathbf{x} \in \mathcal{S}$ as follows,

$$\mathbf{n}_{\mathbf{x}}^{MAP} \approx \frac{1}{\mathrm{Card}(\mathcal{N}_{\mathbf{x}})} \sum_{\mathbf{y}\in\mathcal{N}_{\mathbf{x}}} \vec{E}_{\mathbf{xy}}(\mathbf{d},\mathbf{I},\mathbf{n}_{\mathbf{y}}) \tag{9}$$

where $\mathrm{Card}(\mathcal{N}_{\mathbf{x}})$ is the cardinality of the set $\mathcal{N}_{\mathbf{x}}$. We use ICM algorithm for the maximization here because it allows for the optimization of the normal energy in the continuous domain. It can be seen from equations (9) and (7) that the normals are not determined by a simple mean, but their estimation incorporates disparity information and the piecewise smooth assumption.

3. ALTERNATING MAXIMIZATION

The resulting alternation procedure is the following: at iteration $t = 0$, all the normal field values are assumed to be $\{0, 0, 1\}$ and the first step (1) is performed to get an initial estimate of the disparity map. Then denoting by $\mathbf{n}^{(t)}$ and $\mathbf{d}^{(t)}$ current estimates of the normal and disparity fields, the two steps below are carried out alternately,

1) Update normal field $\mathbf{n}^{(t)}$ into $\mathbf{n}^{(t+1)}$ by applying ICM on (6).
2) Update disparity field $\mathbf{d}^{(t)}$ into $\mathbf{d}^{(t+1)}$ by:
   (i) computing the first order disparity derivatives using $\mathbf{n}^{(t+1)}$ and
   (ii) updating disparity estimates into $\mathbf{d}^{(t+1)}$ with Mean Field applied to the conditional disparity model (3).

This alternation is carried out for a prescribed number of iterations (in our case we obtained good results with 5 iterations) at 4 different scales, ranging from coarse to fine. Each of the Mean Field processes is performed until the total average energy change is less than 0.01 which corresponds to about $4-5$ iterations. The ICM for normals was carried out for 10 iterations.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The examples used to test our models are Corridor (fig. 1), Head (fig. 2) and Cloth (fig. 3) from the Middlebury database. During testing the disparity range and therefore the value of L is fixed to different values depending on each image pair. We use an HSV color-code[3] to represent the normals.The color is obtained by mapping the azimuth and elevation of each normal to hue and saturation, respectively. The first example is the Corridor image (fig. 1(a)) of size 256×256 pixels and the disparity range set to 11 pixels. Figure. 1(c) shows the result by just applying straight forward estimation disparity with fronto-parallel assumption. The estimated disparity and normal maps using our approach are shown respectively in fig. 1(d) and fig. 1(e). These figures indicate that the normals rightly follow the slanted surface of the corridor disparity map. The fig. 1(f) shows the bad pixels map obtained by thresholding the absolute error between the estimated disparity and the groundtruth. The percentage bad-pixel error, which is computed by taking the ratio between the number of bad pixels to overall image pixels, is 3.3%. The Head image (fig. 2(a)) is of size 219×255 pixels and the disparity range of 40 pixels. The fig. 2(b) shows the result obtained when the disparity plane fit is used directly as a post processing. Figures. 2(c) and 2(d) show the results obtained for disparity and normals using our approach. The direct representation of the normals using arrows, corresponding to the color-coded one in fig. 2(d), is shown in fig. 2(e). Our result shows how the proposed procedure captures the surface deformations through the normals which is in-turn used to obtain a consistent disparity map. Finally, we show our results for the Cloth image (fig. 3(a)), of size 370×417 and disparity range 60 pixels, in figures. 3(c) and 3(d). The bad pixels map is shown in fig. 3(e) with percentage bad-pixel error of 5.4%.

In conclusion, we proposed a new joint probabilistic model with the following advantages: 1) it does not require the direct computation of high-order disparity derivatives as in [1], 2) it embeds the estimation of surface properties in the model rather than refining the results using post-processing as done by [4] , 3) the consideration of two conditional models allows for more dependence or independence according to the information to be incorporated 4) unlike [7] where normals are precomputed, the alternating procedure in our approach results in mutual improvement of both disparities and normals. As for the probabilistic setting itself, we first focused on defining a valid unified framework to model cooperations

[2] For color images $\mathbf{I}(\mathbf{x})$ represents the average of the RGB channels.

[3] Note that the same color code is used for all the other normal-maps

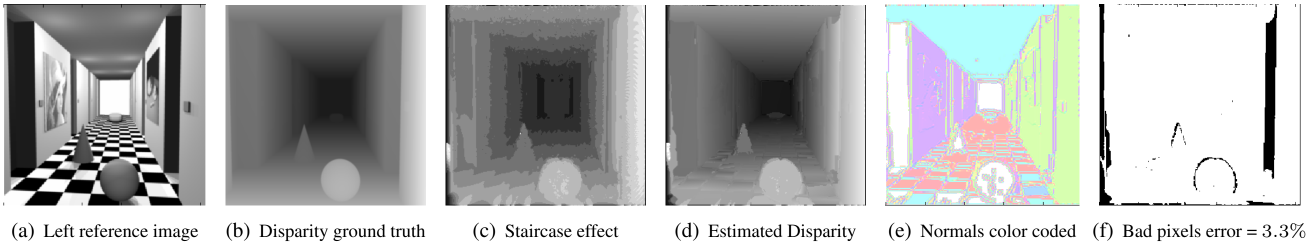
(a) Left reference image (b) Disparity ground truth (c) Staircase effect (d) Estimated Disparity (e) Normals color coded (f) Bad pixels error = 3.3%

Fig. 1. Results using our approach shown in 1(d) and 1(e). 1(f) shows the bad pixels map for error > 1.0.

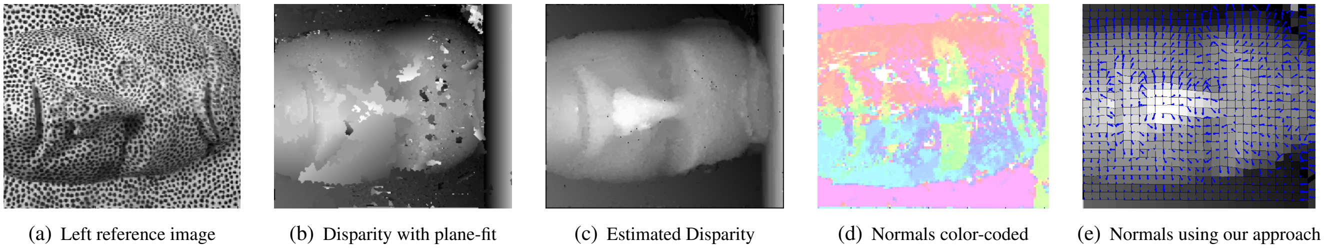
(a) Left reference image (b) Disparity with plane-fit (c) Estimated Disparity (d) Normals color-coded (e) Normals using our approach

Fig. 2. Face Image: Our results are shown in 2(c), 2(d) and 2(e)

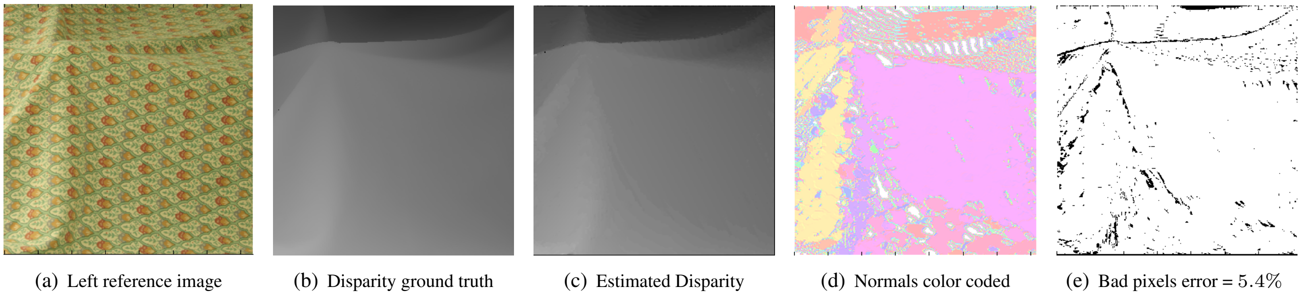
(a) Left reference image (b) Disparity ground truth (c) Estimated Disparity (d) Normals color coded (e) Bad pixels error = 5.4%

Fig. 3. Cloth Image: Our results are shown in 3(c) and 3(d). 3(e) shows the bad pixels map for error > 1.0

and used MAP principle for inference. A natural future direction of research is to investigate the possibility of richer modeling alternatives in which rather than estimating the realizations of fields $\mathbf{D}$ and $\mathbf{N}$, we would be able to estimate their full distributions like Expectation Maximization.

5. REFERENCES

[1] F. Devernay and O.D. Faugeras, “Computing differential properties of 3-D shapes from stereoscopic images without 3-D models,” *CVPR*, vol. 94, 1994.

[2] M.H. Lin and C. Tomasi, “Surfaces with occlusions from layered stereo,” *IEEE PAMI*, vol. 26, no. 8, 2004.

[3] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *IJCV*, vol. 47, no. 1-3, 2002.

[4] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, “Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling,” *IEEE PAMI*, vol. 31, no. 3, 2009.

[5] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon, “Global stereo reconstruction under second-order smoothness priors,” *IEEE PAMI*, vol. 31, no. 12, 2009.

[6] B.M. Smith, L. Zhang, and H.L. Jin, “Stereo matching with nonparametric smoothness priors in feature space,” in *CVPR*, 2009.

[7] G. Li and S.W. Zucker, “Surface geometric constraints for stereo in belief propagation,” in *CVPR*, 2006.

[8] F. Heitz and P. Bouthemy, “Multimodal estimation of discontinuous optical flow using markov random fields,” *IEEE PAMI*, vol. 15, no. 12, 1993.

[9] J. Sun, N.N. Zheng, and H.Y. Shum, “Stereo matching using belief propagation,” *IEEE PAMI*, vol. 25, no. 7, 2003.

[10] R. Narasimha, E. Arnaud, F. Forbes, and R. P. Horaud, “Cooperative disparity and object boundary estimation,” in *ICIP*, 2008.

[11] C. Strecha, R. Fransens, and L.J. Van Gool, “Combined depth and outlier estimation in multi-view stereo,” in *CVPR*, 2006.

[12] D. L. Page, A. Koschan, Y. Sun, J. K. Paik, and M. A. Abidi, “Robust crease detection and curvature estimation of piecewise smooth surfaces from triangle mesh approximations using normal voting,” in *CVPR*, 2001.