Published in IET Computer Vision
Received on 13th December 2006
Revised on 26th October 2007
doi: 10.1049/iet-cvi:20065002

ISSN 1751-9632

3D structure inference by integrating segmentation and reconstruction from a single image

L. Lin[1] *K. Zeng*[2] *Y. Wang*[1] *W. Hu*[1]

[1]*School of Information Science and Technology, Beijing Institute of Technology, Beijing 100081, People's Republic of China*
[2]*Institute of Automation, Chinese Academic of Sciences, Beijing 100080, People's Republic of China*
E-mail: linliang@bit.edu.cn

Abstract: The authors present a hierarchical Bayesian method for inferring the 3D structure of polyhedral man-made objects from a single image by integrating 2D image parsing and 3D reconstruction. In the first stage, the image is parsed into its constituent components – arbitrary shape regions and polygonal shape regions. In the second stage, polygonal shape regions are grouped into man-made polyhedral objects. The 3D structures of these polyhedral objects are further inferred using geometric priors. These two stages are integrated into a Bayesian inference scheme and cooperate to compute the optimal solutions. This method enables the model to correct possible errors and explain ambiguities in the lower level with the help of information from the higher level. The algorithm is applied to the images of indoor scenes, and the experimental results demonstrate satisfactory performance.

1 Introduction

Inferring 3D structures of objects from a single 2D intensity image is a very difficult problem and has been a heavily researched topic in computer vision for many years. There are two major difficulties in this process. First, the intensity value of a pixel in a 2D image does not provide enough information by itself to determine which object it belongs to. To extract an object, we must know how to group related pixels into regions and then into objects by using information such as smoothness, existence of discontinuities and some high-level knowledge about the scene and the objects involved. This segmentation problem is the essential issue for most computer vision systems. Second, no direct 3D shape information can be obtained from a single intensity image. The process of projecting the world into 2D discards information along the third dimension. Therefore we must develop methods to recover the lost 3D information from the remaining 2D information.

Because of these difficulties, many researchers addressing this topic confine themselves to solving a simpler version of the problem. There are three main streams of research in the literature:

Assuming the segmentation has been done and the result is more or less perfect [1–4]: Although these methods differ from each other in how to recover lost information from the remaining 2D information (either using specific object models or generic prior models), such as the classic 'blocks world' [5], they assume the segmentation result is a reliable separation of foreground objects from the background and from each other. However, this assumption may not be true in many cases.

Only using the intensity edges as input features for perceptual grouping [6–9]: These methods extract and use some intensity features (like edges) instead of the raw image data. These features are then sequentially grouped into more and more meaningful parts. However, this data reduction process does not preserve all the information and is prone to errors. In addition, these methods proceed sequentially, so errors in the lower levels may propagate to higher levels without any rectifications.

Introducing human interaction [10–13]: These methods ask the user to provide high-level information to help feature extraction and grouping so that robust computer vision algorithms can concentrate on performing the geometric computations automatically.

It is clear that most of the existing methods either avoid the segmentation problem altogether, like the first and third streams, or perform sequential computing using reduced data. Even as we begin to see more and more robust segmentation algorithms, we should not expect any of them to provide almost perfect results for all cases. In fact, segmentation should not be separated from high-level perception as human vision likely weighs both low-level raw image data and different high-level representations iteratively to explain possible ambiguities in the scene. Besides, it is worthy mentioning a few recent work, approaching some achievements in single-image-based 3D inference. One method using multiple segmentations to extract the structure of the outdoor scene and performing 3D reconstruction from a single image was presented by Hoiem *et al.* [14], as well as its extensive studies [15]. Han and Zhu [16] proposed a grammar-based inference framework to parse a single scene into consistent patterns and the 3D information can be inferred based on 3D-attributed grammar rules.

Being different from those methods mentioned above, in this paper, we present a two-level computation model to extract 3D structures of objects from a single 2D image, as illustrated in Fig. 1. In the first level, the image is parsed into its constituent components – arbitrary shape regions and polygonal shape regions. In the second level, these adjacent polygonal shape regions are partitioned into many groups, where each group corresponds to the visible exterior of a man-made polyhedral object. The occluded 3D structure of each proposed polyhedral object is then inferred. The whole computation proceeds iteratively instead of sequentially, integrating these two levels in a Bayesian framework using data-driven Markov Chain Monte Carlo. In this way, both bottom-up and top-down information cooperate to compute the optimal solution.

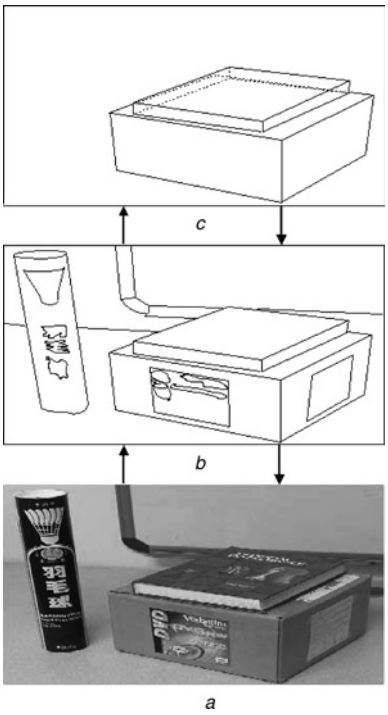

Figure 1 *Two-level computation model to extract 3D structures*

a Input image
b Parsing the input image into its constituent regions
c Grouping regions into polyhedral objects and inferring their 3D structures The arrows show that the whole computation is iterative instead of sequential

2 Problem formulation

In this section, we present generative models for arbitrary shape regions, polygonal shape regions and 3D polyhedral objects. We also formulate the hierarchial inference in a Bayesian framework. Fig. 2 illustrates the two types of regions and the polyhedral objects we are working on.

2.1 Image representation with occlusion model

Each arbitrary shape region R occupies an area $D(R)$ bound by its boundary contour, whereas each polygonal shape region S occupies an area $D(S)$ bound by the convex polygon it represents. We denote an image lattice by

$$\Lambda = \{(i, j): 0 \leq i \leq M, 0 \leq j \leq N\}$$

Then each arbitrary shape region occupies all the pixels lying inside $D(R)$, which are denoted by Λ_R. However, this is not

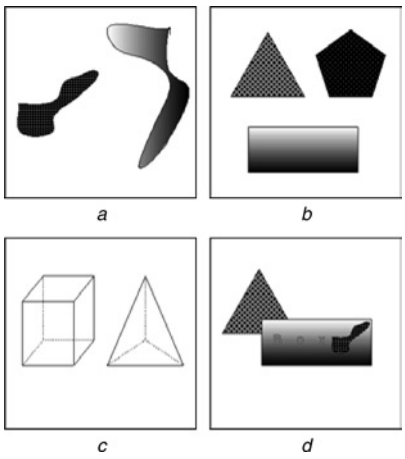

Figure 2 *Two types of regions and the polyhedral objects*

a Arbitrary shape regions
b Polygonal shape regions
c Polyhedral objects
d Possible occlusions

IET Comput. Vis., 2008, Vol. 2, No. 1 pp. 15–22
doi: 10.1049/iet-cvi:20065002

true for polygonal shape regions since they are often occluded either by some arbitrary shape regions produced by appearance patterns on top of them (like texts, textures) or other neighbouring polygonal shape regions, as shown in Fig. 2*d*. Putting all the regions with occlusion relations into a set $\mathcal{A}$, we define a partial order binary occlusion relation $\preceq$ in this set, for example $a \preceq b$ means a occluding b. The total occlusion relation is denoted by $\mathcal{PR} = \langle \mathcal{A}, \preceq \rangle$. Let Λ_S denote the pixels occupied by a polygonal shape S in the 2D discrete lattice Λ of image I. Then, Λ_S are these pixels inside $D(S)$ minus the pixels covered by the occluding regions, which are some polygonal shape region S' or arbitrary shape region R.

$$\Lambda_S = D(S) - \cup_{S' \preceq S} \Lambda_{S'} - \cup_{R \preceq S} \Lambda_R$$

2.2 Generative models of arbitrary shape regions

To model the arbitrary shape regions, we use the same image model and boundary prior model as in [16].

1. Family I_1: This is a uniform region with constant reflectance $\Theta = \mu$.

2. Family I_2: This is a cluttered region with a non-parametric histogram $\Theta = (h_1, h_2, \ldots, h_L)$ modelling its intensity with L being the number of bins.

3. Family I_3: This is a region with smooth variation of reflectance, modelled by a B-spline model as in family D_3.

Assuming the residuals from the fit to be modelled as Gaussian noise, the likelihood probability is

$$p(I_R|R, \theta) \propto \prod_{(i,j)\in\Lambda_R} e^{-(I(i,j)-J(i,j;\ell,\theta))}, \ell \in \{1, 2, 3\}$$

where $J(i, j; \ell, \theta)$ is the fitting result according to model (ℓ, θ).

The prior model for a region R assumes short boundary length ∂R (smoothness) and compact area $|D(R)|$,

$$p(R) \propto \exp\left\{-\gamma_r |D(R)|^{\alpha} - \frac{1}{2}\lambda|\partial R|\right\}$$

where α and λ are constants and γ_r is a scale factor for regions and may vary slightly between different datasets.

2.3 Generative models of polygonal shape regions

To model the polygonal shape regions, we use the same image model as for arbitrary shape regions, but replace the boundary smoothness prior model with a polygonal model [17] defined by a set of vertex points with spatial coordinates $\overrightarrow{\chi}_r$ and edges to form a contour. The vertices are modelled by a Poisson process with the probability

$$p(\overrightarrow{\chi}_r) = (\beta\mathcal{A})^n e^{-\beta\mathcal{A}}/n!, \ 1 \le r \le n$$

where n is the number of vertices, $\mathcal{A}$ is the area of the region and β is the rate at which vertices appear in the image.

This polygonal model puts constraints not on the smoothness, but on the complexity of the region boundaries, which encourages polygonal shape regions since the best region boundary is the one with the fewest points that still approximates the data as well as other models.

2.4 Generative models of polyhedral objects

Some polygonal shape regions can be grouped into an aspect graph for one polyhedral object, whose 3D structure can be inferred in a later step. We denote such a group by

$$P = (n, S_1, \ldots, S_n)$$

where n is the number of polygonal shape regions in this group and S_n represents each individual polygonal shape region in the group. The 3D structure of the polyhedral object corresponds to P and is denoted by a 3D graph $G = (V, E, F)$, which is represented by (1) V – 3D vertices, (2) E – 3D edges and (3) F – surfaces with intensity patterns. Visible surfaces inherit their intensity patterns from their projection regions in P. We set the intensity patterns for hidden surfaces to infinity.

To compensate the lost 3D information and infer 3D structure in a single image, some prior knowledge of polygonal structure is defined manually. Note we perform 3D perceptual inference but not exact geometrical 3D reconstruction in this work.

The prior model for G is based on some prior geometric regularities, defined on its faces and edges. For each face in G, we have two constraints. The first constraint is planarity, that is the edges of each face should lie on a 3D plane. For each face in the polyhedron, f_i, $i = 1, 2, \ldots, |F|$, assume it has a number of 3D lines $l_{ij}, j = 1, 2, \ldots, n_i$. The planarity for all f_i of the polyhedra is enforced by an energy term

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

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. The lengths of the edges of the face should be roughly the same as well. Let θ_{ij}, $j = 1, 2, \ldots n_i$ be the inner angles of face f_i. The regularity

can be enforced by the following two energy terms

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

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

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 $\theta_{ij}, j = 1, 2, \ldots, n_i$ be the angles between all pairs of edges meeting at vertex i. This constraint can be enforced by the following energy term

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

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

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

Third, the lengths of all the edges should be uniformly proportional to those of their 2D projections. Let $e_i, i = 1, 2, \ldots, |E|$ and $e_i', i = 1, 2, \ldots, |E|$ 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 using methods in [10], so we can compute the 2D projection for any W. Then this regularity can be enforced by the following energy term

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

The prior model for one polyhedron is thus defined as

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

Putting the hard constraint that the projection of the polyhedral object is the same as the boundaries of all the regular regions in P, G inherits the likelihood from the regions it covers.

2.5 Bayesian formulation

Given an image I, our objective is to parse it into its constituent components – polygonal shape regions W^s and arbitrary shape regions W^r – with partial order $\mathcal{PR}$, group some polygonal shape regions into objects and infer the 3D structure of these objects W^g.

Thus, a solution W is denoted by

$$W = (W^r, W^s, W^g, \mathcal{PR})$$

The region representation W^r includes the number of regions K^r, and each region R_i has a label $\ell_i \in 1, 2, 3$ and parameter θ_i for its intensity model

$$W^r = (K^r, \{(R_i, \ell_i, \theta_i): i = 1, 2, \ldots, K^r\})$$

Similarly, we have $W^s = (K^s, \{(S_i, \ell_i, \theta_i): i = 1,2, \ldots, K^s\})$, $W^g = (K^g, G_1, G_2, \ldots, G_{K^g})$.

The problem is posed as Bayesian inference in a solution space Ω.

$$W* = \arg\max_{\Omega \ni W} p(I|W)p(W)$$

Except K^r and K^s combining to represent the total region numbers of image I, we assume mutual independence between W^r, W^s and W^g. Thus we have the prior model,

$$p(W) = p(K^r + K^s) \times \prod_{i=1}^{K^r} p(R_i) \prod_{i=1}^{K^s} p(S_i) \left(p(K^g) \prod_{i=1}^{K^g} p(G_i) \right)$$

The priors for individual $p(R)$, $p(S)$ and $p(G)$ are given in the previous subsections, whereas $p(K^r + K^s)$ and $p(K^g)$ are assumed to follow a Poisson distribution.

The likelihood model follows the lattice partition,

$$p(I|W) = \prod_{k=1}^{K^r} \prod_{(i,j)\in\Lambda_{R_k}} e^{-(I(i,j)-J(i,j;\ell,\theta))} * \prod_{l=1}^{K^s} \prod_{(i,j)\in\Lambda_{S_l}} e^{-(I(i,j)-J(i,j;\ell,\theta))} \quad \ell \in \{1, 2, 3\}$$

3 Searching complex solution space by Markov chain

Based on the previous formulation, we see that the solution space $\Omega \ni W$ contains many subspaces of varying dimensions. To compute a globally optimal solution, we design a two-level searching algorithm to get an optimal solution on complex solution space using data-driven Markov chain [18, 19]. An overview of the whole algorithm is shown in Fig. 3.

3.1 Dynamics in the first level

In this level, we parse the image into polygonal shape regions and arbitrary shape regions. Five types of MCMC dynamics used in [20] are adopted to handle arbitrary shape regions: diffusion of region boundary, splitting of a region into two, merging two regions into one, switching the family of

doi: 10.1049/iet-cvi:20065002

In the first level, we not only parse the image into polygonal shape regions and arbitrary shape regions, but also obtain the occlusions of polygonal shape regions. Three types of MCMC dynamics are used,

- Diffuse the boundary of polygonal shape regions;
- Switch between arbitrary shape region and polygonal shape region;
- Reversible jumps for adding/removing a partial order;

In the second level, a two-stage algorithm is used to obtain the polyhedral objects and 3D structure,

- In the first-stage, the polygonal shape regions are grouped into polyhedral objects using the Swendsen-Wang cut algorithm,
- In the second-stage, we infer the hidden structure caused by self-occlusion to get the 3D structure. Three types of MCMC dynamics are used,
  - Split an existing face in $G_{HiddenParts}$ into two faces;
  - Merge faces in $G_{HiddenParts}$ into one;
  - Diffuse z values for both visible and inferred hidden vertices;

Figure 3 *Overview of the two-level searching algorithm*

models and model adaptation for a region. These dynamics also work as in [20], where the jumps are sped up by some bottom-up proposals.

To handle polygonal shape regions and their occlusions, we use the following three dynamics:

1. Diffuse the boundary of polygonal shape regions. To diffuse the boundary of a polygonal shape region, we move each edge segment along the boundary rigidly. This is different from diffusing each node on the boundary freely for arbitrary shape regions. To do this, we discretise each edge segment into many nodes. Then we compute the movement for each of these nodes as for an arbitrary shape region. Finally, all these movements are propagated to the two ending vertices of this edge segment. One example is illustrated in Fig. 4.

2. Switch between arbitrary shape region and polygonal shape region: At each iteration of the computation, an arbitrary shape region can become a polygonal shape region with some occluding arbitrary shape regions on the top of it. Similarly, a polygonal shape region may turn out to be a worse explanation of the image than its alternatives, and can release its boundaries to become an arbitrary shape region. This dynamic is illustrated in Fig. 5.

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

$$PR \Longleftrightarrow PR + / - \{(a, b)\}$$

Or reversing the order

$$(a, b) \Longleftrightarrow (b, a)$$

When an arbitrary shape region is switched into a polyhedral shape region, we have to add the occlusion relations it produces. Oppositely, if one polyhedral shape region is switched into one arbitrary shape region, we have to delete the occlusion relations connected with it. We can also switch the direction of an occlusion relation at each step.

3.2 Dynamics in the second level

3.2.1 Grouping polygonal shape regions into polyhedral objects: This next dynamic is very important because it acts as the connection between the

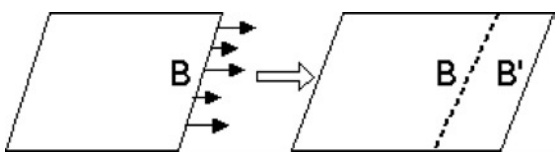

Figure 4 *One part of the boundary of a polygonal shape region, B, is diffused to a new one B′*

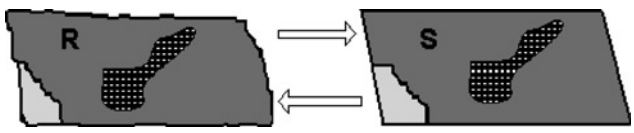

Figure 5 *The arbitrary shape region R can switch into a polygonal shape region S and vice versa*

two representational levels. On one hand, it should be reversible to make the whole computation model reversible. On the other hand, it should efficiently communicate data between our two computational levels. We adopt the Swendsen–Wang cut algorithm [21], which can efficiently perform perceptual grouping based on some bottom-up heuristics. The bottom-up heuristics used here are computed by line labelling, which tries to label each boundary of each polygonal shape region as convex boundary, concave boundary or occluding boundary. These labellings can be used to propose how likely it is that two neighbouring regions are grouped into one object based on the fact that two regions sharing a convex boundary have a higher probability of being on the same object, whereas two regions sharing an occluding boundary have a higher probability of being from two different objects.

3.2.2 *Infer the 3D structure of polyhedral objects:*

After grouping polygonal shape regions into individual polyhedral objects, we try to infer the 3D structure of each object for both the visible and hidden parts. Even though the hidden parts of each surface created by occlusions between neighbouring objects have been recovered by extracting all the polygonal shape regions in the first level, we have to infer the hidden structure caused by self-occlusion.

To see what we should do in this computation process, we separate the 3D graph G for each object into visible parts $G_{\text{VisibleParts}}$ and hidden parts $G_{\text{HiddenParts}}$, and start with $G_{\text{HiddenParts}}$ being set to a single face, making the object spatially closed. Assuming orthographic projection, we only compute the z values for vertices in $G_{\text{VisibleParts}}$ to infer the 3D structure of visible parts. But for $G_{\text{HiddenParts}}$, we should do one more thing: infer the topology of $G_{\text{HiddenParts}}$, which determines the number of vertices, edges, faces for the hidden structure and how they are organised. These two tasks can be accomplished by jump and diffusion dynamics, respectively.

To shed further light on how the jump dynamics work on the topology of $G_{\text{HiddenParts}}$, we show two examples in Fig. 6. From Fig. 6 we can see that, to obtain the correct reconstruction results for the two objects in the figure, the topology of $G_{\text{HiddenParts}}$ should be changed from the left interpretation to the right one. It is clear now that to achieve the goal of changing the topology of $G_{\text{HiddenParts}}$, the jump dynamics can proceed by either splitting an existing face into two or merging two faces into one. Thus, the following dynamics are used:

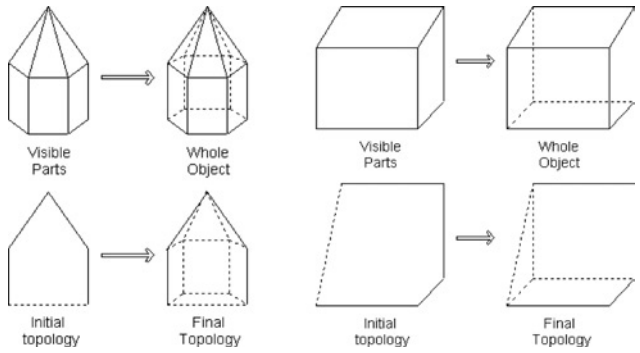

Figure 6 *Two examples are used to show how the topologies should change to achieve the correct result*

1. Split an existing face in $G_{\text{HiddenParts}}$ into two faces by either connecting two existing vertices in $G_{\text{HiddenParts}}$ or inserting a new vertex and connecting it to other two existing vertices. At state W_A, the topology of the hidden structure has one face and three candidates are listed for splitting this face. One of the candidates is proposed probabilistically and we obtain state W_B. At W_B one candidate can be used for merging. Fig. 7 shows an example of this jump between two states W_A and W_B

$$W_A = (W, f_{ik}, V_i, E_i) \rightleftharpoons (W, f_{im}, f_{in}, V_i', E_i') = W_B$$

2. Merge two faces in $G_{\text{HiddenParts}}$ into one.

3. Diffuse z values for both visible and inferred hidden vertices

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

$$p(W_A|I)\,\mathrm{d}W_A\,P(W_A \to \mathrm{d}W_B) = p(W_B|I)\,\mathrm{d}W_B \times P(W_B \to \mathrm{d}W_A)$$

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

$$P(W_A \to \mathrm{d}W_B) = q(W_A \to \mathrm{d}W_B)\,\alpha(W_A \to W_B)$$

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 of proposal probabilities come from [16, 18], which design the proposal probability from bottom-up heuristics $B(I)$: $q(A|B, D(I)) \simeq p(A|I)$ and $q(B|A, D(I)) \simeq p(B|I)$. This produces an acceptance probability close to one.

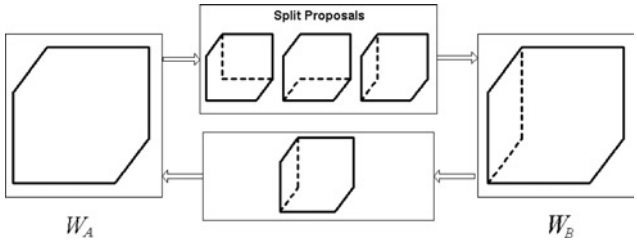

Figure 7 *Example of the split-merge jumps used to infer the hidden structures of polyhedral objects*

doi: 10.1049/iet-cvi:20065002

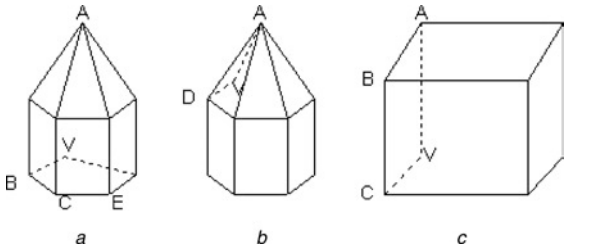

Figure 8 *Three rules are used to propose a new vertex V*

a Mirror symmetry
b Rotational symmetry
c Parallelism

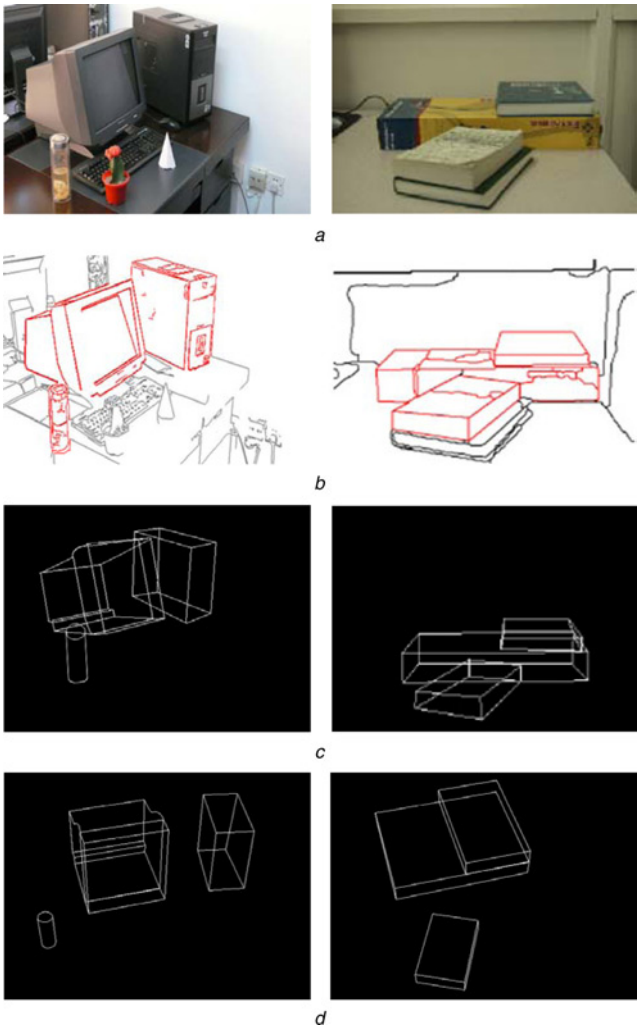

Figure 9 *3D inference on indoor real images*

a Two input images
b Two corresponding segmentation results, in which polygonal shape regions and their occluding regions are displayed in red
c 3D structure of polyhedral objects
d Objects with different viewing position

However, designing a good q for the dynamics that recover the hidden structure is challenging because we have no directly observed data to use. Some geometric rules that human beings often use to represent hidden information are introduced here to produce a good q and act as high-level knowledge. These geometric rules are parallelism, rotational symmetry and mirror symmetry, which are often present in 3D objects. Fig. 8 shows how these rules can be used to propose a new vertex added to G^H.

4 Experimental results

We tested the two-level hierarchy computation model on two real indoor images, as shown in Fig. 9*a*. To highlight the computed polygonal shape regions in the segmentation result, the boundaries of these polygonal shape regions and the regions occluding them are displayed in red (Fig. 9*b*). The final 3D structures inferred for the polyhedral objects in the two examples are shown in wireframe format in OpenGL, as shown in Figs 9*c* and *d* with different view position. Although our main goal is to extract the 3D structures of polyhedral objects in the image, it is clear that the segmentation results are greatly improved at the same time, since the 3D perception can help explain many ambiguities in the segmentation.

5 Summary and future work

In this paper we propose a two-level hierarchical Bayesian computation model for extracting 3D structure from a single image. This hierarchical model is reversible and thus can correct possible errors at lower levels to achieve a global optimum. Currently, we infer the 3D structure of objects individually without considering the spatial relation between them. In reality, objects are organised in a scene with strong spatial regularities, which provide more constraints to recover the lost 3D information. We will add this to the current computational model in the future.

6 Acknowledgments

This work is done at the Lotus Hill Institute and is supported by National Basic Research Program of China (Grant No. 2006AA01Z339 and No. 2006AA01Z121) and National Natural Science Foundation of China (Grant No. 60673198). The authors would like to thank Dr Feng Han for previous work, and Dr Song-Chun Zhu for extensive discussion.

7 References

[1] BROOKS R.A.: 'Intelligence without representation'. 'Preprints of the Workshop in Foundations of Artificial Intelligence', 1987

[2] DICKINSON S.J., PENTLAND A.P., ROSENFELD A.: '3D shape recovery using distributed aspect matching', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1992, **14**, (2), pp. 174–198

[3] DICKINSON S.J., METAXAS D.: 'Integrating qualitative and quantitative shape recovery', *Int. J. Comput. Vis.*, 1994, **13**, (3), pp. 1–20

[4] MUNDY J.L., ZISSERMAN A., FORSYTH D.A., (EDS.): 'Applications of invariance in computer vision', *Lect. Notes Comput. Sci.*, 1994, **825**

[5] ROBERTS L.: 'Machine perception of three-dimensional solids' in Tippett J. (Ed.): 'Optical and electrooptical information processing' (MIT Press, Cambridge, 1965), pp. 157–197

[6] ZERROUG M., NEVATIA R.: 'Three-dimensional descriptions based on the analysis of the invariant and quasi-invariant properties of some curved-axis generalized cylinders', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (3), pp. 237–253

[7] ZERROUG M., NEVATIA R.: 'Volumetric descriptions from a single intensity image', *Int. J. Comput. Vis.*, 1996, **20**, (1–2), pp. 11–42

[8] ZERROUG M., NEVATIA R.: 'Part-based 3D descriptions of complex objects from a single image', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, **21**, (9), pp. 835–848

[9] LIN C.-A.: 'Perception of 3D objects from an intensity image using simple geometric models', USC PhD thesis, 1996

[10] JELINEK D., TAYLOR C.J.: 'Reconstruction of linearly parameterized models from single images with a camera of unknown focal length', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (7), pp. 767–773

[11] FRANCOIS A.R.J., MEDIONI G.G.: 'Interactive 3D model extraction from a single image', *Image Vis. Comput.*, 2001, **19**, (6), pp. 317–328

[12] CRIMINISI A., REID I., ZISSERMAN A.: 'Single view metrology', *Int. J. Comput Vis.*, 2000, **40**, (2), pp. 123–148

[13] DEBEVEC P.E., TAYLOR C.J., MALIK J.: 'Modeling and rendering architecture from photographs'. Proc. ACM. SIGGRAPH. 1996, pp. 11–20

[14] HOIEM D., EFROS A.A., HEBERT M.: 'Automatic photo pop-up'. Proc. ACM SIGGRAPH, 2005, pp. 557–584

[15] HOIEM D., EFROS A.A., HEBERT M.: 'Putting objects in perspective', Proc. Computer Vision and Pattern Recognition, 2006, pp. 2137–2144

[16] HAN F., ZHU S.C.: 'Bottom-up/top-down image parsing by attribute graph grammar'. Int. Conf. Computer Vision, 2005, pp. 1778–1785

[17] BHALERAO A., WILSON R.: 'Unsupervised image segmentation combining region and boundary estimation', *Image Vis. Comput.*, 2001, **19**, (6), pp. 353–368

[18] GREEN P.J.: 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika*, 1995, **82**, pp. 711–732

[19] DICK A., TORR P., CIPOLLA R.: 'A Bayesian estimation of building shape using MCMC'. Proc. 7th European Conference on Computer Vision (ECCV'02), Copenhagen, June 2002, vol. II, pp. 852–866

[20] TU Z., ZHU S.C.: 'Image segmentation by data driven Markov Chain Monte Carlo', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (5), pp. 657–673

[21] BARBU A., ZHU S.C.: 'Graph partition by Swendsen-Wang cuts'. Int. Conf. Computer Vision, 2003, pp. 320–327

doi: 10.1049/iet-cvi:20065002