Incorporating visual knowledge representation in stereo reconstruction

Adrian Barbu and Song-Chun Zhu
University of California, Los Angeles
Departments of Computer Science and Statistics
abarbu@ucla.edu, sczhu@stat.ucla.edu

Abstract

In this paper, we present a two-layer generative model that incorporates generic middle-level visual knowledge for dense stereo reconstruction. The visual knowledge is represented by a dictionary of surface primitives including various categories of boundary discontinuities and junctions in parametric form. Given a stereo pair, we first compute a primal sketch representation which decomposes the image into a structural part for object boundaries and high intensity contrast represented by a 2D sketch graph, and a structureless part represented by Markov random field on pixels. Then we label the sketch graph and compute the 3D sketch (like a wire-frame) by fitting the primitive dictionary to the sketch graph. The surfaces between the 3D sketches are filled in by computing the depth of the MRF model on the structureless part. These two levels interact closely since the MRF is used to propagate information between the primitives, and at the same time, the primitives act as boundary conditions for the MRF. The two processes maximize a Bayesian posterior probability jointly. We propose an MCMC algorithm that simultaneously infers the 3D primitive types and parameters and estimates the depth of the scene. Our experiments show that this representation can infer the depth map with sharp boundaries and junctions for textureless images, curve objects and free-form shapes.

1. Introduction

Stereo matching is an intensively studied problem, and recently there is major interest in studying it using effective algorithms such as BP[11] and graph cuts[4] based on Markov random field models and splines [8]. However, the representations (models) used in these methods are still very limited. As a result, they often produce unsatisfactory results when the images have textureless surfaces, such as indoor walls, or when the images have curve structures which do not fit to the pixel-based MRF representation. Fig.13 dis-

plays two such results using the state-of-the-art graph cut algorithm. For such images, one could see that just matching pixels using Markov Random Field priors is not enough. It is desirable to incorporate some visual knowledge into the surface representation. In the stereo literature, the 3D reconstruction of curves was studied in [14], but the depth was obtained only at the location of curve boundaries and the algorithm does not provide a dense depth estimation. A pure generative model using primitives for the entire scene was developed in [5]. We argue that using such parametric models for the entire scene limits the generality of the model. Instead, we propose to use parametric models only for the places of interest in the image, the rest being modeled by a MRF.

In this paper, we present a two-layer generative model that incorporates generic middle-level visual knowledge for dense stereo reconstruction. The overall data flow for the algorithm is illustrated in Fig. 1. Given a pair of stereo images, we first compute a primal sketch representation[7] which decomposes the image into two layers. (i) A structural layer for object boundaries and high intensity contrast represented by a 2D sketch graph, and (ii) a structureless layer represented by Markov random field on pixels. The sketch graph in the structural layer consists of a number of isolated points, line segments, and junctions which are considered vertices of different degrees of connection.
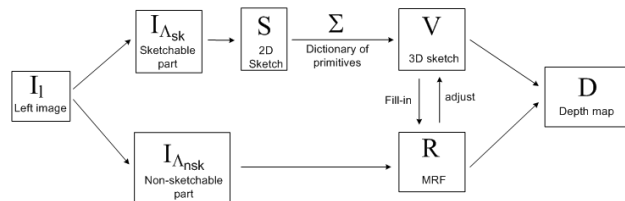

Figure 1. The flow diagram of our algorithm.

We then study the 3D structures for these points, line segments, and junctions and develop a dictionary for different configurations. The boundary primitives correspond to places where the depth map is not smooth, namely the boundaries between objects in the scene (first order discontinuities) and the places where the surface normal experi-

ences large changes in direction (second order discontinuities). The curve primitives describe thin curves of different intensity from the background, and usually represent wire-like 3D objects such as thin branches of a tree or electric cables, etc. The point primitives represent feature points in the image that have reliable depth information. The valid combinations of these 3D primitives is summarized in a dictionary of junctions. Figs. 4 and 6 shows the dictionaries of line segments and junctions respectively. Each is a 3D surface primitive specified by a number of variables. The variable number is reduced for degenerated (accidental) cases.

There are three reasons for us to study these primitives. Firstly, the primitive representation reduces the number of variables (dimension reduction). For example, a boundary primitive may have 11 pixels in width and 30 pixels in length, and is described by 4-7 variables which can be more reliably estimated from data. A similar idea was pursued in motion boundary estimation in[2]. Secondly, these sketches are the most informative parts in the images, and it is computationally more effective to compute their 3D depth early and then propagate the surface information to the textureless layers. Thirdly, the dictionary of junction primitives limits the search space by ruling out invalid configurations, like in line drawing interpretation and perceptual organization[9].

We adopt a probability model in a Bayesian framework, where the likelihood is described in terms of the matching cost of the primitives to images, while the prior has terms for continuity and consistency between the primitives, and a Markov Random Field to fill in the depth information in the structureless areas. This Markov Random Field together with the labeling of the edges can be thought of as a Mixed Markov Model [6], in which the neighborhood structure of the MRF depends upon the primitive types, and changes dynamically during the computation.

The inference algorithm simultaneously finds the types of the 3D primitives, their parameters and the full depth map. To make-up for the slowdown given by the long range interactions between the primitives through the MRF, the algorithm makes use of data driven techniques to propose local changes (updates) in the structureless areas.

The paper is organized as follows. We present the representation and dictionary of primitives in Section 2, and then discuss the algorithm in Section 3. We show results in Section 4, and conclude the paper by a discussion in Section 5.
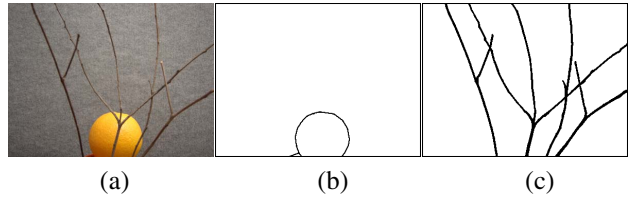
2. A two layer representation

Given a stereo pair I_l, I_r of images, we are required to find the depth of all pixels in I_l. Assuming that the camera parameters are known, this is equivalent to finding for each pixel, the horizontal disparity that matches it to a corresponding pixel in I_r. Let D be the disparity map that needs to be inferred and Λ be the pixel lattice.

We assume the disparity map D is generally continuous and differentiable, with the exception of a number of curves Λ_{sk} where the continuity or differentiability assumption does not hold. These curves are augmented with disparity values and are considered to form a 3D sketch D_s that acts as boundary conditions for the Markov Random Field modeling the disparity on $\Lambda_{nsk} = \Lambda \setminus \Lambda_{sk}$.

2.1. The sketch layer – from 2D to 3D

We assume that the places where the disparity is discontinuous or non-differentiable are among the places of intensity discontinuity. The intensity discontinuities are given in the form of a sketch S consisting of a region layer S_R and a curve S_C layer, as illustrated in figure 2. These sketches can be obtained as in [7, 13]. The sketch edges are approximated with line segments $S = \{s_i, i = 1, .., n_e\}$. The segments that originated from the region layer $s_i \in S_R$ will be named *edge segments* while the segments originating from the curve layer $s_i \in S_C$ will be named *curve segments*.

Figure 2. Our algorithm starts from a two layer sketch representation. (a) input image, (b) region layer, (c) curve layer.

Each edge segment $s_i \in S_R$ from the region layer is assigned two 5 pixel wide regions l_i, r_i, on the left respectively on the right of s, as shown in Figure 3. Each curve segment $s \in S_C$ is assigned a region $r(s)$ along the segment, of width equal to the width of the curve. Then all line segments $s_i \in S$ are augmented parameters to become 3D sketch primitives, as shown in figure 4. Depending on the type of segments they originated from, there are *boundary primitives* and *curve primitives*.
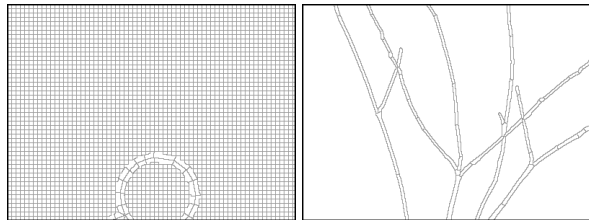
Figure 3. Division of the image in Fig. 2 into sketch primitives and 6x6 pixel regions. Region layer (left) and curve layer (right).

Because away from the places of discontinuity, the surfaces are in general very smooth and to reduce the dimensionality of the problem, the pixels of Λ_{nsk} are grouped into square regions of size 6×6 pixels, by intersecting Λ_{nsk} with

a 6×6 rectangular grid. Small regions at the boundary between the primitives and the rectangular grid are merged to the primitives. This way, the entire lattice Λ is divided into atomic regions that either are along the sketch S, or are on the rectangular grid, as shown in Figure 3. This structure allows the use of the thin plate spline model for the MRF and also enables implementation of good boundary conditions by the 3D primitives.
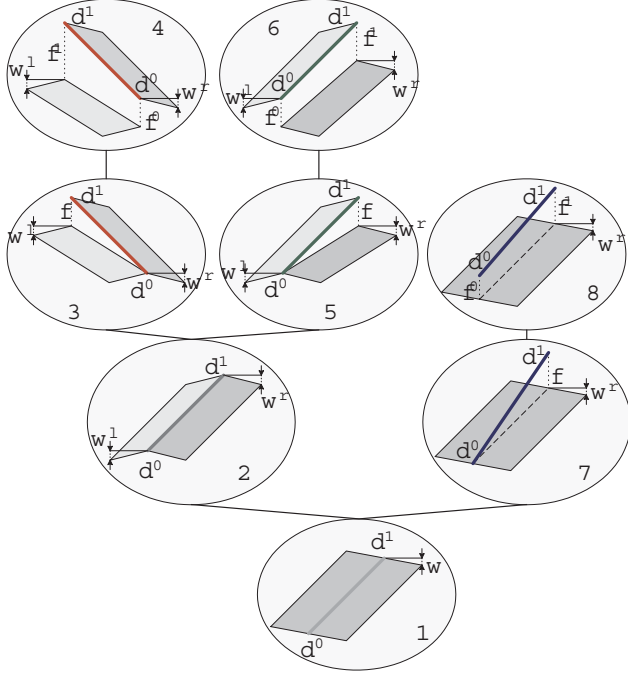
Figure 4. Each sketch segment is augmented to a primitive from the following dictionary, order by generality.

Let
$$V_1 = \{\pi_i = (s_i, [l_i], r_i, t_i, d_i, w_i[, f_i]), i = 1, .., n_e\} \quad (1)$$
be the set of all primitives, where the parameters in brackets might be missing, depending on the primitive type, as described below:

1. an occlusion label o_i^l, o_i^r for each of the regions l_i, r_i, representing whether they are occluded (0) or not (1).

2. a label $t_i = t(\pi_i) \in \{1, .., 8\}$ indexing the type of the primitive from the primitive dictionary with the restriction that edge segments $s_i \in S_R$ can only be assigned labels from $\{1, .., 6\}$ while curve segments $s_i \in S_C$ can only be assigned labels from $\{1, 7, 8\}$.

   - Type 1 represents edges or curves that are on the surface of the objects.
   - Type 2 represents places where the surface is continuous but the normal is discontinuous.
   - Types 3, 4, 5, 6 represent occluding edges where the occluded surface is on the left (types 3, 4) or on the right (types 5, 6) of the edge.

   - Types 7, 8 represent 3D curves, either connected with one end to the surface behind, or totally disconnected.

3. a label p_i specifying whether this primitive is a control point (value 1) of the thin plate spline or not (value 0). All horizontal edges have $p_i = 0$ at all times.

4. the disparities $d_i = d(\pi_i) = (d_i^0, d_i^1)$ at the endpoints of the segment.

5. the left and right control arm $w_i = w(\pi_i) = (w_i^l, w_i^r)$ representing the slope of the disparity map D in the direction perpendicular to the line segment.

6. for types 3-6, the disparity $f_i = f(\pi_i) = (f_i^0, f_i^1)$ of the occluded surface at the ends of the segment.

Each of the regions l_i, r_i of a primitive $\pi_i = (s_i, [l_i], r_i, t_i, d_i, w_i[, f_i])$ is assigned a matching cost
$$c(r_i, d) = \begin{cases} \sum_{v \in r_i} |I_l(v) - I_r(v - d_v(\pi_i))| & \text{if } o_i^r = 1 \\ \alpha & \text{else} \end{cases} \quad (2)$$

where for each pixel $v \in r_i$, $d_v(\pi_i)$ is the linear interpolation based on the disparity d at the endpoints of the region, in the assumption that $w = 0$. Then the matching cost of the primitive π_i is
$$c(\pi_i) = \begin{cases} c(r_i, d_i) & \text{if } t_i = 7, 8, 1(\text{curve}) \\ c(l_i, d_i) + c(r_i, d_i) & \text{if } t_i = 2, 1(\text{region}) \\ c(l_i, f_i) + c(r_i, d_i) & \text{if } t_i = 3, 4 \\ c(l_i, d_i) + c(r_i, f_i) & \text{if } t_i = 5, 6 \end{cases} \quad (3)$$
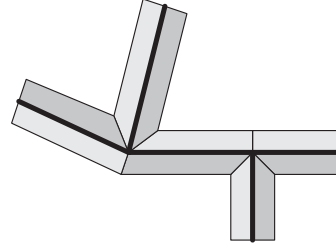

Figure 5. A set of edge segments (black) and their 3D primitives (gray). The primitives form a graph by the adjacency of the segments.

The boundary primitives form a graph by the natural adjacency relation between the underlying edge segments, as can be seen in Figure 5. To further restrict the search space, the junction points of two or more primitives are modeled. Similar to [9] we will have certain types of possible junctions depending on the degree (number of primitives) of the junction, as mentioned below and illustrated in Fig. 6.

- junctions of 2 boundary primitives have three main types: Surface junctions, beginning of occlusion and occlusion junctions.

- junctions of 3 boundary primitives have three main types: Surface junctions, Y-junctions and T-junctions.

- junctions of 4 or more boundary primitives are accidental and are assumed to be all surface junctions.

- we assume there are no junctions between one or two curve primitives and one boundary primitive

- junctions of 1 curve primitive with two boundary primitives have three main types: curve beginning, Y-junctions and T-junctions.

- junctions of 2 curve primitives have only one type.

- junctions of 3 curve primitives have only one type, namely bifurcation.

- junctions of 4 curve primitives have two types, namely curve crossing or curve overlapping. In both cases, the opposite primitives can be seen as part of the same 3D curve.
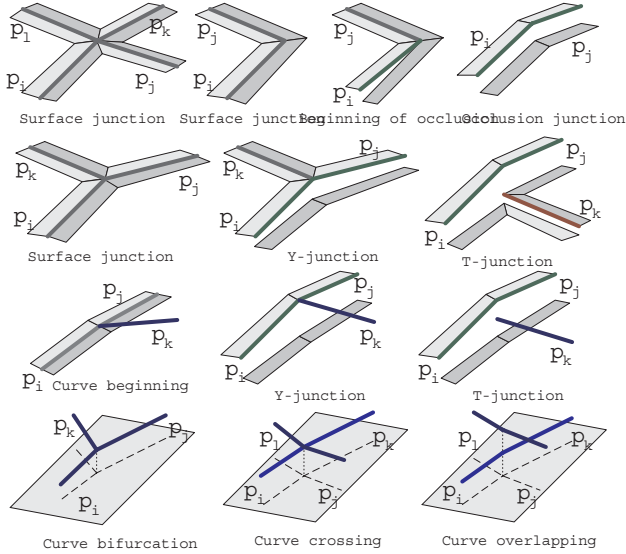
Figure 6. These are the main types of junctions between boundary and curve primitives.

Let $J = \{\phi_i = (k, \pi_{i_1}, ..., \pi_{i_k}), \pi_{i_1}, ..., \pi_{i_k} \in V_1, i = 1, ..., n_J\}$ be the set of junctions, each containing the list of primitives that are adjacent to it. Each junction $\phi_i \in J$ is assigned a prior model defined in terms of the junction type and the directions of the 3D primitives $\pi_{i_1}, ..., \pi_{i_k}$ meeting in this junction, as illustrated in Figure 7,b). This prior model assigns a probability $P(\phi_i)$ to each combination of primitive parameters at junction ϕ_i. This prior is composed of a 3D geometric prior on the primitives and a 2D occurrence prior of each particular junction type. Thus

$$P(\phi) \propto P(\phi^{3D}|t, \phi^{2D})P(t|\phi^{2D}) \qquad (4)$$

since the 2d geometry ϕ^{2D} of the junction is fixed.

The prior $P(\phi^{3D}|t, \phi^{2D})$ is specific for each junction type, and depends on the two continuity priors:

$$p_c(\pi_i, \pi_j) \propto e^{-\beta_c|d_i^\phi - d_j^\phi|^2}$$

$$p_s(\pi_i, \pi_j) \propto e^{-\beta_c|d_i^\phi - d_j^\phi|^2 - \beta_s(|d_i^{\overline{\phi}} - 2d_i^\phi + d_j^\phi|^2 - |d_i^{\overline{\phi}} - 2d_j^\phi + d_j^{\overline{\phi}}|^2)} \qquad (5)$$

where $d_i = (d_i^\phi, d_i^{\overline{\phi}})$ is the disparity of the primitive π_i, with d_j^ϕ being the disparity at the junction ϕ endpoint. For example, the surface junction from Fig. 7 has

$$P(\phi^{3D}|t, \phi^{2D}) \propto p_c(\pi_j, \pi_k)p_c(\pi_j, \pi_l)p_c(\pi_k, \pi_l) \qquad (6)$$
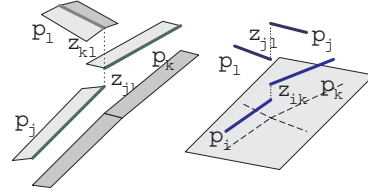
Figure 7. The prior at each junction between primitives, encourages 3D continuity of the line and curve primitives.

while the curve overlapping junction from Fig. 7 has

$$P(\phi^{3D}|t, \phi^{2D}) \propto p_s(\pi_i, \pi_k)p_s(\pi_j, \pi_l) \qquad (7)$$

2.2. The free-form layer

The primitives $\pi \in V_1$ discussed in the previous section are elongated primitives corresponding to line segments, so they can be considered of dimension 1. Other sketch primitives that are involved in the free form layer are the zero dimensional primitives corresponding to feature points with reliable disparity information, i.e. *point primitives*. These primitives are a subset of the rectangular atomic regions, and together with the one dimensional boundary primitives are the control points of the thin plate spline. The curve primitives are not involved in the MRF computation.

Let R be the set of all rectangular atomic regions.

For each region $r \in R$, we compute a saliency map

$$\rho_r(d) \propto \exp(-\sum_{v \in r} |I_l(v) - I_r(v - d)|/10) \qquad (8)$$

to all possible disparities $d \in [d_{\min}, d_{\max}]$
Then the rectangular regions

$$R = \{r_i = (d_i, s_i, \mu_i, \sigma_i^2), i = 1, .., n_r\} \qquad (9)$$

have the following parameters:

1. the disparity $d_i = d(r_i)$ of the center of the region

2. a label o_i specifying whether this region is occluded (value 0) or not (value 1).

3. a label $p_i = p(r_i) \in \{0, 1\}$ representing whether the region is a point primitive (i.e. control point for the thin plate spline) or not.

4. the mean μ_i and variance σ_i^2 of the saliency map ρ_{r_i}.

Following [1] the occlusion labels are deterministically decided. The matching cost for each region $r_i \in R$ is

$$c(r_i) = \begin{cases} \alpha & \text{if } o_i = 0 \\ \sum_{v \in r_i} |I_l(v) - I_r(v - d_i))| & \text{if } o_i = 1 \end{cases} \qquad (10)$$

The set of point primitives is denoted by

$$V_0 = \{r_i \in R, p_i = 1\}. \tag{11}$$

In Figure 8 are illustrated the point and boundary primitives that act as control points for the Λ_{nsk} part. The depth and disparity maps obtained this way are shown in Figure 14, obtained from V_1 and R by interpolation using the MRF.
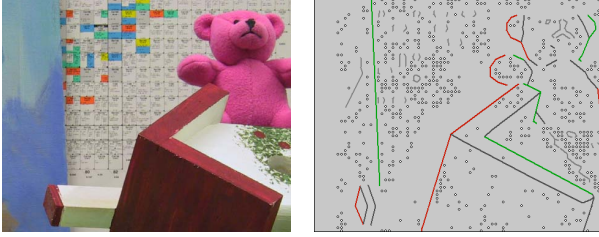
Figure 8. Left image of a stereo sequence and the control points (point and boundary primitives) of the thin plate spline.

By using the boundary primitives to model the places of discontinuity, the obtained disparity map has crisp discontinuities at the object boundaries and is smooth everywhere else, as shown in Figure 14.

2.3. Bayesian formulation

We formulate our model using the Bayes rule:

$$P(V_1, R|I_l, I_r) = P(I_l|I_r, V_1, R)P(R - V_0|V_0, V_1)P(V_0, V_1) \tag{12}$$

The likelihood $P(I_l|I_r, V_1, R)$ is expressed in terms of the matching cost $c(\pi_i), c(r_j)$ of the sketch primitives.
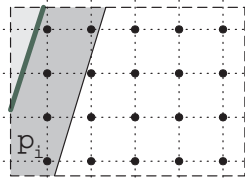
$$P(I_l|I_r, V_1, R) \propto \exp[-\sum_{i=1}^{n_e} c(\pi_i) - \sum_{r_j \in R} c(r_j)] \tag{13}$$

The prior

$$P(R - V_0|V_0, V_1) \propto \exp[-E_c(R) - \beta_b E_b(R, V_1)] \tag{14}$$

is defined in terms of the energy of the soft control points:

$$E_c(R) = \sum_{r_j \in V_0} (d_j - \mu_j)/2\sigma_j^2 \tag{15}$$

Figure 9. The region not covered by boundary primitives has a thin plate spline prior, computed on a rectangular grid that intersects the wings (atomic regions) of the primitives.

and the thin plate bending energy:

$$E_b(R, V_1) = \sum_{(x,y) \in G} [d_{xx}^2(x, y) + 2d_{xy}(x, y)^2 + d_{yy}^2(x, y)] \tag{16}$$

which is computed on a 6×6 grid G containing the centers of all the rectangular regions and neighboring grid points on the boundary primitives. For example, if the point $(x, y) \in G$ is the center of $r_j \in R$ and $r_N, r_{NW}, r_W, r_{SW}, r_S, r_{SE}, r_E, r_{NE}$ are the 8 neighbors of r_j, then

$$d_{xx}(x, y) = d_W - 2d_j + d_E$$
$$d_{yy}(x, y) = d_N - 2d_j + d_S$$
$$d_{xy}(x, y) = (d_{NE} + d_{SW} - d_{NW} - d_{SE})/4$$

Similar terms in the bending energy $E_b(R, V_1)$ can be written for cases where one or many of the neighbors are boundary primitives. However, there are no terms involving the left and right atomic regions $l_i, r_i \in \pi_i$ belonging to the same edge primitive π_i.

The prior $P(V_0, V_1) = P(V_0)P(V_1)$ assumes a uniform prior on V_0 while $P(V_1)$ encourages continuity of the 3D boundary primitives.

$$P(V_1) = \prod_{\phi_i \in J} P(\phi_i) \tag{17}$$

where the junction prior $P(\phi_i)$ was defined in Sect. 2.2.

3. The inference algorithm

There are two types of variables that exist in this problem formulation, discrete and continuous. The discrete variables are
$$\Delta = V_1^d \cup R^d \tag{18}$$
consisting of $V_1^d = \{(t(\pi), o^l(\pi), o^r(\pi), p(\pi)), \forall \pi \in V_1\}$ and $R^d = \{(o(r), p(r)), \forall r \in R\}$. All other variables are continuous, namely $V_1^c = V_1 \setminus V_1^d$ and $R^c = R - R^d$, and can be divided into the boundary conditions

$$\Gamma = V_0^c \cup \{d(\pi), \forall \pi \in V_1, p(\pi) = 1\} \tag{19}$$

and the fill-in variables
$$\Psi = \{([w(\pi)], [f(\pi)]), \forall \pi \in V_1\} \cup$$
$$\{d(\pi), \forall \pi \in V_1, t(\pi) = 1\} \cup R^c - V_0^c. \tag{20}$$

The posterior probability can then be written as

$$p(V_1, R|I_l, I_r) = p(\Delta, \Gamma, \Psi|I_l, I_r) \tag{21}$$

In a MAP formulation, our algorithm performs three tasks simultaneously:

- Reconstructs the 3D sketch to infer the parameters Γ of the primitives.

- labels the graph to infer the discrete parameters Δ, i.e. associates the primitives with the appropriate types. This represents the detection of surface boundaries and of the feature points of the image.

- Performs "fill in" of the remaining parts of the image, using the MRF and Γ, Δ as boundary conditions, to infer Ψ and obtain a dense disparity map D.

The system is initialized by using an approximation of the posterior that only takes into account the matching cost of the edge regions $\pi_i \in V_1$ and the junction prior.

$$P(V_1|I_l, I_r) \propto \prod_{i=1}^{n_e} L_{\pi_i}(t_i) \prod_{\phi_i \in J} P(\phi_i) \qquad (22)$$

The initialization algorithm contains three types of moves:

- a single node move changing one variable d_i at a time.

- a move that simultaneously shifts all d_i at a junction ϕ by the same value.

- a labeling move similar to the MCMC algorithm described below, accepted based on the posterior probability from Eq. (22).

It quickly propagates the depth information along the sketch and obtains a good initialization as seen in Fig. 10.
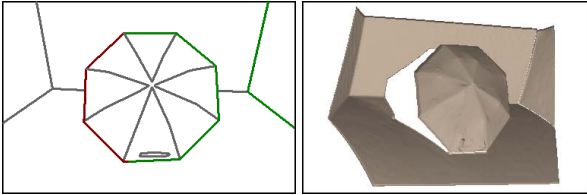
**Figure 10. Initialization is obtained by propagating the junction priors along the sketch.**

The fill-in variables Ψ can be computed analytically, since for fixed Δ, Γ, the conditional probability $\log(P(\Psi|\Delta,\Gamma))$ is a quadratic function in Ψ. This implies that Ψ can be regarded as a function on Δ, Γ, $\Psi = \psi(\Delta, \Gamma)$. This restricts the problem to maximizing the probability $P(\Delta, \Gamma, \psi(\Delta, \Gamma)|I_l, I_r)$, of much smaller dimensionality.
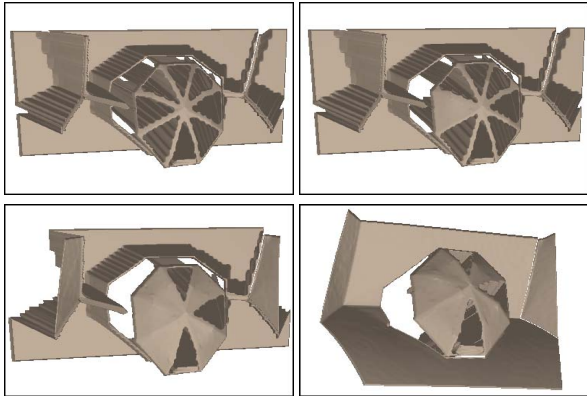
Figure 11. The fill-in can be restricted to regions bounded by control point boundary primitives. In a few steps, the initial 3D reconstruction before graph labeling is obtained.

However, minimizing $-\log(P(\Psi|\Delta,\Gamma))$ analytically involves finding the inverse of a $n \times n$ matrix where $n = |\Psi|$. This can be computationally expensive if all Ψ is updated, since Ψ is on the order of $|\Psi| \sim 4000$. But since inside each

of the regions bounded by the control point sketch primitives, the variables depend only on the control points inside and on the boundary of this region, the computation can be localized to each of regions independently, as shown in Figure 11, and the computation demand will be much lower. Shown in Fig. 11 are the 3D reconstructions after 0,1,4,5 connected components have been updated. The horizontal edges change the disparity at the same time with the interior lattice, because they are not control points.
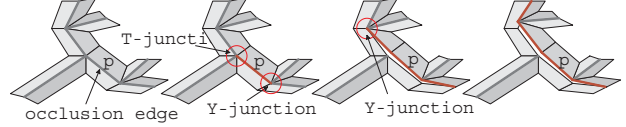
Figure 12. Each graph labeling move changes the types of a set of primitives in a consistent manner in a growing process. Illustrated is the left side of the umbrella image.

After the initialization, the depth of the sketch primitives $\{d(\pi) \in \Gamma\}$ will be fixed.

To maximize $P(\Delta, \Gamma, \psi(\Delta, \Gamma)|I_l, I_r)$ we will use a Markov chain Monte Carlo algorithm that will sample $P(\Delta, \Gamma, \psi(\Delta, \Gamma)|I_l, I_r)$, and this way obtain the most probable solutions.

Based on the matching cost, for each primitive we compute a likelihood $L_\pi(\pi)$ toward the different primitive types. Using this intensity-driven likelihood, we construct a likelihood, driven simultaneously by the image intensity and the geometry (relative position of primitives), for each junction $\phi = \{\pi_1, ..., \pi_k\}$:

$$L_\phi(t) = P(\phi)L_{\pi_1}(t_1)...L_{\pi_k}(t_k) \qquad (23)$$

These likelihoods are used to propose a coherent set of primitives in each MCMC move.

At each step, the algorithm proposes, as shown in Figure 12, new types for a set of primitives N and junctions J in one move, as follows:

1. Grow a set N of primitives as follows:
   1. Choose a random non-horizontal primitive π
   2. Initialize $N = \{\pi\}$ and $J = \{\phi_1, \phi_2\}$ where ϕ_1, ϕ_2 are the two junctions adjacent to π.
   3. Sample the primitive type $t(\pi)$ from the likelihood $L_\pi(t)$.
   4. Sample the type of $\phi \in J$ from $L_\phi(t)$, conditional on the primitive type $t(\pi)$. This determines the types of all primitives of $N_n = \{\pi' \notin N, \pi' \sim \phi$ for some $\phi \in J\}$,where $\pi \sim \phi$ means π is adjacent to ϕ.
   5. Set $N \leftarrow N \cup N_n$.
   6. Initialize $J_n = \emptyset$.
   7. For each $\pi \in N_n$, pick the adjacent junction $\phi \notin J$. If π changed its type at step 4, set $J_n \leftarrow J_n \cup \{\phi\}$, else set $J_n \leftarrow J_n \cup \{\phi\}$ with probability 0.5.
   8. Set $J \leftarrow J \cup J_n$.
   9. Repeat steps 4-8 for each $\pi \in N_n$ and each $\phi \in J_n, \pi \sim \phi$.
2. Update the fill-in variables $\Psi(\Delta, \Gamma)$ where it is necessary.
3. Accept the move based on the posterior $p(V_1, R|I_l, I_r)$.

4. Experimental results

The experiments are presented in Fig.14 where five typical images for stereo matching are shown. The first two have textureless surfaces and the most information is from the surface boundaries. The fourth image has curves (twigs). For these three images, it is not a surprise to see that the graph cut method with simple MRF models on pixels produce unsatisfactory results (see Fig.13 for comparison). The third and fifth images have free-form surfaces with or without textures from [8]) and [10].
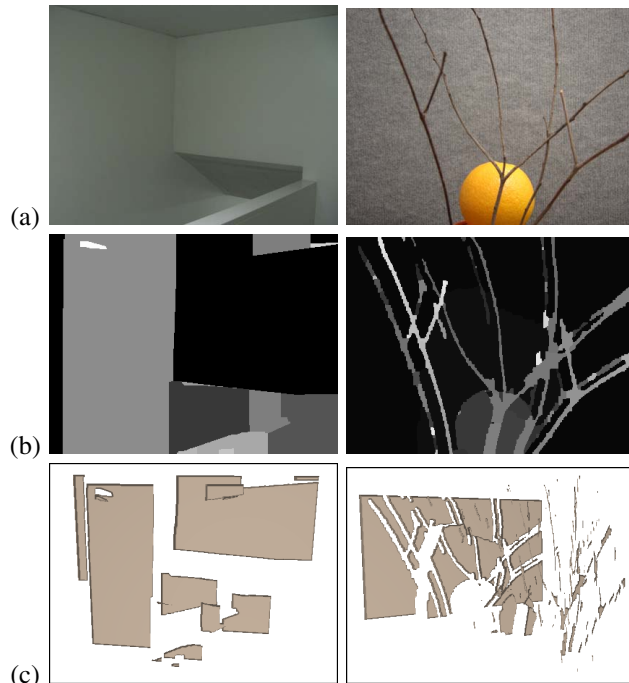
Figure 13. Two comparison examples using the *graph cuts algorithm* **on scenes containing textureless surfaces and thin curve structures. (a) left image, (b) disparity map, (c) 3d map. Our results are shown in Fig.14.**

We have also shown the interactions of the two layers and the effects of sketch labeling in Fig. 11.

5. Discussion

In this paper we presented a two-layer generative representation for dense stereo reconstruction. By studying a dictionary of boundary and junction primitives, we incorporate some generic middle level visual knowledge to obtain accurate depth at object boundaries, junctions, and to handle 3D curves structures such as twigs, cables etc., thus our method being applicable on a broad range of images. The two layers of Markov random field models– one on the sketch layer and one on the pixel layer work collaboratively in optimizing a Bayesian posterior probability. Our method achieve satisfactory results in comparison to the state of the art graph cut method.

References

[1] P. Belhumeur, "A Bayesian Approach to Binocular Stereopsis" *IJCV* 19(3)237-260, 1996

[2] M. Black and D.J. Fleet, "Probabilistic detection and tracking of motion boundaries", *IJCV* 38(3) 231-245, 2000

[3] A. Blake, A. Zisserman, "Visual Reconstruction", em MIT Press, 1987

[4] Y. Boykov, O. Veksler, and R. Zabih. "Fast approximate energy minimization via graph cuts". *IEEE Trans. on PAMI* vol. 23, no. 11, pp 1222-1239, 2001

[5] A. R. Dick, P.H.S. Torr and R. Cipolla, "Modelling and Interpretation of Architecture from Several Images", *IJCV*, **60** (2), 111-134, 2004.

[6] A. Fridman, "Mixed Markov Models", *Proc Nat Acad Sci*, 100 (14) 8092-8096, 2003

[7] C. Guo, S.C. Zhu and Y.N. Wu, "A Mathematical Theory of Primal Sketch and Sketchability", *ICCV* 2003.

[8] M. Lin and C. Tomasi, "Surfaces with occlusions from layered stereo". *IEEE Trans. on PAMI*, 26 (8), 710–717, 2004

[9] E. Saund, "Perceptual organization of occluding contours of opaque surfaces", *Computer Vision and Image Understanding* 1999 October; 76 (1): 70-82.

[10] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light", *CVPR* 2003

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

[12] H. Tao, H. S. Sawhney and R. Kumar, "A global matching framework for stereo computation", *ICCV*, 2001

[13] Z. Tu and S.C. Zhu, "Parsing Images into Regions, Curves, and Curve Groups", IJCV(under review), http://www.stat.ucla.edu/∼ztu/.

[14] G. Li and S. W. Zucker, "A Differential Geometrical Model for Contour-Based Stereo Correspondence" *IEEE Workshop on Variational, Geometric, and Level Set Methods in Computer Vision*, Nice, 2003.
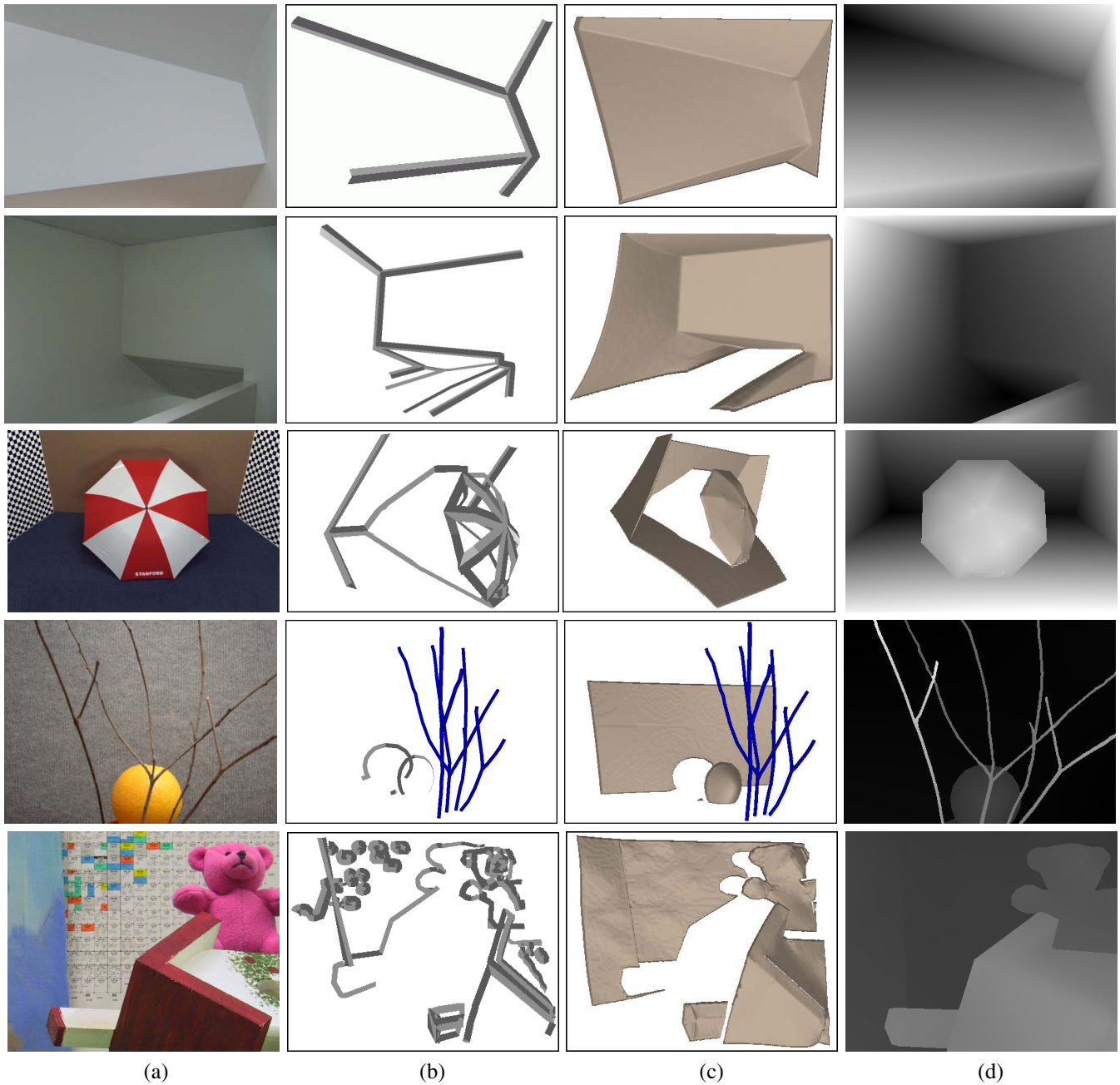
Figure 14. Results obtained using our method. (a) left image of the stereo pair, (b) 3D sketch using the primitives, (c) 3D depth map, (d) disparity map.