Multi-view urban scene reconstruction in non-uniform volume

Runchao Mao, Qiang Wu, Yu Qiao, Li Bai, Jie Yang

Runchao Mao, Qiang Wu, Yu Qiao, Li Bai, Jie Yang, "Multi-view urban scene reconstruction in non-uniform volume," Proc. SPIE 9067, Sixth International Conference on Machine Vision (ICMV 2013), 90670A (24 December 2013); doi: 10.1117/12.2049800

SPIE.

Event: Sixth International Conference on Machine Vision (ICMV 13), 2013, London, United Kingdom

Multi-view Urban Scene Reconstruction in Non-uniform Volume

Run-Chao Mao[1], Qiang Wu[2], Yu Qiao[1], Li Bai[3], Jie Yang[1]
[1]Department of Automation, Shanghai Jiao Tong University, Shanghai, China
[2]School of Computing and Communications, University of Technology Sydney, Sydney, Australia
[3]School of Computer Science, University of Nottingham UK, Nottingham, UK

ABSTRACT

This paper presents a new fully automatic approach for multi-view urban scene reconstruction. Our algorithm is based on the Manhattan-World assumption, which can provide compact models while preserving fidelity of synthetic architectures. Starting from a dense point cloud, we extract its main axes by global optimization, and construct a non-uniform volume based on them. A graph model is created from volume facets rather than voxels. Appropriate edge weights are defined to ensure the validity and quality of the surface reconstruction. Compared with the common point-cloud-to-model methods, the proposed methodology exploits image information to unveil the real structures of holes in the point cloud. Experiments demonstrate the encouraging performance of the algorithm.

Keywords: multi-view stereo, 3D reconstruction, structure from motion, Manhattan-World, graph-cut.

1. INTRODUCTION

3D reconstruction is one of the core problems in computer vision. Past decades have witnessed the development of various modeling techniques, from depth cameras to aerial LiDAR. However, the inflexibility, high cost, and difficulty in capturing color information have limited the applications of these techniques in many areas. Through widely available internet applications such as Flickr and Google image, it is not difficult to obtain images of a typical object such as a building from various views. The availability of such resources also make image based 3D reconstruction more popular. Multi-view reconstruction is however difficult. Most solutions rely on manual interactive process which is acceptable for reconstructing small scenes, but not for a large scale urban scene. This paper describes a novel 3D reconstruction method we have developed to tackle the challenges of reconstructing large scale urban scenes.

Several automatic 3D reconstruction methods have been developed. Volumetric approaches are widely used in image-based reconstruction because of their flexibility and simplicity. The approaches often use 2D projections of object contours and discretization of the space into uniform voxels [1, 2]. They are however limited by the tradeoff between volume resolution and computational cost, rendering them unsuitable for reconstructing large scale scenes.

Other approaches separate the overall 3D reconstruction into two sub-processes: Computer Vision (CV) process and Computer Graphics (CG) process. The CV step reconstructs a dense point cloud via structure from motion (SfM) [3-5], and the CG step reconstructs the surface based on the dense point cloud. Point clouds are very suitable for large scale scene reconstruction, as points can be handled more easily. The CG step requires the point cloud to form a manifold and to be dense everywhere. However, some observing angles are difficult to capture. Thus the obtained surfaces are open surface rather than manifolds. In addition, urban scenes are usually lacking of textures which lead to large holes in the point cloud created by SfM.

Automatic urban scene reconstruction algorithms usually follow the CV and CG process. They require certain assumptions to be made according to the common structures of synthetic architectures. Manhattan-World assumption [6], piece-wise planar priors [7], or triangle meshes combined with various shape primitives [8] are widely used to make the models succinct. Urban scene reconstruction based on these assumptions is closer to real situations. However, their complexity depends on predefined model parameters, rather than the complexity of the scene itself. Methods for processing large scale urban scene data can be very complex [3].

In this paper, a new algorithm is proposed, which is based on the Manhattan-World assumption. It extracts plane hypotheses from the point cloud, and constructs a non-uniform volume, as discussed in section 2. This volume can be used to improve the accuracy of 3D reconstruction and tackle the limitation of scalability widely seen in common volumetric methods. In section3, image information is utilized to guide the reconstruction and identify whether those

Sixth International Conference on Machine Vision (ICMV 2013), edited by Antanas Verikas, Branislav Vuksanovic, Jianhong Zhou, Proc. of SPIE Vol. 9067, 90670A · © 2013 SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2049800

holes in the point cloud should be filled as parts of the surface or be left as genera on the manifold. The surface is finally obtained through graph-cut. Experimental results are shown in section 4, and conclusion is drawn in section 5.

2. NON-UNIFORM VOLUME CONFIGURATION

SfM is employed in our algorithm first to create a point cloud $\{P_i\}$. Most of the SfM algorithms can be used, such as those described in [3] and [4]. The volume planes are extracted by clustering and are non-uniform, which means that the voxels in the volume are cuboids rather than cubes. This does not affect the accuracy of the solution. Most of object planes in urban scene can be located on a volume plane.

2.1 Main Direction Calculation

Under Manhattan-World assumption, the scene is composed of planes whose normal vectors are parallel to three mutually orthonormal directions respectively, which are called as main directions. We calculate normal vectors of points from their neighbors through PCA and choose the main directions as below:

$$\arg\max_{R} \sum_{i=1}^{3} \sum_{j=1}^{N} \delta(\vec{n}_j, Re_i), \tag{1}$$

where e_i is one of the initial orthonormal basis and $\vec{n}_j$ is the normal vector of a point. R represents the rotation of initial basis. $\delta(\square)$ is the voting function implying that two normal vectors are close enough.

The problem is non-convex and the search space is large. Thus it is hard to get a globally optimal solution. We solve this problem by branch-and-bound algorithm [9], which utilizes interval analysis to effectively reduce the scale of search space and guarantees a global optimum. After a proper rotation is determined according to (1), the original point cloud is aligned with the proper coordinate system based on this selected rotation.

2.2 Plane Hypothesis

Given the axes of the volume, our algorithm needs to determine plane hypotheses in order to construct the non-uniform volume. The surface of the object is represented by collections of those plane hypotheses at different positions of different areas. As each plane is perpendicular to only one main direction, its position can be indicated according to the intersection of the plane and the corresponding main direction.

Many clustering algorithms can be applied. In this paper, improved mean-shift [10] is employed, because it is efficient and can position the cluster centers more accurately. The choice of parameters should be to get more clusters in order not to miss many planes. The bandwidth B is the mean of nearest neighbor distance of all points. To further speed up the process, KD-tree [11] is exploited to accelerate the nearest neighbor query.

3. MODEL EXTRACTION

Traditional volumetric algorithms [1, 12] regard reconstruction as voxel selection. Inspired by [13], our reconstruction algorithm focuses on the facets of the volume. A graph model is created based on the facets, where edges reflect the smoothness of adjacent facets and the consistency of those facets to the point cloud.

Such graph structure can generate more reliable surfaces. If the graph model is built on voxels, it has to distinguish graph nodes inside or outside of surfaces in order to represent the proper manifold. In our method, the graph model only needs to tell out the nodes on or off the surface which is indicated directly by photo consistency.

3.1 Boundary Surfaces

We first roughly calculate inner and outer boundaries because graph-cut needs separated sink and source nodes. Simple heuristic is exploited in our algorithm. The inner boundary is defined as the surface containing the largest volume which does not occlude any 3D point from its visible view. The outer one is the surface with smallest volume which is not occluded by any 3D point. They can be easily extracted by binary Markov Random Field from the volume, which regards voxels as nodes and labels them as inside or outside. To get inner boundary, occluding voxels are penalized and inside nodes are assigned small negative cost values to get a maximum volume. For outer boundary, occluded voxels are

penalized and outside nodes are assigned small positive cost values. In both cases, there are non-smoothness costs for edges linking neighboring nodes with different labels. The facets between inside and outside voxels form the boundary surfaces.

3.2 Surface Reconstruction

Given the boundaries of the surface, we construct a graph on the volume voxels between them, where voxel facets F_i correspond to nodes, and intersecting facets are connected by undirected edges e_{ij}. Note that two coincident facets of different orientations are represented by two separated graph nodes which are not connected.

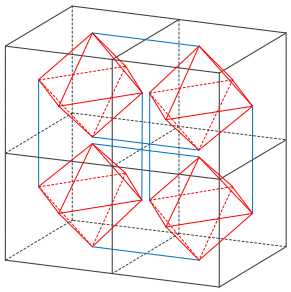

Figure 1. Graph structure of the non-uniform volume

We can formulate the original surface reconstruction as an optimization problem, where the cost function to be minimized can be expressed as:

$$E(S) = \int NC(S) \cdot dS + \int NS(S) \cdot dS, \tag{2}$$

where S represents the surface, $NC(S)$ expresses the inconsistency between the surface and point cloud, and $NS(S)$ measures its non-smoothness [14]. They can be discretized into $NC(F_i)$ for facets and $NS(e_{ij})$ for edges.

Visibility conflicts are reflected by $NC(F_i)$. If a facet occludes a point along its visible view or is occluded by it, facet conflicts are present. We query each point's visibility on all input images. Conflicts can be determined by simple depth comparison and $NC(F_i)$ can be defined as:

$$NC(F_i) = \left(1 - e^{-\sum C_{ij}}\right) \cdot S_i, \tag{3}$$

where S_i is the area of facet F_i. C_{ij} equals to the confidence of point P_j if it conflicts with F_j, and zero otherwise. Note that both visibility and confidence are calculated directly from the point cloud and input images [4].

$NS(e_{ij})$ is normally constant α in most algorithms to maintain the smoothness of the surface. However, it imposes naïve surface shrinkage which will cause artifacts. In this paper, we keep the non-smooth ridges of the surface along the contours on input images. As shown in figure 1, there are two different edges: non-coplanar edges (red lines) and coplanar edges (blue lines). They are considered differently because the former ones correspond to non-smooth surface ridges, while the latter ones only imply the surfaces. Thus a non-coplanar edges which is not consistent with input images should be penalized more heavily than others. We first query all facets' visibility [4], and use V_i to represent the set of visible views of F_i. $NS(e_{ij})$ is set to be 10α if and only if e_{ij} is non-coplanar and the projection of $F_i \cup F_j$ onto any view of $V_i \cap V_j$ does not intersects its image contours. Therefore, for those facets in the middle of surface planes, this term imposes shrinkage and ridges will be restricted along the locations of image contours. Those contours should be parallel to one of the three main directions under Manhattan-World assumption. As vanishing points are known for all images, they can easily be extracted by edge detection and simplified Hough transformation because there are only three possible common points of those edges and the only degree of freedom is the polar angle.

Edge weights are formulated as:

$$w_{ij} = \frac{1}{2}\left(NC(F_i) + NC(F_j)\right) + NS(e_{ij}) \cdot \tag{4}$$

Facets coincide inner or outer boundaries are linked to sink or source respectively. They are assigned with infinite weights. Then the surface reconstruction is achieved through a min-cut. Since this cost function is sub-modular, graph-cut can be solved within polynomial time as [14] has proved.

4. EXPERIMENTS AND DISCUSSION

We have tested our algorithm on several real datasets. Figure 2 shows typical examples. It shows that large portion of areas is lack of textures, and only two facades are captured by the photos.

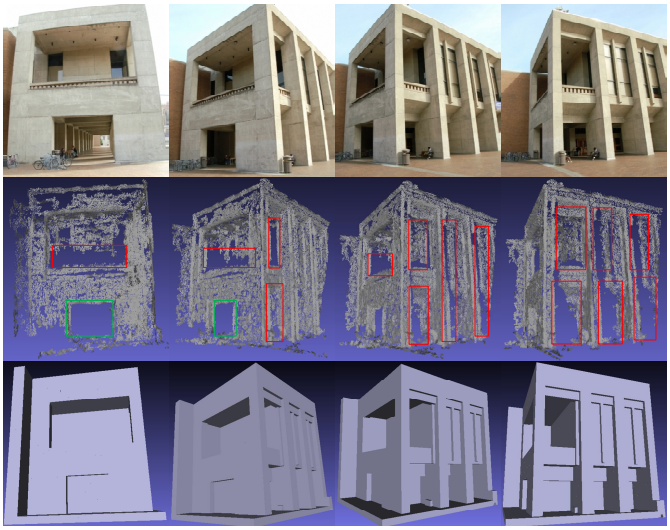

Figure 2. 4 of 10 input images (first row) and corresponding screen shots of the point cloud (second row) and the 3D model (third row)

PMVS [4] is used as the initial SfM algorithm in our approach. Figure 2 shows that the point cloud contains a lot of large holes, which will cause great uncertainty during surface reconstruction. The rectangles highlight holes in the point cloud. Some of them are real empty spaces on the building and others are fake holes which are supposed to be part of surfaces. This will cause extreme difficulties to surface reconstruction because it is difficult to determine whether it is necessary to fill in these holes. Manhattan-World assumption allows us to expand scattered points into surface because it assigns simple structures (planes with three orientations) to the surface areas, and locates the ridges only at certain positions. The surface can be reconstructed with more certainty because there are few possible structures for those holes in the point cloud under the smoothness prior: genera or smooth surface. In addition, image information is exploited in our algorithm to help determine those structures. On the contrary, typical CG algorithms cannot sort out the problems caused by such holes since they do not have clues for surface structure. The result shows that our algorithm is able to restore most of its topology correctly. However, there are still some areas with wrong topology (green rectangles) because only one photo contains the inner part of that hole, and it is impossible for SfM to restore 3D information from a single view.

Moreover, the views only contain two sides of the building, which will lead to an open surface for many surface reconstruction methods. However, our approach works well because clusters naturally form a tight bounding box to complete the manifold in our algorithm. Although there are no views of the rooftop and other two facades, points on the edge of the existing facades can generate planes to bound the 3D model and form a closed surface.

Our volume structure reduces the complexity of reconstruction significantly because it is non-uniform and is generated from clustering. Small details can be captured, while large structures do not need to be separated into massive voxels, as uniform volume does. In the experiments, there are only 67 planes in the volume extracted from mean-shift clustering, which makes the volume and graph model quite tractable. On the contrary, uniform volume needs over 100000 voxels for this example to maintain those details in our results, since the scale of the smallest facet in our volume

is 1/50 of the scale of the whole model. This implies that the complexity of the algorithm depends on the scene complexity rather than predefined modeling parameters. It makes the proposed method scalable to handle large scenes, because they usually contain both large compact structures and small details. Sampling the whole scene uniformly by fixed parameters in such situation will either miss more details or lead to huge complexity for the large structures, which may just consist of several planes. In addition planes in our volume can represent real planes more accurately, while common methods based on uniform volumes need further steps to adjust plane aliases [6].

5. CONCLUSION

We propose a multi-view 3D reconstruction algorithm for urban scenes based on Manhattan-World assumption in this paper. Non-uniform volume is exploited to ensure efficiency and accuracy. The graph structure is constructed from voxel faces to avoid uncertainty. Edge weights are defined according to images to guide surface reconstruction for large holes. Experimental results show that the proposed method can handle main difficulties for scene reconstruction.

ACKNOWLEDGEMENT

This research is partly supported by NSFC, China (No: 61273258), Ph.D. Programs Foundation of Ministry of Education of China (No.20120073110018).

REFERENCES

[1] C.H. Esteben and F.Schmitt,"Silhouette and stereo fusion for 3D object modeling,"CVIU. **96**, 367-392, (2004)

[2] B. Curless and M. Levoy,"A volumetric method for building complex models from range images,"Proc. Computer Graphics and Interactive Techniques, 303-312, (1996)

[3] G. Zeng, S. Paris, L. Quan and F. Sillion,"Accurate and scalable surface representation and reconstruction from images,"IEEE Trans. PAMI. **29**, 141-158, (2007)

[4] Y. Furukawa and J. Ponce,"Accurate, dense, and robust multiview stereopsis,"IEEE Trans. PAMI. 32, 1362-1376, (2010)

[5] M. Vergauwen and L. Van Gool,"Web-based 3d reconstruction service,"Machine Vision and Applications. 17, 411-426, (2006)

[6] Y. Furukawa, B. Curless, S. Seitz and R. Szeliski,"Reconstructing building interiors from images,"Proc. IEEE ICCV, 80-87, (2009)

[7] B. Mičušík and J. Košecká,"Multi-view superpixel stereo in urban environments,"IJCV, 89, 106-119, (2010)

[8] F. Lafarge, R. Keriven, M. Brédif and V.H. Hiep,"Hybrid multi-view reconstruction by jump-diffusion,"Proc. IEEE CVPR, 350-357, (2010)

[9] J.C. Bazin, et al,"Globally optimal line clustering and vanishing point estimation in Manhattan world,"Proc. IEEE CVPR, 638-645, (2012)

[10] D. Comaniciu and P. Meer,"Mean shift: A robust approach toward feature space analysis,"IEEE Trans. PAMI. **24**, 603-619, (2002)

[11] W. Hunt, W.R. Mark and G. Stoll,"Fast kd-tree construction with an adaptive error-bounded heuristic," IEEE Symposium on Interactive Ray Tracing, 81-88, (2006)

[12] C. Hernández, G. Vogiatzis and R. Cipolla,"Probabilistic visibility for multi-view stereo,"Proc. IEEE CVPR, 1-8, (2007)

[13] A. Hornung and L. Kobbelt,"Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding," Proc. IEEE CVPR, 503-510, (2006)

[14] V. Kolmogorov and R. Zabin,"What energy functions can be minimized via graph cuts?"IEEE Trans. PAMI. **26**, 147-159, (2004)