Automatic 3D Reconstruction from a Single Catadioptric Omnidirectional Image

Hui Gao, Wei Xu and Maojun Zhang

Multimedia and Virtual Reality Laboratory
College of Information System and Management
National University of Defense Technology, Changsha, China, 410073

Abstract—**In this paper we present a method for 3D urban reconstruction from a single catadioptric omnidirectional image. Firstly, we classify the catadioptric omnidirectional image to horizontal ground, vertical building surface and vertical background surface through the registration between catadioptric omnidirectional image and remote sensing image. According to the classification results, we recover the geometry based on the catadioptric projection model. The experiment shows that our method is feasible and realizes a precise 3D reconstruction for the city scenes.**

Index Terms—**Catadioptric Omnidirectional Image, 3D Reconstruction, Geometry Recovery, Remote Sensing Image.**

I. INTRODUCTION

Image-based rendering (IBR) is a technique to generate novel views from a set of reference images. It produces photo-realistic output without a complex lighting model [1]. A potential usage of IBR is the walkthrough of 3D city, which may have many applications including car navigation, shopping guides, city planning and virtual tourism. However, IBR generally requires a large number of reference images. And it will have high laboring cost and consume a significant amount of time in the image capturing process. Meanwhile, the capturing of catadioptric omnidirectional image (COI) has become a standard technique. The COI has a 360 degrees field of view, which makes it possible to accomplish a very complete 3D reconstruction from a few images. So, there have been many research projects for 3D reconstruction by using the COIs.

In the literature, the conventional schemes proposed for COI based reconstruction can be divided into two categories: multiple-COIs based method and single-COI based method. Reference [2] estimates the full geometry and accomplishes the 3D model from the COIs sequence by using a Structure from Motion (SFM) method. In [3], depth is obtained from pairs of COIs using a stereo vision technique, and is fused in a probabilistic framework to obtain a more accurate result. Reference [4] describes a single view point catadioptric depth extraction system. The system consists of a camera and two hyperboloids which are coaxially aligned. The depth is

extracted by calculating the disparity of the two COIs. Because the overlap of the images reflected from two hyperboloids is small, the field of view is limited.

For the single-COI based methods, the stereo vision system in [5] is consisted by two components: an omnidirectional camera and a laser projector. The geometry can be recovered according to the structured lights from the laser projector. This method can obtain the more accurate results but higher cost. Reference [6] presents a 3D objects reconstruction from a single panoramic image. The approach is based on the user-provided coplanarity, perpendicularity and parallelism constraints. Reference [7] uses a semi-automatic reconstruction process, in which the user marks the room corners in the panoramic images. The corners can be translated into the viewing-angle measurements, from which the exact sizes of the walls can be computed. This method can realize a coarse reconstruction of the indoor environment.

This paper presents an automatic 3D reconstruction approach for urban scenes from a single COI. For capturing a COI, we use an omnidirectional camera (shown in Fig. 1(a)) consisting of a standard digital camera, a paraboloid mirror located in front of the digital camera and a GPS. The processing pipeline is shown in Fig.1 (b), and entails the following steps: classify the scenes in COI as the horizontal ground, vertical buildings and vertical background scenes by using registration between COI and remote sensing image. With the classification results, calculate the geometry and reconstruct the 3D scenes based on catadioptric projection model. Compared with the previous approaches, the proposed method has the advantages of simple and compact hardware system, low cost and automatic processing.

II. GEOMETRY RECOVERY FROM A SINGLE CATADIOPTRIC OMNIDIRECTIONAL IMAGE

Here, we develop a geometric recovery algorithm from an omnidirectional image based on the catadioptric projection model. Without loss of generality, we assume that the vertex of the paraboloid mirror lies at the origin **O** of a Cartesian coordinate system (Fig. 2). Suppose that the effective pinhole is located at point **F**. The y-axis lies in the direction $\overrightarrow{FO}$ and make sure that $\overrightarrow{FO}$ is vertical to the ground. Since the perspective projection is rotationally symmetric about any line through **F**,

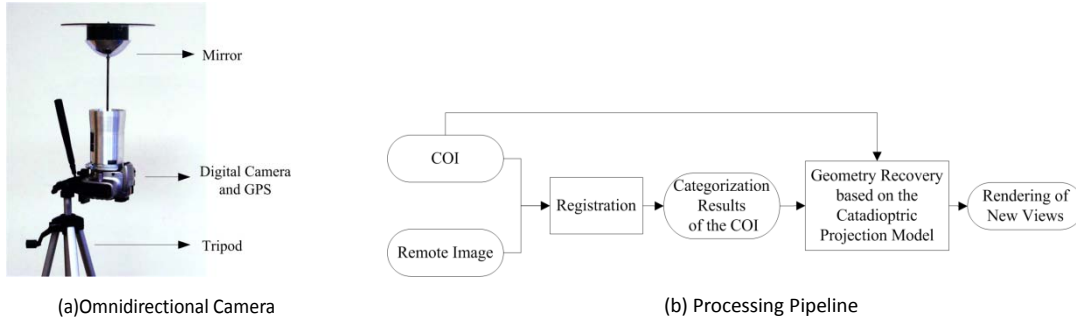we work in the 2-D Cartesian frame $(O, \hat{x}, y)$ where $\hat{x}$ is a unitvector orthogonal to $\hat{y}$.

(a)Omnidirectional Camera   (b) Processing Pipeline

Fig. 1. Omnidirectional camera and processing pipeline

(a) Geometry calculation of the points on the ground   (b)Geometry calculation of the points on vertical surfaces
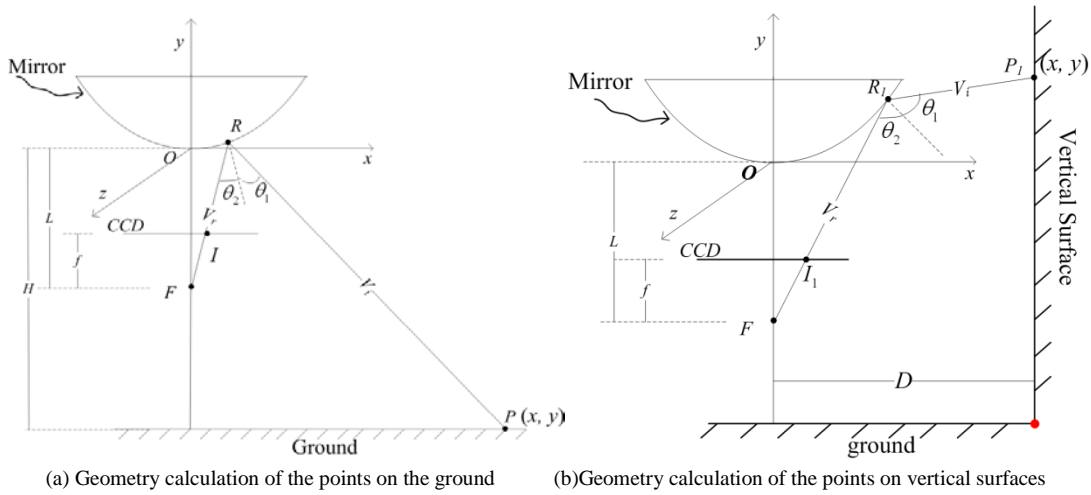
Fig. 2. 3D Reconstruction based on the Catadioptric Projection Model

2.1 Geometry Recovery of Points on a Horizontal Surface

In the 3D reconstruction of urban scenes, the horizontal surface mainly indicates the ground. We assume that an incoming ray $\vec{vi}$ from a world point $\mathbf{P}$ on the ground is reflected at $\mathbf{R}$, and images a pixel $\mathbf{I}$ by the reflected ray $\vec{vr}$. We can calculate the 3D position of $\mathbf{P}$ according to the position of $\mathbf{I}(i, j)$ in the COI. In Fig.1(a), H is the height of the mirror from the ground which is a constant value and could be measured previously. L is the distance of the mirror from F and f is the camera focal length. Note that the intersection of $\vec{vi}$ and the ground (ground equation is $y=-H$) is what we calculate. The detailed conductions are shown as follows.

The slope and equation of $\vec{vr}$ are $\tan\beta = \dfrac{f}{i}$ and $y = \dfrac{f}{i}x - L$, respectively. The equation of mirror shape is defined as $y=ax^2$. We can easily get the coordinates of $\mathbf{R}$ which is the intersection

of $\vec{vr}$ and the mirror surface:

$$R(x_1, y_1) = \left(\left(\frac{f}{i} + \sqrt{(\frac{f}{i})^2 + 4aL} \right) \Big/ 2a, a\left(\left(\frac{f}{i} + \sqrt{(\frac{f}{i})^2 + 4aL} \right) \Big/ 2a \right)^2 \right) (1)$$

The slope of the parabola at $\mathbf{R}$ is $\tan\alpha = 2ax|_{x=x_1} = 2ax_1$. According to the reversibleness of rays, and the reflection angle θ_2 should be equal to the incidence angle θ_1. We can calculate the slope of $\vec{vi}$: $a_1 = \tan(2\alpha - \beta)$. And the equation of $\vec{vi}$ is:

$$y = a_1x + (y_1 - a_1x_1) \qquad (2)$$

By solving the equations that consist of $y=-H$ and (2), we can obtain the position of world point $\mathbf{P}$:

$$P(x, y) = \left((-H - (y_1 - a_1x_1))/a_1, -H \right) \qquad (3)$$

2.2 Geometry Recovery of Points on a Vertical Surface

We assume that a incoming ray from a point $\mathbf{P_1}$ on vertical surface is reflected at $\mathbf{R_1}$, and images a pixel $\mathbf{I_1}$. Now we calculate the geometry of $\mathbf{P_1}$ according to the position of $\mathbf{I_1}$ (i, j) in the COI. We define the distance from the vertical surface to

y-axis as *D*. Compared with section 2.1, similar deductions are applied to the points on vertical surfaces. So we can obtain the position of **P₁** by solving the equations of *x=D* and (3):

$$P_1(x, y) = \left(D, a_1 D + (y_1 - a_1 x_1)\right) \qquad (4)$$

We find out that there is also an unknown parameter *D* in (4), which is the distance from the vertical surface to y-axis. Fortunately, using (3) we can calculate the distance from any point on the ground to y-axis. So, if we find out a special point which is both on the vertical surface and the ground, we could obtain *D* by using (3). The next section describes a method to find out the borderlines between the vertical surfaces and the ground where the points satisfy our requirement.

The 3D geometry of a world point calculated from the 2D position of P and P₁ is as follows:

$$\begin{cases} x_P = \dfrac{|x_I - W_I/2|}{x_I - W_I/2} \times \left| \dfrac{(x_I - W_I/2)x}{\sqrt{(x_I - W_I/2)^2 + (y_I - H_I/2)^2}} \right| \\[4mm] y_P = y \\[2mm] z_P = \dfrac{|y_I - H_I/2|}{y_I - H_I/2} \times \left| \dfrac{(y_I - H_I/2)x}{\sqrt{(x_I - W_I/2)^2 + (y_I - H_I/2)^2}} \right| \end{cases} \qquad (5)$$

Where (x_I, y_I) is the position of I in COI; W_I and H_I are the width and height of COI, respectively. However, the precondition is that we must identify which kind of surface (horizontal or vertical) the pixel in COI comes from. Next section presents a region classification method based on the registration between COI and remote sensing image. The method can automatically classify the scenes in COI as the horizontal ground, vertical building and background scenes.

III. REGION CLASSIFICATION OF CATADIOPTRIC OMNIDIRECTIONAL IMAGES

This section describes the method to classify a COI into ground, buildings and background and extract the borderlines among the regions. The classification results are the basis of the geometry recovery in Section 2, because they can tell us which kind of surface (horizontal or vertical) the pixel in COI comes from.

With the developments of remote sensing and network techniques, there are more and more precise remote sensing images on Internet. The viewpoint of a remote sensing image is as high as outer space, so it shows us a top view which could describe the object positions on ground more clearly. So we try to classify the COI through the registration between COI and remote sensing image. We first extract the horizon, borderlines between the buildings and the ground, and the skyline in a COI. We consider the region between the skyline and the bottom borderline of building to be region of building, the region between the skyline and the horizon to be region of background scenes, and the other region inside the horizon to be the ground. The background scenes are defined as what are so far from the viewpoint that cannot be calculated the geometry, for example the buildings and the trees in the distance. In this paper we use the 16-bit integer to represents the COI geometry with Centimeter as the unit. So the biggest distance we can represent is 655.36m. We consider the background scenes as a vertical surface in a circle round the view point with an arbitrary

distance of 655.36m.

Our method consists of three steps: the registration between the omnidirectional image and remote sensing images, the extraction of the horizon and the bottom borderline of building and the extraction of skyline.

3.1 Registration between the Ground Panorama and Remote Sensing Images

Through registration, we can obtain the position correspondences between the COI and remote sensing images. In order to weaken the distortion in COI (Fig. 3(a)) and obtain an accurate registration, we transform the COI from catadioptric vision model to perspective model. In Section 2.1, if we consider the ground as a projection plane, (3) will give us pixels correspondences between the COI and the projection plane. And the result transformed into the projection plane is a perspective image, which is referred to as *Ground Panorama*. Fig. 3(b) shows the ground panorama transformed from the COI of Fig. 3(a).

The GPS records a rough position where the COI was captured. It can help us reduce the searching area during registration. Fig. 3(c) shows a remote sensing image downloaded from Google Earth according to the GPS positional data (North Latitude: 28.1391°, East Longitude: 113.0324°). With GPS positional data as the center, a square of 80m × 80m is chosen as the searching area (the red block in Fig. 3(c)). And we have also learned the following properties in the registration:

- Remote sensing images from Google Earth have been appended the direction information, and we can also get the direction information of ground panorama with the help of GPS. Thus, we can guarantee the same direction during the processing of registration.
- Remote sensing images from Google Earth have been appended the scale information, and we can also calculate the scale information of ground panorama according to (5) and relevant parameters in Section 2. Thus, we can guarantee the same size during the processing of registration.
- Make sure that the direction of omnidirectional camera is approximately vertical and the affine distortion of ground panorama will be effectively reduced or omitted.

The above properties guarantee that we can choose a template based method to accomplish the registration. So, we apply the algorithm of normalized cross correlation (NCC) [8] for our registration. In order to registry accurately, the ground panorama is cut out as Fig. 3(d), this is because there are so much badly distortion in the outer part of Fig. 3(b). And the red point **O_c** in Fig. 3(d) represents the optical center of the omnidirectional camera. Registry Fig. 3(d) in the searching area in Fig. 3(c), and the result is shown in Fig. 3(e). The corresponding position of **O_c** in remote sensing image is referred to as *View Point Position* (VPP) where the COI was taken.

3.2 Extraction of the Horizon and the Bottom borderline of building

According to the correspondences between the COI and the

projection plane given by (3) and with VPP (seen in section 3.1) as the viewpoint, we can transform the remote sensing image to catadioptric vision model (Fig. 4(a)). In the transformed image the ground has shrunk to a circular form whose edge is exactly the horizon.

We now begin to extract the building bottom borderline. Between Fig. 4(a) and a COI, the earth could match well, while the buildings cannot match because a COI (Fig. 3(c)) includes the side information of the vertical scenes which a remote sensing image is lack. We calculate the subtractive image

between the COI and Fig. 4(a) in gray scale where the higher intensity indicates what can not match. Define a threshold with Otsu method [14] and convert the subtractive image to a binary image where 1 represents the building and 0 represents earth. Scan from the center of the binary image, and it will be defined as the borderline when meet the 1 at the first time. Finally, we obtain the bottom borderlines of building shown in Fig. 4(b). And in Fig. 4(c), the bottom borderlines of buildings and the horizons are respectively labeled in red and yellow.
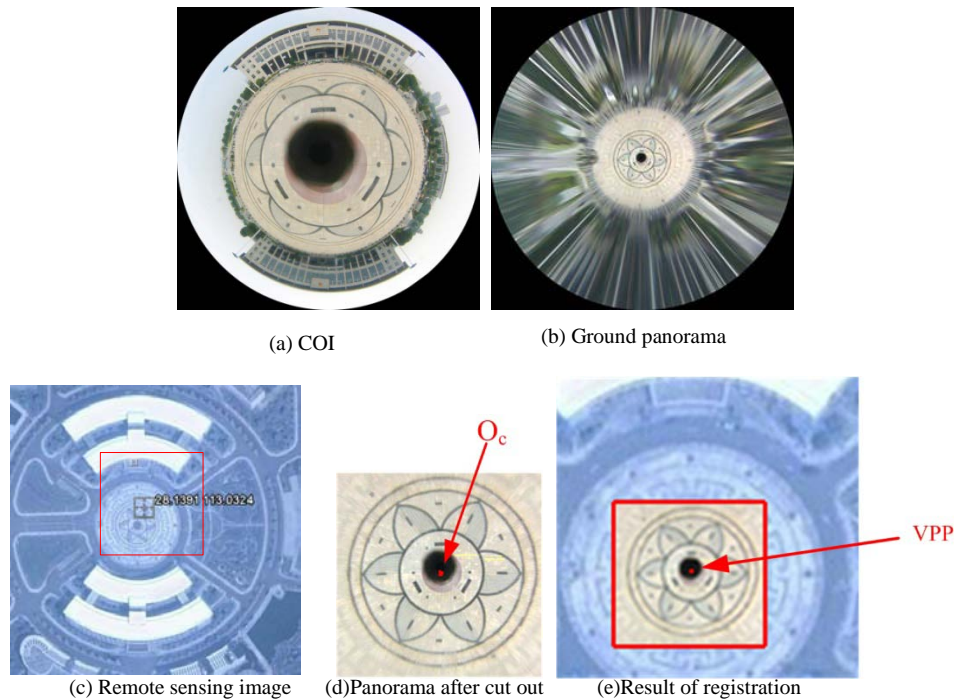

(a) COI   (b) Ground panorama

(c) Remote sensing image   (d)Panorama after cut out   (e)Result of registration

Fig. 3. Registration between catadioptric omnidirectional image and remote sensing image
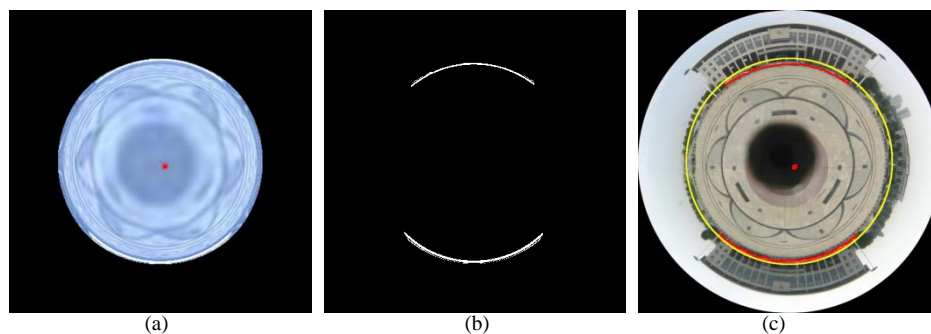

(a)   (b)   (c)

Fig. 4. Extraction of the Horizon and the Bottom Borderline of Building

3.3 Skyline Extraction

In this section, we firstly oversegment the COI (Fig. 5) using [12], and then take SVM [9]-[11] method to classify each segment as 'sky' or 'non-sky'. We choose 3218 segments from 5 COIs as the training samples. In the training and prediction process, we select the color, position, percentage of area and texture as classification features. Train and predict by using LIBSVM [13]. Finally, fuse the 'sky' labeled segments to big

regions and extract its borderline as skyline (Fig. 5).

Once we have obtained the skylines, horizons and bottom borderlines of building, we can classify a COI as regions of ground scenes, building facade and background scenes. The result is shown in Fig.6. The blue region represents the ground; red region represents the building facade; and green region represents the background scenes. Thus, we could identify which kind of face (horizon ground or vertical surface) the pixel in COI comes from. And, we could also find out the points which are on both the vertical surface and the ground according

to the borderline between the building and the ground. By now it has been an easy work to accomplish the 3D reconstruction according to the algorithm in Section 2.
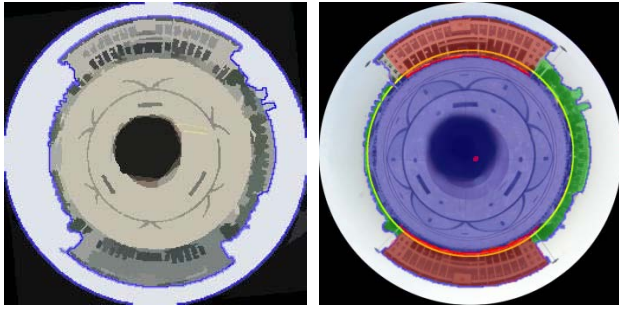

Fig. 5. Extraction of the Skyline   Fig. 6. Classification of the COI

IV.   EXPERIMENTS

In this section we experiment with the pipeline shown in Fig.1(b). The parameter a of our omnidirectional camera (Fig. 1(a)) is a=1/7. The COI and the GPS positional data were taken at a plaza in Changsha of China, when *L* was 20cm and *H* was 400cm. According to the classification results, we recover the geometry of the COI by using the algorithm in Section 2, and the result is represented as a depth map shown in Fig. 7(a) (the gray scale is brighter, the distance is farther). The distance and height of both buildings are respectively 40m and 17.28m. The errors of distance and height are respectively 4.7% and 4% compared with the real ones (42m and 18m). It is noted that there are some artifacts in the rendering: 1) since it is difficult to recover the geometry of sky, we give the sky a constant color when rendering; 2) the background scenes are rendered as a vertical surface in a circle round the viewpoint with the biggest distance (D =655.36m). We can conclude that the depth given by our method is smooth, and include a 360° field of view.

Applying the geometry to the rendering processing, we can obtain the walking through results shown in Fig.7. Fig. 7(b), Fig. 7(c) and Fig. 7(d) show a sequence of views when walking forward. Fig. 7(e) represents the image when turning left from Fig. 7(d). Fig. 7(f) shows a image of large angle of view. The buildings are arced in real world (Fig. 7(g)), and the results faithfully present this shape. Therefore, our method can overcome the disadvantage in [7] and [6].

(a)   (b)   (c)

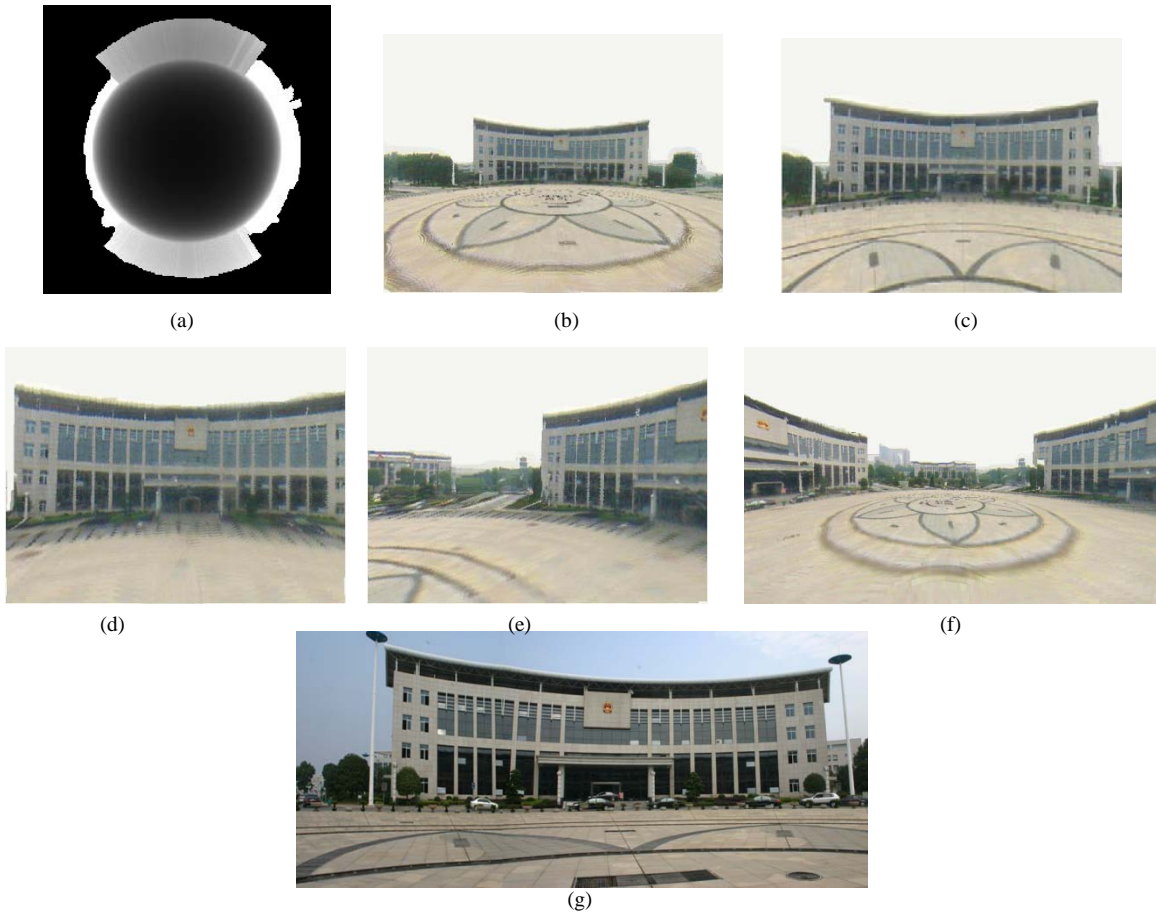(d)   (e)   (f)

(g)

Fig. 7. Results of 3D reconstruction for urban environments

V.   CONCLUSION

In this paper, we have described a method to reconstruct 3D scenes from a single catadioptric omnidirectional image. We did not choose the stereo vision technique [2]-[4] during the geometry recovery, but employed the registration with the remote sensing image from the Internet. The registration helps us classify a COI as regions of ground scenes, building facade and background scenes. We next recovered the geometry according to the catadioptric projection model. The whole processing is fully automatic. This method can obtain the geometry in a larger angle of view compared with [2]-[4], can reconstruct non-planar scenes compared with [6] and [7], and also has a lower cost than [5]. However, the precision of our method is influenced by the registration precision. And there is not a method can obtain a good registration result in any situation, so we have to improve the robustness of the method next step.

ACKNOWLEDGEMENT

REFERENCES

[1]   A.M.K. Siu, A.S.K. Wan, and R.W.H. Lau. Modeling and Rendering of Walkthrough Environments with Panoramic Images, in *VRST'04*, Hong Kong, pp.114-121, 2004.

[2]   M. Lhuillier. Toward Flexible 3D Modeling using a Catadioptric Camera, in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, pp.1-8, 2007.

[3]   R. Bunschoten and B. Kröse. 3D scene reconstruction from cylindrical panoramic images, *Robotics and Autonomous Systems*, vol.41, pp.111–118, 2002

[4]   G. Jang, S. Kim and I. Kweon. Single Camera Catadioptric Stereo System, in *The 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras*, Beijing, 2005.

[5]   R. Orghidan. Catadioptric Stereo based on Structured Light Projection, PhD thesis, University of Girona, 2005.

[6]   P. Sturm. A Method for 3D Reconstruction of Piecewise Planar Objects from Single Panoramic Images, in *Proc. IEEE Workshop on Omnidirectional Vision table of contents*, Washington, DC, pp.119, 2000

[7]   D. Farin, W. Effelsberg and P.H.N. With. Floor-Plan Reconstruction from Panoramic Images. In *Proc. the 15th international conference on Multimedia table of contents*, New York, pp.823 – 826, 2007.

[8]   Q. Li and B. Zhang. A Fast Matching Algorithm Based on Image Gray Value, *Journal of Software*, 17(2), pp.216-222 2006. (in Chinese)

[9]   Z.Q. Bian and X.G. Zhang. Pattern Recognition. Tsinghua University Press, 2000.

[10] Y.X. Chen and J.Z. Wang. Support vector learning for fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 1(11), pp.716-728, 2003.

[11] P. Mitra, C.A. Murthy and S.K. Pal. A probabilistic active support vector learning algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3), pp.413-418, 2004.

[12] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*. 159(2), pp.167-181, 2004.

[13] C.C. Chang and C.J. Lin. LIBSVM: a library for support vector machines. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.

[14] N. Otsu. A Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp.62-66, 1979

Hui Gao was born in HeBei province, China, in 1981. He received his B.S. degrees in system engineering from National University of Defense Technology, Changsha, China, in 2007. He is currently pursuing the Ph.D. degree in system engineering. His research interests include image processing, omnidirectional image, 3D urban reconstruction and virtual reality.

Wei Xu was born in Nanchang, Jiangxi province, China, in 1973. He received the B.S. and Ph.D. degrees respectively in electronic science &technology and system engineering from National University of Defense Technology, Changsha, China, in 1999 and 2007 respectively. He is currently an associate professor in the department of system engineering, National University of Defense Technology. His research interests include image/video processing and virtual reality technology.

Maojun Zhang was born in Huanggang, HuBei province, China, in 1971. He received the B.S. and Ph.D. degrees in system engineering from National University of Defense Technology, Changsha, China, in 1992 and 1997 respectively. He is currently a professor in the department of system engineering, National University of Defense Technology. His research interests include image/video processing, information system engineering, system simulation and virtual reality technology.