

# Bayesian Reconstruction of 3D Shapes and Scenes From A Single Image

Feng Han and Song-Chun Zhu  
University of California, Los Angeles  
Departments of Computer Science and Statistics  
hanf@cs.ucla.edu, sczhu@stat.ucla.edu

## Abstract

It's common experience for human vision to perceive full 3D shape and scene from a single 2D image with the occluded parts "filled-in" by prior visual knowledge. In this paper we represent prior knowledge of 3D shapes and scenes by probabilistic models at two levels – both are defined on graphs. The first level model is built on a graph representation for single objects, and it is a mixture model for both man-made block objects and natural objects such as trees and grasses. It assumes surface and boundary smoothness, 3D angle symmetry etc. The second level model is built on the relation graph of all objects in a scene. It assumes that objects should be supported for maximum stability with global bounding surfaces, such as ground, sky and walls. Given an input image, we extract the geometry and photometric structures through image segmentation and sketching, and represent them in a big graph. Then we partition the graph into subgraphs each being an object, infer the 3D shape and recover occluded surfaces, edges and vertices in each subgraph, and infer the scene structures between the recovered 3D sub-graphs. The inference algorithm samples from the prior model under the constraint that it reproduces the observed image/sketch under projective geometry.

## 1. Introduction

Computing 3D object shapes and complex scene structures from 2D image(s) is a fundamental problem in computer vision and has been extensively studied. There are three main streams of research in the literature.

The first is 3D reconstruction from line drawings [8, 9, 11, 10, 4]. Early work used deterministic rules for categorizing boundary and vertex, and lately the focus has been shifted to compute 3D shapes through energy minimization. Existing line drawing work assumes single man-made object, such as polyhedra, from manual input. Especially all hidden lines have to be inputted and their occlusion relation-

ship specified manually. These assumptions prevent the algorithms from working on real images.

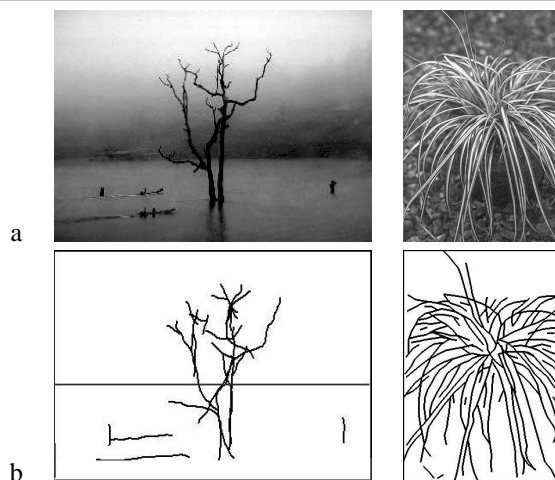
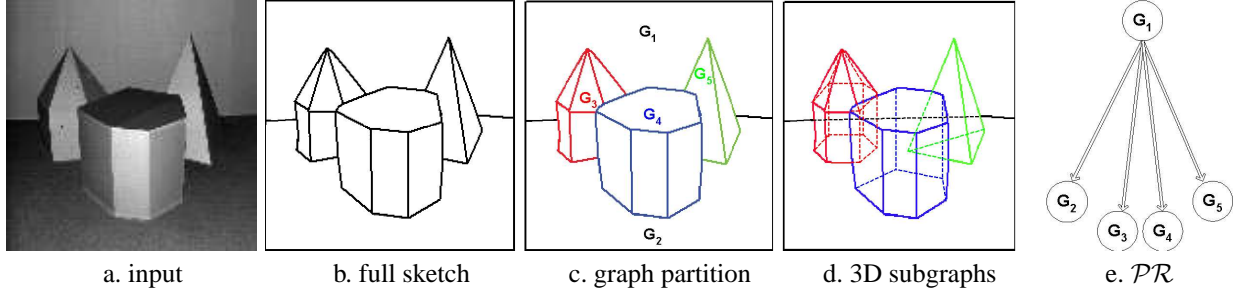


Figure 1. a. Two input images. b. Their primal sketches obtained by a segmentation and sketching algorithm [3, 6].

The second is to compute 2.5D depth map using shape-from-shading, texture and defocus etc or compute 3D models with user-interactions from a single image [13, 12, 14, 15, 19]. For the methods to compute depth map, the result has yet to be parsed into objects and 3D shapes have yet to be computed. Also the photometric cues are rather weak and some global prior models must be engaged to yield useful results. For the methods using some user-interactions, they can't meet the needs to be fully automatic.

The third is to compute 3D shape from multiple images. For example, structure from Motion (SFM), multiple-view stereo, space carving etc [18, 17, 16, 5, 20]. As it is well stated in the photo-hull theory [16], these methods seek constraints from a large number of images, taken from well-



**Figure 2. a. input image. b. a graph representation by segmentation algorithm in [2]. c. The graph is partitioned into five subgraphs for the five objects: wall, ground, and three block objects. d. Hidden structures ((dashed lines) are recovered for each subgraph. e. Spatial support relation between five objects in a graph. (Please view the graph in color on screen).**

controlled camera positions, to minimize the uncertainty in 3D reconstruction. Most of such algorithms often use very little prior knowledge about object shape and scene, even though there are a few impressive trials [20].

Existing methods can hardly be applicable to generic natural scenes like the images shown in Fig.1. Due to the complexity and concavity of the objects, like trees and grasses, one has to capture an enormous number of images to constrain their 3D shapes, and some views are impossible or impractical to access without disturbing the objects, such as water and grasses.

These difficulties are in sharp contrast to human vision perception. One can perceive full 3D shapes and scene structures from a single 2D image, like those in Fig.1, with the occluded parts “filled-in” using prior visual knowledge. Our perceived 3D structures may only be approximately correct, but they are sufficient for many vision and graphics tasks. This contradiction between human vision and machine vision suggests that we should make use of prior models on 3D shapes and scenes.

In this paper we represent prior knowledge of 3D shapes and scenes by probabilistic models at two levels – both are defined on graphs.

The first level model is built on a graph representation for single object, and it is a mixture model for both man-made block objects and natural objects such as trees and grasses. It assumes surface and boundary smoothness, 3D angle symmetry, like line drawing work [9, 11].

The second level model is built on the relation graph of all objects in a scene. It is a mixture model for both indoor and outdoor scenes. It assumes that objects should be supported for maximum stability, i.e. maximum surface contact and alignment, and use global bounding surfaces, such as ground, sky and walls. This supporting relation is partially ordered and represented by a graph.

Fig.2 shows a simple example for illustration, and more

examples are shown in Figs.6 and 7. Given an input image in Fig.2.a, we extract the geometry and photometric structures through image segmentation and sketching, and represent them in a big graph in Fig.2.b. Then we partition the graph into subgraphs each being an object in Fig.2.c. Meanwhile we infer the 3D shape and recover occluded surfaces, edges and vertices in each subgraph which are shown by the dashed lines in Fig.2.d. We also infer the scene structures with a supporting relation graph in Fig.2.e for the recovered 3D sub-graphs. These 3D structures can be used for scene editing, augment reality, image rendering etc.

The inference algorithm samples from the prior model under the constraint that it reproduces the observed image under projective geometry. We adopt stochastic algorithms for graph partition, death-birth of hidden vertices, edges, and surfaces, which are implemented by reversible jumps [24].

## 2. Problem formulation

We formulate the problem as Bayesian inference and proceed in two steps.

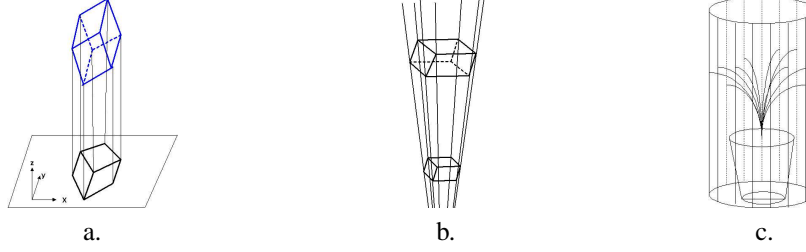
### 2.1. From an input image to a full primal sketch

Let  $\mathbf{I}$  be an input 2D image, we first apply an image segmentation and sketching algorithm [2, 3, 6] to compute a full primal sketch  $\mathbf{S}$  in a Bayesian formulation.  $\mathbf{S}$  contains two layers: a region layer  $\mathbf{S}^r$  in the background and a curve layer  $\mathbf{S}^c$  in the foreground.

$$\mathbf{S} = (\mathbf{S}^r, \mathbf{S}^c) \sim p(\mathbf{I}|\mathbf{S})p(\mathbf{S}) \quad (1)$$

The region layer partitions the image lattice into a number of  $K^r$  regions and is represented by a planar graph

$$\mathbf{S}^r = (V, E, F), \quad V = \{p_i = (x_i, y_i) : i = 1, 2, \dots, |V|, \}.$$



**Figure 3. a. Orthogonal projection, b. Perspective projection. c. Spatial model of grass/tree**

where  $V$ ,  $E$  and  $F$  are respectively the set of vertices, edges and faces. For polyhedra objects, the vertices  $V$  are junctions, and edges in  $E$  are line segments. But for natural scenes,  $V$  also includes knots with degree  $d(v) = 2$  to represent region boundaries. Each face of the planar graph is a region  $R$  and we record the image intensity in each region

$$F = \{(R_i, \mathbf{I}_{R_i}) : i = 1, 2, \dots, K^r\}.$$

The curve layer consists of a number of  $K^c$  curves

$$\mathbf{S}^c = (K^c, C_1, C_2, \dots, C_{K^c}).$$

Each curve  $C$  is actually a degenerated 1D region and is represented by a sequences of  $L$  knots (vertices) with attributes such as curve width  $w$  and intensity profile  $\rho$  perpendicular to the curve.

$$C_i = (L_i, \{(p_{ij}, w_{ij}, \rho_{ij}) : j = 1, 2, \dots, L_i\})$$

As the regions and curve profiles preserve the intensity information,  $\mathbf{S}$  is considered an augmentation from  $\mathbf{I}$  with almost no loss of information. The prior  $p(\mathbf{S})$  specifies curve and boundary smoothness etc. The algorithms of segmentation and sketching are referred to [2, 3, 6]. For example, in Fig.1, the lake-tree image has two regions in  $\mathbf{S}^r$  – the lake and the sky, and  $\mathbf{S}^c$  includes some free curves and two trees. The grass image has only one background region and the rest are curves. The polyhedra image in Fig. 2 has only  $\mathbf{S}^r$  and  $\mathbf{S}^c$  is empty.

## 2.2. From full primal sketch to 3D

The full 3D scene is represented by  $K$  objects each being a 3D graph  $G_i$ , and by relation  $\mathcal{PR}$  among the 3D graphs.

$$W = (K, \{G_1, G_2, \dots, G_K\}, \mathcal{PR})$$

a graph  $G = (V_i, E_i, F_i)$  is represented by (1)  $V_i$  – 3D vertices, (2).  $E_i$  – edges that may include attributes such as width and intensity profile for the curve processes, and (3).  $F_i$  – surfaces with albedo or intensity patterns (or 3D curves for degenerated surfaces).

$$V = \{v_i = (x_i, y_i, z_i) : i = 1, 2, \dots, |V|\}$$

$$E = \{(e_m, w_m, \rho_m) : e_m = (v_s, v_t), m = 1, 2, \dots, |E|\}$$

$$F = \{(f_n, \rho_n) : f_n = (e_{n1}, e_{n2}, \dots, e_{nk})\}$$

$\mathcal{PR}$  is a partially ordered relation for the set of objects

$$\mathcal{PR} = (\{G_1, \dots, G_K\}; \preceq).$$

An element  $\langle G_i, G_j \rangle \in \mathcal{PR}$  means  $G_i \preceq G_j$ , i.e. object  $G_i$  supports object  $G_j$ . Figs.2.d-e show a representation of  $W$  with  $K = 5$  objects.

Our objective is to compute an optimal representation by maximizing (or sampling) a Bayesian posterior in a solution space  $\Omega$ ,

$$W \sim p(\mathbf{S}|W)p(W), \quad W \in \Omega.$$

To preserve the 2D information in  $\mathbf{S}$  or equivalently in  $\mathbf{I}$ , we put the likelihood  $p(\mathbf{S}|W)$  as hard constraints. Then it becomes

$$\begin{aligned} &W \sim p(W) \\ &\text{subject to } \Pi(W) = \mathbf{S}. \end{aligned}$$

where  $\Pi$  represents the projection matrix in the image formation. Orthogonal projection (see Fig.3.a) was assumed in almost all line drawings work. This will cause major artifacts in dealing with some real images. In this paper, we also use the perspective projection shown in Fig. 3.b to deal with the polyhedra scene.

## 2.3. Theoretical connection of our method with Julesz ensemble and photo-hull

The prior model  $p(W)$  should be essentially learned from a large training set of real world shapes and scenes, for example, by a maximum entropy principle [21]. Thus it is a minimally biased summary of real world regularities expressed in statistical constraints over a number of features  $\phi_i(\cdot)$ .

$$E[\phi_i(W)] = \phi_i^{\text{obs}}, \quad i = 1, 2, \dots, M$$

$E[\cdot]$  is the expectation over the prior  $p$ , and  $\phi_i^{\text{obs}}$  is the average over the training ensemble. Thus the problem becomes to sample from the following ensemble

$$W \sim \text{ensemble}\{W : E[\phi_i(W)] = \phi_i^{\text{obs}} \prod_{i=1}^M, \Pi(W) = \mathbf{S}\}.$$

This ensemble is constructed based on both statistical (soft) constraints and hard constraints. It is an integration of the Julesz and Gibbs ensemble in texture modeling [7] and the photo-hull in space carving [16]. Any  $W$  sampled from the ensemble will be a reasonable explanation of the observation image  $I$  or full primal sketch  $S$ .

Intuitively, each object graph is divided into two sub-graphs

$$G_i = G_i^O \cup G_i^H, \quad i = 1, 2, \dots, K.$$

$G_i^O$  includes the observable vertices, edges and surfaces, which have to be consistent with the image after projection. This means that each 3D vertex in  $G_i^O$  slides along a projection line in Fig. 3.b and the photometric attributes in the surfaces have to be interpolated. Some 3D structures will be preferred by the prior  $p(W)$ . This preference is further propagated to compute the hidden graph  $G_i^H$ .

However, it's very hard to learn the prior model  $p(W)$  in practice because we don't have enough 3D data about real world available at current stage of computer vision. Fortunately, some manually defined prior models haven been shown to work well to some extent [20]. So we will follow this way to define our prior models in the next section.

### 3. Prior knowledge on graphs

We represent prior knowledge in two levels. One is on the 3D object graphs and the other is on the relation graph  $\mathcal{PR}$ . Thus the total prior model is as follows,

$$p(W) = p(K) \cdot \prod_{i=1}^K p(G_i) \cdot p(\mathcal{PR})$$

where  $p(K)$  is assumed to be a Poisson distribution. The 3D objects are limited to polyhedron, trees, and grasses. We also assume that the type of each component in the full primal sketch  $S$  is known, e.g. it is the projection of a 3D curve or surface.

#### 3.1. Prior Model for Single Object — p(G)

**3.1.1. Prior model for polyhedra** For each face of any polyhedron, we have two regularities. The first regularity is planarity that the lines for each face should lie on a 3D plane. For each face in the polyhedra,  $f_i, i = 1, 2, \dots, |F|$ , assume it has a number of 3D lines  $l_{ij}, j = 1, 2, \dots, n_i$ . The planarity for all  $f_i$  of the polyhedra is enforced by an energy term,

$$E_1^{face} = \sum_{i=1}^{|F|} \sum_{j=1}^{n_i} \left( 1 - \frac{(l_{i,j-1} \times l_{ij}) \cdot (l_{ij} \times l_{i,j+1})}{\|l_{i,j-1} \times l_{ij}\| \|l_{ij} \times l_{i,j+1}\|} \right)^2$$

where  $\cdot$  and  $\times$  are inner and outer product respectively.

The second regularity is that the inner angles of the face should be more or less the same. This is also the case for the lengths of edges of the face. Let  $\theta_{ij}, j = 1, 2, \dots, n_i$  be the inner angles of face  $f_i$ . The regularity can be enforced by the following two energy terms,

$$E_2^{face} = \sum_{i=1}^{|F|} \sum_{j=1}^{n_i} \frac{1}{n_i} (\theta_{ij} - \bar{\theta}_i)^2, \quad \bar{\theta}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \theta_{ij}$$

$$E_3^{face} = \sum_{i=1}^{|F|} \sum_{j=1}^{n_i} \frac{1}{n_i} (\|l_{ij}\| - \|\bar{l}_i\|)^2, \quad \|\bar{l}_i\| = \frac{1}{n_i} \sum_{j=1}^{n_i} \|l_{ij}\|$$

We also define the prior on all the edges  $E$  using the following three regularities. First, all angles between all pairs of edges meeting at each vertex must be similar. Let  $\theta_{ij}, j = 1, 2, \dots, n_i$  be the angles between all pairs of edges meeting at vertex  $i$  and the regularity can be enforced by the following energy term,

$$E_4^{edge} = \sum_{i=1}^{|V|} \sum_{j=1}^{n_i} \frac{1}{n_i} (\theta_{ij} - \bar{\theta}_i)^2, \quad \bar{\theta}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \theta_{ij}$$

Second, the lengths of all the edges meeting at each vertex must be similar. Let  $e_{ij}, j = 1, 2, \dots, m_i$  be all the edges meeting at vertex  $i$ . This regularity can be enforced by the following energy term,

$$E_5^{edge} = \sum_{i=1}^{|V|} \sum_{j=1}^{m_i} \frac{1}{m_i} (\|e_{ij}\| - \|\bar{e}_i\|)^2, \quad \|\bar{e}_i\| = \frac{1}{m_i} \sum_{j=1}^{m_i} \|e_{ij}\|$$

Third, the lengths of all the edges should be uniformly proportional to those of their 2D projections. Let  $e_i, i = 1, 2, \dots, |E|$  and  $e'_i, i = 1, 2, \dots, |E|$  be the edges in 3D space and their 2D projections respectively. In this paper we assume the projection is either orthogonal or perspective with the projection matrix known by some methods in [22], so we can compute the 2D projection for any  $W$ . Then this regularity can be enforced by the following energy term,

$$E_6^{edge} = \sum_{i=1}^{|E|} \frac{1}{|E|} \left( \frac{\|e_i\|}{\|e'_i\|} - \bar{r} \right)^2, \quad \bar{r} = \frac{1}{|E|} \sum_{i=1}^{|E|} \frac{\|e_i\|}{\|e'_i\|}$$

The prior model for one polyhedron is thus defined as,

$$p(G) \propto \exp\left(-\sum_{i=1}^6 \lambda_i E_i\right)$$

**3.1.2. Prior model for trees and grasses** For trees and grasses, the face surfaces are degenerated into 3D curves. To explicitly represent this changing, we replace  $F$  and  $f$  in  $G$  with  $C$  and  $c$  here. We define the prior model on  $G$  using two regularities. The first one is that each curve in  $C$  should be smooth. To enforce this regularity, we use a

Markov chain model in [23] which forces the smoothness of 2D curves, but extend it to 3D case. Let  $c_i, i = 1, 2, \dots, |C|$  be all the curves in  $C$  and  $v_{ij}, j = 1, 2, \dots, |c_i|$  be all the vertices on curve  $c_i$ . The smoothness prior model for curve  $c_i$  can be represented as,

$$p(c_i) = p(v_{i1}, v_{i2})p(v_{i3}|v_{i1}, v_{i2}) \prod_{j=4}^{n_i} p(v_{ij}|v_{i,j-1}, v_{i,j-2}, v_{i,j-3})$$

The probability  $p(v_{i1}, v_{i2})$  is assumed to be uniform,  $p(v_{i3}|v_{i1}, v_{i2})$  is a two gram represented by a 2-way joint histogram and  $p(v_{ij}|v_{i,j-1}, v_{i,j-2}, v_{i,j-3})$  is a trigram representation by three way joint histogram. The first histogram is learned as in [23], while the second histogram is learned from some manually obtained data by computing three variables:

1. the angle between  $(v_{i,j-1}, v_{i,j-2})$  and  $(v_{i,j-2}, v_{i,j-3})$ ,
2. the angle between  $(v_{i,j-1}, v_{i,j-2})$  and  $(v_{i,j-1}, v_{i,j})$ ,
3. the distance from  $v_{i,j}$  to the plane fitting through  $v_{i,j-1}, v_{i,j-2}$  and  $v_{i,j-3}$ .

The second regularity is that these curves should evenly spread in the 3D space as shown in Figure 3.c. To enforce this regularity, we fit one plane through each of these long curves and force the angles between these planes more or less the same. Let  $\theta_i, i = 1, 2, \dots, N$  be these angles. The regularity can be enforced by the following energy term,

$$E = \sum_{i=1}^N \frac{1}{N} (\theta_i - \bar{\theta})^2, \quad \bar{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i$$

The prior model for grasses is thus defined as,

$$p(G) \propto \prod_{i=1}^{|C|} p(c_i) \exp\{-E\}$$

However, the prior model for trees is more complex since the tree curves are recursively evenly spread in the 3D space. That means not only the overall tree curves are evenly spread in the 3D space, but also are the tree curves of the subtrees. Taking the grass prior model as the prior model for the first level of tree, we can define the tree prior model as:

$$p(G) \propto \prod_{i=1}^L \prod_{j=1}^{M_i} \prod_{k=1}^{|C_{ij}|} p(c_k) \exp\{-E_{ij}\}$$

where  $L$  is number of levels of the tree,  $M_i$  is the number of subtrees at level  $i$ ,  $C_{ij}$  is the number of curves in the  $j$ th subtree at level  $i$ , and  $E_{ij}$  is the regularity energy term for  $j$ th subtree at level  $i$ .

In the computing process, these two families of prior models will compete with each other to explain the 3D shape of one object.

### 3.2. Prior model for scenes – $p(\mathcal{PR})$

For each  $G_i \preceq G_j$  that means  $G_i$  supports  $G_j$ , the interface between them should be as large as possible to be stable. For example, when one box is lying on the ground, we should expect one of its face is totally touching the ground instead of part of it. However, if the interface between  $G_i$  and  $G_j$  is not a surface, but degenerated line or points (e.g. the interface between the wall plane and floor plane or that between grass curves and the pot), we should expect the overall structure of  $G_j$  is perpendicular to  $G_i$ . Since we know each component in  $S$  represents a 3D curve or surface, the type of the interface between  $G_i$  and  $G_j$  can be classified deterministically. Thus, this regularity is enforced by the following energy term,

$$E(G_i \preceq G_j) = \begin{cases} \frac{\sum_{m=1}^{|V_j|} (1 - \delta(D(v_{jm}, G_i) < T))}{\|A(G_i, G_j) - \frac{\pi}{2}\|} & G_i \cap G_j \text{ is a surface} \\ \text{otherwise} & \end{cases}$$

where  $D(v_{jm}, G_i)$  is a function to compute the minimum distance between vertex  $v_{jm}$  and object  $G_i$ , and  $A(G_i, G_j)$  is a function to compute the angle between the two overall structures of  $G_i$  and  $G_j$ .  $T$  is a threshold to judge the distance is small enough for two objects touching each other.

The prior model for spatial relation  $\mathcal{PR}$  is thus defined as,

$$p(\mathcal{PR}) \propto \exp - \left\{ \sum_{m=1}^{|\mathcal{PR}|} E(G_{\mathcal{PR}(m).1} \preceq G_{\mathcal{PR}(m).2}) \right\}$$

where  $\mathcal{PR}(m).1$  and  $\mathcal{PR}(m).2$  denote the first and second element in the  $m$ th set of the poset  $\mathcal{PR}$  respectively.

## 4. Computation and inference

### 4.1. Random walk in manifold of different dimensions

Given the two levels of prior models and our interest in obtaining nearly globally optimal solutions, we design Markov chain to simulate walks in the solution space  $\Omega$ . Recently there are some ideas to speed up the MCMC search by data driven techniques for image segmentation and perceptual grouping [2, 23] which compute importance proposal probabilities to drive the Markov chain and make the search focus on promising subspaces. We adopt the DDM-CMC paradigm in [2]. However, only data driven may not be enough to recover the hidden information since the given data provides no direct information about the invisible parts. Here we introduce some geometric rules that human beings often use to imagine the hidden information to speed up the recovering of hidden information.

## 4.2. Diffusion of vertices and objects

When each subgraph in  $W$  is fixed for both visible and inferred hidden vertices and edges, we try to diffuse the  $z$  value for each visible vertex and the  $x, y, z$  values for each inferred hidden vertex. We also have another diffusion process to adapt the overall position of each object to make all the objects in a scene have maximum stabilities, e.g. the wall and the floor would be moved to be perpendicular to each other. The diffusion dynamics can be easily implemented by Gibbs sampler.

## 4.3. Reversible jumps

Structure changes in the solution  $W$  are realized by Markov chain jumps, which are bridges between subspaces of different dimensions in  $\Omega$  and made reversible [24] to avoid greedy search. To obtain the full 3D information of the scene, we design the following reversible jumps that form ergodic Markov chain to search in the space  $\Omega$ .

1. Reversible jumps to switch between the two families of prior models for single object,

$$(G; \ell_1) \leftrightarrow (G; \ell_2)$$

where  $\ell_i, i \in \{1, 2\}$  is the label for these families of prior models.

2. Reversible jumps for adding/removing a partial order  $\{G_i, G_j\}$  element in the spatial relation set  $\mathcal{PR}$ ,

$$PR \leftrightarrow PR + / - \{(G_i, G_j)\}$$

Or reversing the order,

$$(G_i, G_j) \leftrightarrow (G_j, G_i)$$

3. Reversible jumps for adding/removing a vertex in the inferred hidden part  $G^H$

$$G^H \leftrightarrow G^H + / - v$$

4. Reversible jumps for adding/removing an edge in the inferred hidden part  $G^H$

$$G^H \leftrightarrow G^H + / - e$$

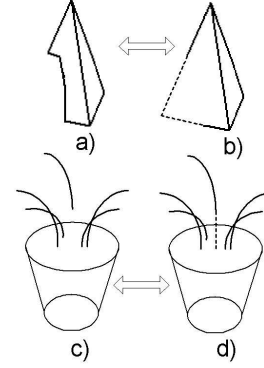
5. Reversible jumps for splitting/merging surfaces in the inferred hidden part  $G^H$

$$f_k \leftrightarrow f_i \cup f_j$$

6. Reversible jumps for splitting/merging subgraphs

$$G_k \leftrightarrow G_i \cup G_j$$

7. Reversible jumps for growth/shrink of the boundary of occluded face (or curve for tree and grass). Figures 4 shows how this pair of jumps work, where the occluded boundary and curve shown in a) and c) are grown to the state as shown in b) and e) respectively.



**Figure 4. a) The state of a polyhedra before growth of its occluded boundary, b) the state of the polyhedra in b) after growth of its occluded boundary, c) the state of a grass before growth of its occluded curve, d) the state of the grass in c) after growth of the occluded curve.**

## 4.4. Data driven and knowledge driven

Each reversible jump connects two states  $W_A$  and  $W_B$ , and observes the detailed balance equation

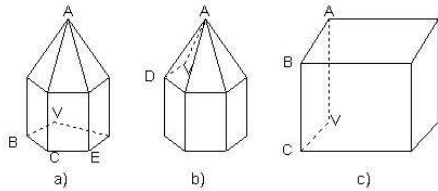
$$p(W_A|I)dW_AP(W_A \rightarrow dW_B) = p(W_B|I)dW_BP(W_B \rightarrow dW_A)$$

where  $p(W_A|I), p(W_B|I)$  are the posterior probabilities and for  $W_A \neq W_B$

$$P(W_A \rightarrow dW_B) = q(W_A \rightarrow dW_B)\alpha(W_A \rightarrow W_B)$$

is the transition (conditional) probability from  $W_A$  to  $W_B$ , and  $q()$  and  $\alpha()$  are respectively the proposal and acceptable probabilities. A good design of the proposal probability is crucial to speed up the search and two recent successful examples are [2, 23] which design the proposal probability from bottom-up heuristics  $B(I): q(A|B, D(I)) \approx p(A|I)$  and  $q(B|A, D(I)) \approx p(B|I)$ . So the acceptance probability is close to one. To speed up the splitting and merging jumps on subgraphs using this idea, we compute *importance proposal*,  $q$ , from the relation between the shapes of the component that always should be grouped together. This is especially true for the tree and grass since we should never group a curve with a face, but instead one curve with another curve.

However, how to design a good  $q$  for the dynamics recovering the hidden structure is challenging because we have no directly observed data to use. Some geometric rules that human being often use to imagine the hidden information are introduced here to produce a good  $q$  and act as Knowledge Driven. These geometric rules are: parallelism,



**Figure 5. Three rules are used to propose a new vertex  $V$  for the hidden structure  $G^H$ . a) mirror symmetry, b) rotational symmetry, c) parallelism.**

rotational symmetry and mirror symmetry, which are often present in the 3D objects. Fig. 5 shows how these rules can be used to propose a new vertex added to  $G^H$ .

## 5. Experimental results

We test our whole algorithm on two types of data sets. The first data set are images about polyhedra and we use image segmentation [2] to work on these images to get the sketch. Besides the vertices and edges information of sketch, the segmentation result also provides the face information that means which vertices and edges are combined to form a face. The second data set are images about objects with strong stochastic 3D shapes, like trees and grasses. We use primal sketch algorithms [3, 6] to work on these images to get the sketch. As the segmentation result can provide the face information, the primal sketch algorithm also provides the curve information about which edges are connected to form a long curve. Therefore, for each component in  $S$ , we can know it should represent a 3D curve or surface. In addition, to show the reconstruction results effectively in OpenGL with texture mapping, we intently extend the curves into narrow surfaces.

Fig. 6 shows some reconstruction results on the polyhedra data set. The first two rows in the figure show three images of polyhedra scene and their sketches obtained by the algorithm in [2]. The rest three rows show the reconstruction results in three different viewpoints.

Fig. 7 shows some reconstruction results on trees and grasses in the second data set. The first row shows three input images of tree or grass. The sketches of these images, obtained by the algorithm in [3] for the tree image and by the algorithm in [6] for grass images, are shown in the second row. The reconstruction results are also shown in three different viewpoints in the last three rows.

## 6. Discussion

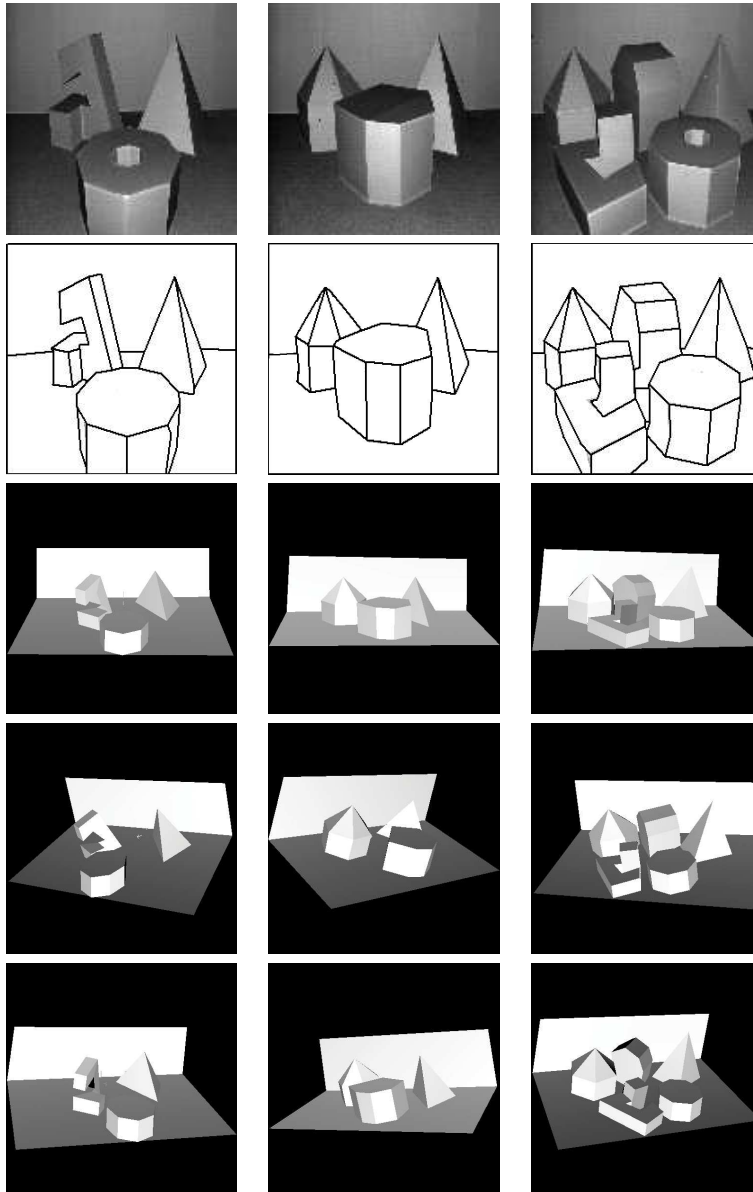
Our algorithm is built on a full primal sketch from other segmentation and sketching algorithm using MCMC [2, 3, 6]. Currently we are integrating the two stages as they are all formulated in the Bayesian framework with MCMC computing. Thus we solve  $W$  from  $I$  directly with  $S$  being an intermediate representation.

$$(W, S) \sim p(I|S)p(S|W)p(W)$$

Thus we can use 3D priors to resolve ambiguities in segmentation and sketching.

## References

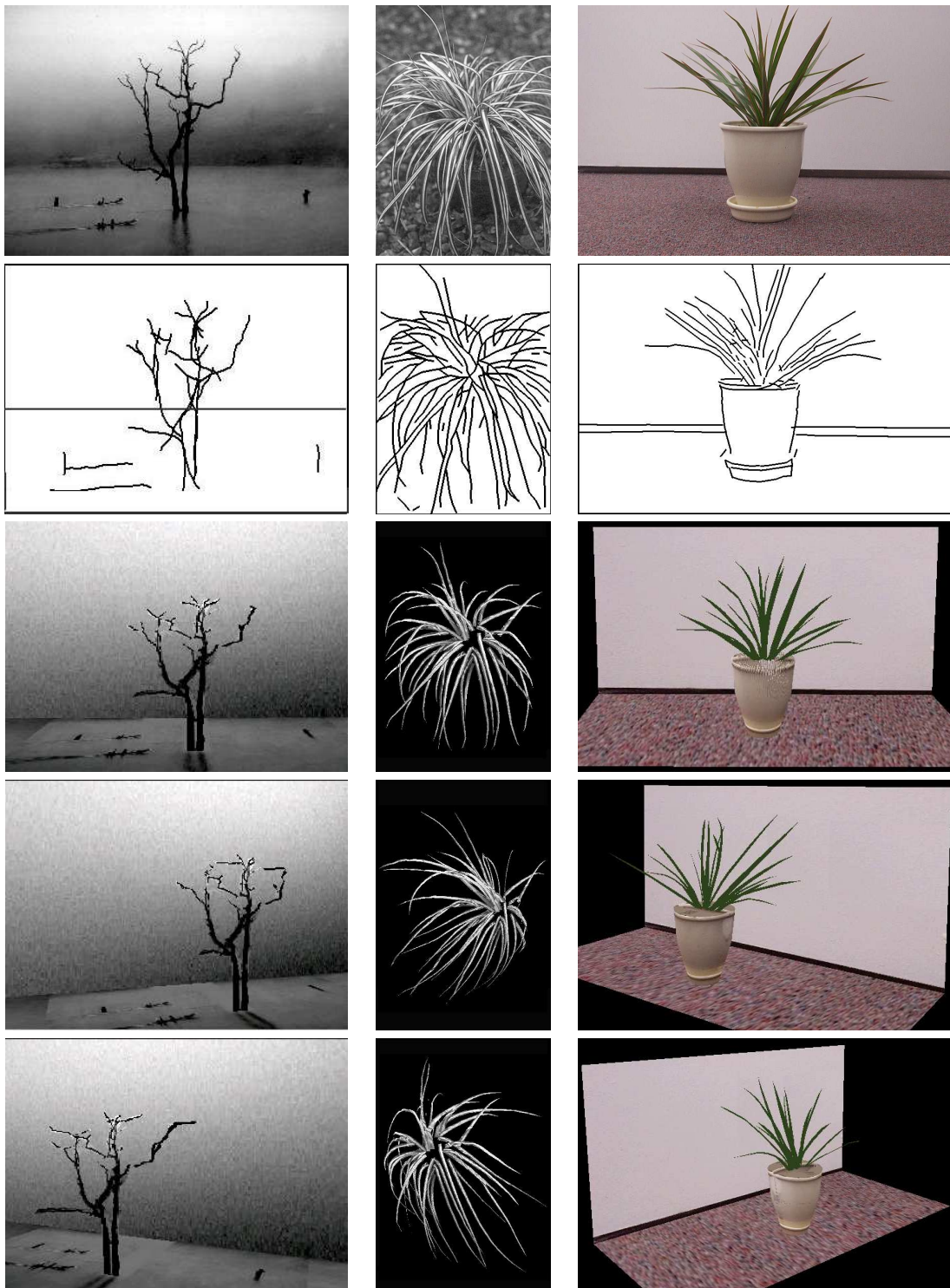
- [1] P. A. C. Varley and R. R. Martin, "A system for constructing boundary representation solid models from a two-dimensional sketch", *Proc. GMP*, 2000.
- [2] Z. W. Tu and S. C. Zhu, "Image segmentation by data driven Markov chain Monte Carlo." *IEEE TPAMI*, 24(5), 2002.
- [3] Z. W. Tu and S. C. Zhu, "Parsing images into region and curve processes." *ECCV*, 2002.
- [4] H. Lipson and M. Shpitalni, "Optimization-based reconstruction of a 3D object from a single freehand line drawing", *Computer-Aided Design*, Vol. 28, No. 8, pp. 651-663, 1996
- [5] A. Yezzi and S. Soatto, "Stereoscopic Segmentation", *IJCV*, pp. 59-66, 2001.
- [6] Cheng-en Guo, Song-Chun Zhu, and Ying Nian Wu, "A mathematical theory of primal sketch", *ICCV*, 2003
- [7] S. C. Zhu, X. W. Liu and Y. N. Wu, "Exploring Texture Ensembles by Efficient Markov Chain Monte Carlo", *IEEE TPAMI*, 22(6), 2000.
- [8] K. Sugihara, "Mathematical structure of line drawings of polyhedrons - towards man-machine communication by means of line drawings", *IEEE TPAMI*, 04(5), 1982.
- [9] T. Marill, "Emulating the human interpretation of line drawings as three-dimensional objects", *IJCV*, 06(2), 1991.
- [10] E. W. Brown and P. Wang, "Three-dimensional object recovery from two-dimensional images: a new approach", *SPIE*, vol. 2904, pp. 138-147, 1996.
- [11] Y. G. Leclerc and M. A. Fischler, "An optimization-based approach to the interpretation of single line drawings as 3d wire frames", *IJCV*, 09(2), 1992.
- [12] B. K. P. Horn, "Height and gradient from shading", *IJCV*, 05(1), 1990.
- [13] J. D. Foley, A. van Dam, S. K. Feiner and J. F. Hughes, "Computer Graphics: Principles and Practice", Addison-Wesley Publishing Company, Inc., pp. 72-91, 1990.
- [14] S. K. Nayar and Y. Nakagawa, "Shape from focus", *IEEE TPAMI*, 16(8), 1994.
- [15] A. Criminisi, I. Reid and A. Zisserman, "Single view metrology", *IJCV*, 40(2), 2000.
- [16] N. K. Kutulakos and S. M. Seitz, "A theory of shape by space carving", *IJCV*, 38:199-218, 2000.



**Figure 6. Three input images and their sketches are shown in the first two rows and the reconstruction results are shown in three different viewpoints in the rest three rows.**

- [17] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun, "Structure from Motion without Correspondence", *CVPR*, 2000.
- [18] M. Pollefeys, R. Koch, and L. Van Gool. "Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters", *Proc. ICCV*, pp. 9096, 1998.
- [19] L. Zhang, G. Dugas-Phocionz, J-S. Samsonz, and S. M. Seitz "Single View Modeling of Free-Form Scenes", *CVPR*, 2001.
- [20] A. R. Dick, P. H. S. Torr and R. Cipolla, "A bayesian estimation of building shape using MCMC", *ECCV*, 2002.
- [21] S. C. Zhu, Y. N. Wu and D.B. Mumford, "Minimax Entropy Principle and Its Applications to Texture Modeling", *Neural Computation* Vol. 9, no 8, pp 1627-1660, Nov. 1997.
- [22] D. Jelinek and C. J. Taylor "Reconstruction of linearly parameterized models from single images with a camera of unknown focal length", *IEEE TPAMI*, 23(7), 2001.
- [23] A. Barbu and S.C. Zhu, "Graph partition by Swendsen-Wang Cuts", *ICCV*, 2003.
- [24] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, vol.82, 711-732, 1995.





**Figure 7. Three input images are shown in the first row and their sketches are shown in the second row. The reconstruction results are shown in three different viewpoints in the rest three rows.**