Cooperative disparity and object boundary estimation

Ramya Narasimha, Élise Arnaud, Florence Forbes, Radu Horaud

COOPERATIVE DISPARITY AND OBJECT BOUNDARY ESTIMATION

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

[1] INRIA Rhône-Alpes [2] Université Joseph Fourier
655, avenue de l'Europe 38330 Montbonnot, France

ABSTRACT

In this paper we carry out cooperatively both disparity and object boundary estimation by setting the two tasks in a unified Markovian framework. We introduce a new joint probabilistic model that allows to estimate disparities through a Markov random field model. Boundary estimation then cooperates with disparity estimation to gradually and jointly improve accuracy. The feedback from boundary estimation to disparity estimation is made through the use of an auxiliary field referred to as a displacement field. This field suggests the corrections that need to be applied at disparity discontinuities in order that they align with object boundaries. The joint model reduces to a Markov random field model when considering disparities while it reduces to a Markov chain when focusing on the displacement field. The performance of our approach is illustrated on real stereo images sets, demonstrating the power of this cooperative framework.

Index Terms— Stereo Disparity Estimation, Markov Random Field, Markov Chain

1. INTRODUCTION

Most disparity estimation algorithms result in disparity discontinuities occurring at improper locations. By improper we mean locations which are not at the actual depth discontinuities. Recent approaches such as [1], [2] overcome this problem by treating it seperately using colour image segmentation. In this paper we propose a method that simultaneously estimates disparity and corrects the disparity at discontinuities in mathematically sound framework. Our approach relies on two assumptions: (i) that the discontinuities in depth are usually at object boundaries (ii) that the disparity discontinuities obtained from naive disparity estimation are usually at the vicinity of actual depth discontinuities. Thus, if we locate the object boundaries which are in the vicinity of the disparity discontinuities — using the gradient map[1] of the image as evidence —we can correct the disparity values so that they fit closer to the object boundaries. We thus suggest a new joint model to estimate both disparity and object boundaries cooperatively *i.e.*, the presence of disparity discontinuities aids the detection of object boundaries and vice versa. This is done by introducing a *displacement field*, which suggests the corrections that need to be applied at disparity discontinuities in order to align them with object boundaries. The disparities are modeled as a Markov Random Field (MRF) while the displacement field reduces to a Markov chain during estimation. The resulting procedure involves alternating between estimation of disparity and displacement fields in an iterative framework.

The idea of combining multiple fields in order to estimate one or more variables has been previously used in several contexts, including motion analysis [6], background/foreground separation [7], depth estimation [8], [4] and boundary detection [9]. More recent work by Chung et al. [10] suggests a technique for estimating boundaries in conjunction with segmentation using a dual-MRF framework. As in [10], we *do not* use a line process for boundary representation (as in case of [6], [9]). This means that the field representing the boundary does not exist between the pixels (as in line processes) but on the pixel locations themselves. In contrast to [10], we do not focus on the boundary locations themselves but provide the directions toward which discontinuities must be displaced based on observed image gradient values. Hoiem et al [11] also address the problem of recovering depth boundaries combining traditional edge and region cues with depth and surface information, but using only a single image. Recently the same problem was addressed by Mattoccia et al [12] for stereo image pairs using Scanline Optimization. Our approach has the further advantage of making a clear distinction between the probabilistic model and the subsequent optimization.

2. JOINT DISPARITY-DISPLACEMENT MODEL

We consider a finite set $\mathcal{S}$ of N pixels on a regular two-dimensional grid. The observed data are made of a left and a right images $\mathbf{I}_L$ and $\mathbf{I}_R$, where $\mathbf{I}$ represents the pair. We denote by $\mathbf{D} = \{D_{\mathbf{x}}, \mathbf{x} \in \mathcal{S}\}$ the unknown disparity values at each pixel. The $D_{\mathbf{x}}$'s are considered as random variables that take their values in a finite discrete set $\mathcal{L}$. $\mathbf{D}$ is referred to as the disparity field or disparity map and takes its values in $\mathcal{D} = \mathcal{L}^N$. In addition, we consider a *displacement* field denoted by $\mathbf{A} = \{A_{\mathbf{x}}, \mathbf{x} \in \mathcal{S}\}$ where each $A_{\mathbf{x}}$ is made of two random variables denoted by $M_{\mathbf{x}}$ and $E_{\mathbf{x}}$. For each pixel location $\mathbf{x}$, $E_{\mathbf{x}}$ can be interpreted as the normal vector

[1] All gradients are computed from the left intensity image.

to the disparity discontinuity contour and $M_{\mathbf{x}} \in \{-1, 0, 1\}$ indicates the direction of displacement along this normal.

Our goal is then to estimate realizations of $\mathbf{D}$ and $\mathbf{A}$ from the observed stereo image pair. Ideally we are interested in finding the MAP (Maximum A Posteriori) estimates of $\mathbf{D}$ and $\mathbf{A}$ using the posterior probability $p(\mathbf{d}, \mathbf{a}|\mathbf{I})$. Note that we use $\mathbf{d}$ and $\mathbf{a} = (\mathbf{m}, \mathbf{e})$ to denote specific realizations of the random fields $\mathbf{D}$ and $\mathbf{A} = (\mathbf{M}, \mathbf{E})$. This global optimization problem has in general no straightforward solution. Thus, we consider instead an iterative approach consisting in maximizing the posterior probability alternatively in the first and second variable. Starting from current estimates $\mathbf{d}^{(q)}$ and $\mathbf{a}^{(q)}$ at iteration q, it can be shown using Bayes' theorem that the alternation procedure can be written as :

(i) $\mathbf{d}^{(q+1)} = \arg\max_{\mathbf{d}} p(\mathbf{d}|\mathbf{a}^{(q)}, \mathbf{I})$

(ii) $\mathbf{a}^{(q+1)} = \arg\max_{\mathbf{a}} p(\mathbf{a}|\mathbf{d}^{(q+1)}, \mathbf{I})$

Hence, we now focus on the definitions of the conditional distributions $p(\mathbf{d}|\mathbf{a}, \mathbf{I})$ and $p(\mathbf{a}|\mathbf{d}, \mathbf{I})$.

2.1. Displacement conditional disparity model

We first specify the disparity distribution conditionally to the displacement field and the observed data. We define $\mathbf{d}$ as a MRF with the following energy function, where β_d is an interaction parameter to be fixed:

$$p(\mathbf{d}|\mathbf{a}, \mathbf{I}) \propto \mathrm{e}^{\left(-\sum_{\mathbf{x}\in\mathcal{S}} U_d(d_{\mathbf{x}}, \mathbf{I}) - \beta_d \sum_{\mathbf{x}\in\mathcal{S}} \sum_{\mathbf{y}\in\mathcal{N}_{\mathbf{x}}} V_d(d_{\mathbf{x}}, d_{\mathbf{y}}, \mathbf{a})\right)}. \quad (1)$$

The data term U_d in (1) assigns a cost for each disparity value chosen at location $\mathbf{x}$ based on the intensity difference between the left and the right images. However to account for outliers, it is formulated as a robust function:

$$U_d(d_{\mathbf{x}}, \mathbf{I}) = \min\left(\lambda_1 \left|\mathbf{I}_L(\mathbf{x}) - \mathbf{I}_R(\mathbf{y})\right|, T_1\right) \quad (2)$$

for every $\mathbf{x} = (i, j), \mathbf{y} = (i, j + d_{\mathbf{x}})$, where λ_1 and T_1 are the robust function's parameters and $\mathbf{I}_L(\mathbf{x})$ (resp. $\mathbf{I}_R(\mathbf{x})$) denotes the pixel value at location $\mathbf{x}$ in the left (resp. right) image.

The interaction term V_d in (1) defines how the disparity at a location is influenced by its neighbors. In usual MRF-based disparity models the neighborhood $\mathcal{N}_{\mathbf{x}}$ is defined as the set of the 4 (or 8) nearest pixels of $\mathbf{x}$, all influencing the value at $\mathbf{x}$. In the displacement conditional model, only some of the neighbors in $\mathcal{N}_{\mathbf{x}}$ actually interact with $\mathbf{x}$, depending on their displacement values defined by the displacement field $\mathbf{A}$. An intermediate *Active neighborhood* field $\mathbf{H}(\mathbf{A}) = \{\mathbf{H}_{\mathbf{x}}(\mathbf{A}), \mathbf{x} \in \mathcal{S}\}$ is then built to encode this specificity. Let $\mathcal{N}^0_{\mathbf{x}}$ denote the neighbors of $\mathbf{x}$ for which the displacement field is 0.Then, define: $\forall \mathbf{a}$,

$$\mathbf{H}_{\mathbf{x}}(\mathbf{a}) = \begin{cases} \mathcal{N}^0_{\mathbf{x}} & \text{if } m_{\mathbf{x}} = 0 \\ \{\lfloor \mathbf{x} - m_{\mathbf{x}}\mathbf{e}_{\mathbf{x}} \rfloor\} \cap \mathcal{N}^0_{\mathbf{x}} & \text{if } |m_{\mathbf{x}}| = 1 \end{cases} \quad (3)$$

where $\mathbf{e}_{\mathbf{x}}$ is the unit vector at location $\mathbf{x}$ when $\mathbf{E} = \mathbf{e}$. Note that $\lfloor \mathbf{x} - m_{\mathbf{x}}\mathbf{e}_{\mathbf{x}} \rfloor$ denotes the closest point in $\mathcal{S}$, when $\mathbf{x} - m_{\mathbf{x}}\mathbf{e}_{\mathbf{x}}$ does not correspond exactly to a site in $\mathcal{S}$. In particular, $\mathbf{H}_{\mathbf{x}}(\mathbf{a})$ can be the empty set if $|m_{\mathbf{x}}| = 1$ and $|m_{\lfloor \mathbf{x} - m_{\mathbf{x}}\mathbf{e}_{\mathbf{x}} \rfloor}| = 1$. Figure1 illustrates the three cases that can possibly occur in this active neighborhood construction. **Case I**: suggests that, when there is no evidence that a displacement is required at the location under consideration $\mathbf{x}$, *ie* $m_{\mathbf{x}} = 0$, then the only *reliable* neighbors are those which themselves do not require displacement. For this case, the interaction potential is then defined as a robust function, that enforces the smoothness of the disparity field. **Case II**: indicates that when $|m_{\mathbf{x}}| = 1$ the disparity at this location has to be changed in order to better agree with object boundaries. In such a case, only a neighbor whose location is $\mathbf{x} - m_{\mathbf{x}}\mathbf{e}_{\mathbf{x}}$ corresponding to the one which is located at the opposite direction to the normal to the discontinuity located at $\mathbf{x}$ should interact. **Case III**: illustrates the case when $|m_{\mathbf{x}}| = 1$ and the selected active neighbor has itself been assigned a non zero displacement value, meaning that its corresponding disparity value is likely to change and should then not be used. In that case, the set $\mathbf{H}_{\mathbf{x}}(\mathbf{a})$ is empty and the interaction potential is set to 0.

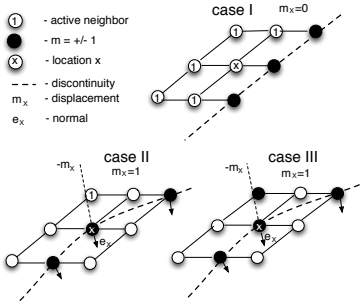

Fig. 1. The figure shows how the neighborhood is activated based on $\mathbf{A} = (\mathbf{M}, \mathbf{E})$

Denoting by $\mathbb{1}_{\{D_{\mathbf{x}}=D_{\mathbf{y}}\}}(d_{\mathbf{x}}, d_{\mathbf{y}})$ the function of $d_{\mathbf{x}}$ and $d_{\mathbf{y}}$ that is 1 when $d_{\mathbf{x}} = d_{\mathbf{y}}$ and 0 otherwise[2], V_d is defined as below, where λ_2 and T_2 are the robust function's parameters

$$V_d(d_{\mathbf{x}}, d_{\mathbf{y}}, \mathbf{a}) = \quad (4)$$
$$\begin{cases} \min(\lambda_2 |d_{\mathbf{x}} - d_{\mathbf{y}}|, T_2) \text{ if } \mathbf{x} \in \mathbf{H}_{\mathbf{y}}(\mathbf{a}) \ \& \ \mathbf{y} \in \mathbf{H}_{\mathbf{x}}(\mathbf{a}) \\ 0 \qquad\qquad\qquad\qquad \text{if } \mathbf{x} \notin \mathbf{H}_{\mathbf{y}}(\mathbf{a}) \ \& \ \mathbf{y} \notin \mathbf{H}_{\mathbf{x}}(\mathbf{a}) \\ T_2\big(1 - \mathbb{1}_{\{D_{\mathbf{x}}=D_{\mathbf{y}}\}}(d_{\mathbf{x}}, d_{\mathbf{y}})\big) \text{ otherwise.} \end{cases}$$

2.2. Disparity conditional displacement model

Preliminary steps are involved in the modeling of the distribution of $\mathbf{A}$ conditionally to $\mathbf{D}$ and $\mathbf{I}$. It requires the definition of *discontinuity chains* representing the locations at which a disparity discontinuity occurs considering a current value $\mathbf{d}$ of the disparity map $\mathbf{D}$. This extraction is done by retaining the $\mathbf{x}$ as a discontinuity, if the difference between the disparity at $\mathbf{x}$ and the disparity at any one of the neighboring locations lies above the threshold value T_2 (first case of (4)). This procedure gives a binary map of disparity discontinuity locations which is then converted

[2]More generally for a variable $\mathbf{z}$ taking values in a set $\mathcal{Z}$, we denote by $\mathbb{1}_{\mathcal{Z}_1}(\mathbf{z})$ the indicator function which is 1 when $\mathbf{z} \in \mathcal{Z}_1 \subset \mathcal{Z}$ and 0 otherwise.

into a set of discontinuity chains denoted by $\mathcal{C}(\mathbf{D})$. The set $\mathcal{C}(\mathbf{D})$ is made of a number $T(\mathbf{D})$ of connected components $\mathcal{C}(\mathbf{D}) = \{\mathcal{C}^1(\mathbf{D}), \ldots \mathcal{C}^t(\mathbf{D}) \ldots, \mathcal{C}^{T(\mathbf{D})}(\mathbf{D})\}$ which can be of different sizes. Each discontinuity chain $\mathcal{C}^t(\mathbf{D})$ is itself made up of two sets; $\mathcal{S}^t(\mathbf{D}) = \{\mathbf{x}_1^t, \ldots, \mathbf{x}_K^t\} \subset \mathcal{S}$ is a set of K locations which are connected, namely for all $k = 2, \ldots K-1$, $\mathbf{x}_k^t$ is a neighbor of $\mathbf{x}_{k-1}^t$ and $\mathbf{x}_{k+1}^t$ while the first and last locations have only one neighbor and the set $\mathcal{W}^t = \{\mathbf{w}_1^t, \ldots, \mathbf{w}_K^t\}$ represents the normals to the chain associated to each location $\mathbf{x}_k^t$. The conditional distribution $p(\mathbf{a}|\mathbf{d}, \mathbf{I})$, defined using the discontinuity chains $\mathcal{C}(\mathbf{D})$, is reduced to $\prod_{t=1}^{T(\mathbf{d})} p(\mathbf{m}^t|\mathcal{C}^t(\mathbf{d}), \mathbf{I})$ due t; (i) at the discontinuity locations the displacement field normals $\mathbf{E}_\mathbf{x}$ are the same as the discontinuity chains normals $\mathbf{w}_\mathbf{x}$ and (ii) that non zero displacements can occur only at discontinuity locations. The probability distribution of each of the chains is expressed as a second order Markov chain as follows:

$$p\big(\mathbf{m}^t|\,\mathcal{C}^t(\mathbf{d}), \mathbf{I}\big) = P\big(m_1^t, m_2^t|\mathcal{C}^t(\mathbf{d}), \mathbf{I}\big) \prod_{k=3}^{K} \underbrace{p\big(m_k^t|m_{k-1}^t m_{k-2}^t, \mathcal{C}^t(\mathbf{d}), \mathbf{I}\big)}_{A} \quad (5)$$

where m_k^t is the displacement at location $\mathbf{x}_k^t$ so that $\mathbf{m}^t = \{m_1^t, \ldots, m_K^t\}$. From now on we drop the superscript t, as each of the chain distributions will be defined in a similar manner. We define A in (5) as follows, where β_c is the interaction parameter:

$$A \propto \exp\big(-U_c(m_k, \mathcal{C}(\mathbf{d}), \nabla\mathbf{I}_L) - \beta_c\, V_c(m_k, m_{k-1}, m_{k-2}, \mathcal{C}(\mathbf{d}))\big) \quad (6)$$

The data term U_c in (6) associates a cost for moving the point $\mathbf{x}_k$ on either side of the normal $\mathbf{w}_k$. In order to determine this cost, we first choose points on both sides of the normal using a range of locations distant from $-\epsilon$ to ϵ. Then, we determine the difference in the gradient magnitude between the current position and the points chosen along the normal. If this difference is negative, it means that moving the current chain point to that position along the normal leads to a higher gradient region. Hence, the direction of motion towards this position is favored. (see Figure 2). This can be written as:

$$U_c\big(m_k, \mathcal{C}(\mathbf{d}), \nabla\mathbf{I}_L\big) = 1 - 2\ \mathbb{1}_{\{M_k = s_k\left(\nabla\mathbf{I}_L, \mathbf{w}_k\right)\}}(m_k) \quad (7)$$

Denoting $\mathbf{y}_l = \mathbf{x}_k + l\,\mathbf{w}_k$, $|\nabla\mathbf{I}_L|$ the gradient magnitude in the reference image $\mathbf{I}_L$ and sgn as sign function we have

$$s_k\left(\nabla\mathbf{I}_L, \mathbf{w}_k\right) = \mathrm{sgn}\Big(\arg\min_{l\in[-\epsilon,\epsilon]}\big(|\nabla\mathbf{I}_L(\mathbf{x}_k)| - |\nabla\mathbf{I}_L(\mathbf{y}_l)|\big)\Big)$$

The interaction term V_c in (6) assigns a score based on the angle between vectors defined by the positions $(\mathbf{x}_{k-2}, \mathbf{x}_{k-1})$ and $(\mathbf{x}_{k-1}, \mathbf{x}_k)$. The interaction term enforces smoothness by favoring those m-values which make the angle between these vectors as close to zero as possible. This second-order Markov chain can be easily turned into a a first-order Markov chain so that optimal displacement values can be found using the standard Viterbi algorithm.

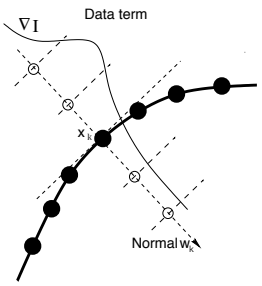

Fig. 2. The figure shows the calculation of U_c using gradient information

3. ALTERNATING MAXIMIZATION PROCEDURE

Using the defined joint probabilistic model, we then estimate the MAP estimates of $\mathbf{D}$ and $\mathbf{A}$ given the observed image set $\mathbf{I}$., with the alternation procedure defined in the previous section. The resulting iterative procedure is described below. At iteration $q = 0$, the displacement field values are assumed to be zero and the calculation of $\mathbf{d}^{(0)}$ is performed from $p(\mathbf{d}|\mathbf{a}, \mathbf{I})$ using a Variational mean field approach [3]. Then the two following steps are carried out alternatively, in a multi-scale framework [5]. The alternation is carried out until a large percentage of the displacement values are equal to zero.

1. Update Displacement field $\mathbf{a}^{(q)}$ into $\mathbf{a}^{(q+1)} = (\mathbf{m}^{(q+1)}, \mathbf{e}^{(q+1)})$:
   (i) Extract disparity discontinuities from current disparity map $\mathbf{d}^{(q)}$
   (ii) Convert the binary map into a set of discontinuity chains $\mathcal{C}(\mathbf{d}^{(q)})$
   (iii) At discontinuity locations, *ie.* for $\mathbf{x} \in \mathcal{S}(\mathbf{d}^{(q)})$, set $\mathbf{e}_\mathbf{x}^{(q+1)}$ to the normal $\mathbf{w}_\mathbf{x}$ in $\mathcal{C}(\mathbf{d}^{(q)})$. For the remaining locations, $m_\mathbf{x}^{(q+1)}$ is set to zero so that $\mathbf{e}_\mathbf{x}^{(q+1)}$ can be set arbitrarily
   (iv) Estimate new m-values for each discontinuity chain $\mathcal{C}^t(\mathbf{d}^{(q)})$ in $\mathcal{C}(\mathbf{d}^{(q)})$ using the Viterbi algorithm for the model defined in (5)
2. Update Disparity field $\mathbf{d}^{(q)}$ into $\mathbf{d}^{(q+1)}$:
   (i) Determine the active neighborhood system using $\mathbf{a}^{(q+1)}$
   (ii) Obtain $\mathbf{d}^{(q+1)}$ using Variational mean field for the MRF (1).

4. RESULTS AND DISCUSSION

We report an experimental evaluation of our algorithm using four stereo image pairs in Figure 3. The row `(a)` shows the left images of the stereo image pairs used in the experiment. Row `(b)` shows the disparity maps obtained using Variational Mean Field without using the proposed approach. Row `(c)` shows the disparity discontinuities obtained from row `(b)`. This corresponds to the initialization of our alternation procedure.The results of our algorithm are illustrated in rows `(d)` and `(e)`, where the boundary and disparity estimation results are respectively presented.

The results clearly indicate the advantage of including boundary estimation where the standard boundary detector usually fails to segregate the objects. Row `(c)` shows that the discontinuities of the disparity maps in `(b)` are displaced from physical object boundaries. However, using our approach, we obtain both object boundaries and improved disparity maps as in `(d-e)`.These results show the value of the joint model, and the effectiveness of the cooperative algorithm, even with basic image information such as gradient

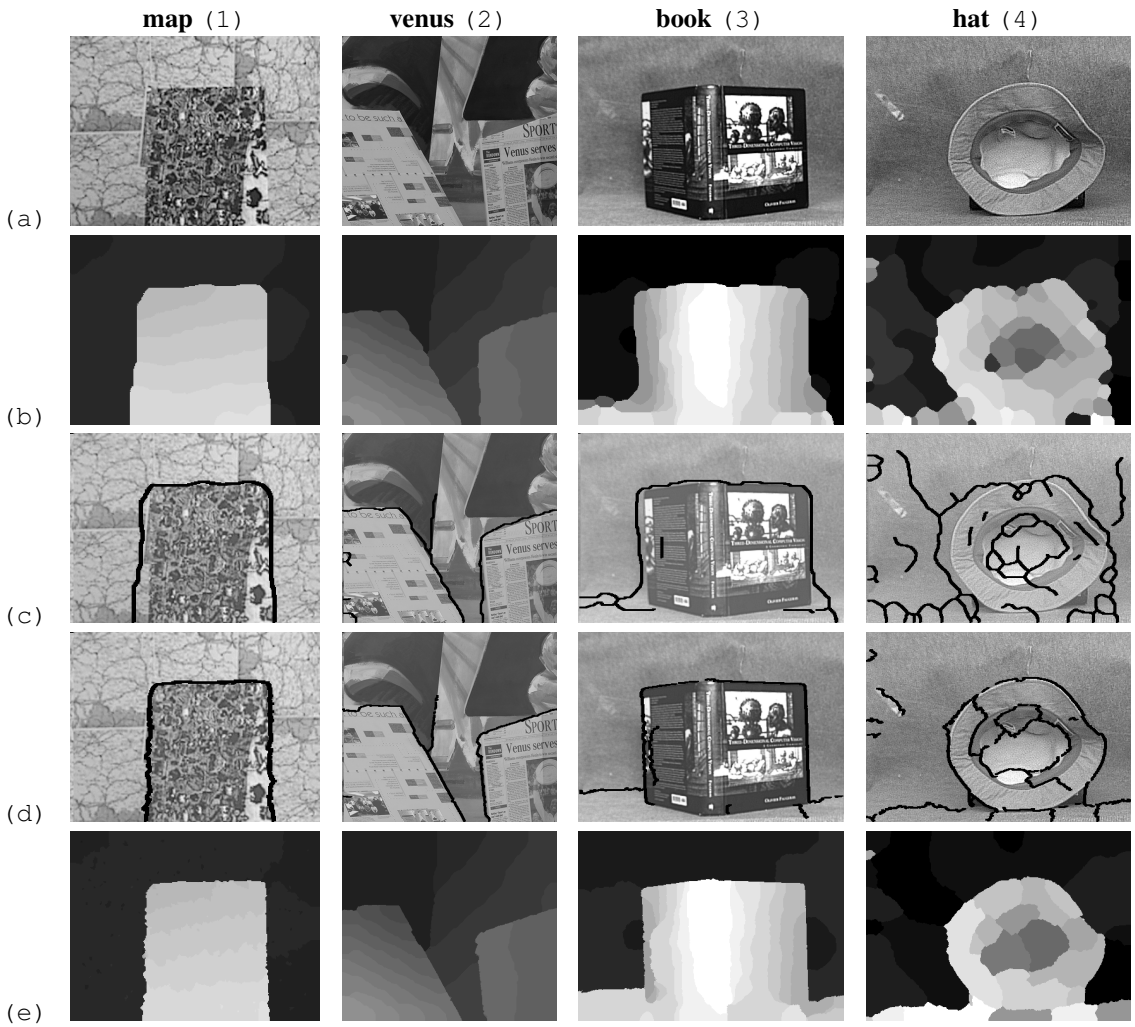

Fig. 3. (a) Original Images; (b) Disparity maps using Variational Mean-field only; (c) Disparity discontinuities corresponding to (b); (d) Corrected disparity discontinuities using our approach; (e) Corresponding disparity maps.

maps. The main originality of our work has been to define a fully Markovian model that makes explicit the relationships between disparity and object boundaries, through the introduction of a displacement field.

5. REFERENCES

[1] Andreas Klaus, Mario Sormann, and Konrad F. Karner, "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure," in *ICPR*, 2006.

[2] Q. Yang, L. Wang, R. Yang, H. Stewnius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling," in *CVPR*, 2006.

[3] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola and Lawrence K. Saul, "An Introduction to Variational Methods for Graphical Models" in *Machine Learning*, 37(2),1999.

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

[5] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient belief propagation for early vision," *IJCV*, 70(1), 2006.

[6] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov random fields," *PAMI*, 15 (12), 1993.

[7] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother, "Probabilistic fusion of stereo with color and contrast for bilayer segmentation," *PAMI*, 28 (9), 2006.

[8] J. Sun, N.N. Zheng, and H.Y. Shum, "Stereo matching using belief propagation," *PAMI*, 25 (7), 2003.

[9] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *PAMI*, 12 (7), 1990.

[10] J. Wu and A.C.S. Chung, "A segmentation model using compound Markov random fields based on a boundary model," *IP*, 16 (1), 2007.

[11] Derek Hoiem, Andrew Stein, Alexei A. Efros, and Martial Hebert, "Recovering occlusion boundaries from a single image," in *ICCV*, 2007.

[12] S. Mattoccia, F. Tombari, and L. Di Stefano, "Stereo vision enabling precise border localization within a scanline optimization framework," in *ACCV*, 2007.