J Intell Robot Syst (2011) 62:383–396
DOI 10.1007/s10846-010-9456-1

An Automation System of Rooftop Detection and 3D Building Modeling from Aerial Images

Fanhuai Shi · Yongjian Xi · Xiaoling Li · Ye Duan

Received: 9 November 2009 / Accepted: 26 July 2010 / Published online: 7 August 2010

Abstract This paper presents a prototype system of rooftop detection and 3D building modeling from aerial images. In this system, without the knowledge of the position and orientation information of the aerial vehicle a priori, the parameters of the camera pose and ground plane are first estimated by simple human–computer interaction. Next, after an over-segmentation of the aerial image by the Mean-Shift algorithm, the rooftop regions are coarsely detected by integrating multi-scale SIFT-like feature vectors with SVM-based visual object recognition. 2D cues alone however might not always be sufficient to separate regions such as parking lots from building roofs. Thus in order to further refine the accuracy of the roof-detection result and remove the misclassified non-rooftop regions such as parking lots, we further resort to 3D depth information estimated based on multi-view geometry. More specifically, we determine whether a candidate region is a rooftop or not according to its height information relative to the ground plane, whereas the candidate region's height information is obtained by a novel, hierarchical, asymmetry correlation-based corner matching scheme. The output of the system will be a water-tight triangle mesh based 3D building model texture mapped with the aerial images. We developed an interactive 3D viewer based on OpenGL and C++ to allow the user to virtually navigate the reconstructed 3D scene with mouse and keyboard. Experimental results are shown on real aerial scenes.

F. Shi (✉) · Y. Xi · X. Li · Y. Duan
University of Missouri-Columbia, 321 Engineering Building West,
Columbia, MO 65201, USA
e-mail: fhshi@sjtu.edu.cn

F. Shi
Shanghai Jiao Tong University, Shanghai 200240, China

Keywords Image-based modeling · Asymmetry corner matching · 3D reconstruction

1 Introduction and Related Work

3D building reconstruction from aerial images has many useful applications such as urban planning, cartography, surveillance, robotic path planning, entertainment, virtual tourism, geo-spatial web browsing, etc. In the earlier works, building detection and reconstruction are typically conducted interactively, which has been well studied by researchers in the computer vision and photogrammetry communities.

In the recent two decades, automatic approaches to create 3D geometric models of the buildings from aerial images has received more and more attention from researchers in photogrammetry and computer vision (see [15] for a survey). Many of the existing works either compute a dense digital elevation map (DEM) from high-resolution images [10, 11, 19] or directly use light detection and ranging (LIDAR) data [18]. Hu et al. [9] integrated LIDAR, aerial image and ground images together for urban building modeling.

For building detection, a popular approach to generate the hypothesis of the presence of building rooftops is to first detect low-level image primitives such as edges, lines or junctions and then grouping these primitives using perceptual grouping principles (i.e. the Gestalt Law) based on either geometric heuristics or some known statistical model [15]. For example, in 1999 Baillard et al. [1] developed a method of automatically computing a piecewise planar reconstruction based on line matching that can generate complete roof reconstructions from multiple images. This type of building detection approach however strongly depends on the accuracy of the 3D reconstruction and thus requires very high image resolution for the aerial images which might not always be available.

To summarize, existing aerial image based automatic 3D building modeling techniques mainly differs in the choices of building detection, segmentation and reconstruction techniques. Wherein, rooftop detection is a key and very challenging step in this process. Rooftop detection from aerial images has two prominent features that differentiate them from other natural images:

Longer Range of Size Buildings in aerial images exist at very different sizes, from large blocks of buildings to small, individual house. It is very difficult if not impossible to detect and model them successfully at a single scale.

Variable Shape Configurations Unlike images used for object detection that usually contains a few objects with relatively consistent shape configurations, building rooftops in aerial images can have variable shape configurations with potentially infinite spatial layouts.

Most recently, with the advancement in visual object recognition, more and more machine learning technologies are explored to improve the building detection or rooftops detection in aerial images. Earlier works include nearest-neighbor method and naive Bayesian classifier [14], hierarchical and contextual model [17] etc. These works fully utilized the 2D information of the aerial images and can greatly improve the robustness of building detection.

In this paper, we combine both the 2D visual object recognition and 3D visual computing for rooftop detection and 3D building modeling. The contribution of this work is two folds: (1) one is that the rooftop regions are coarsely detected by integrating multi-scale SIFT-like feature vectors with SVM-based visual object recognition. The overall idea of rooftop detection is to segment the image into regions, similar with the approach in [20], to treat all extracted regions as candidates, and employ visual object recognition to select correct building regions; (2) the other development is a method of rooftop refinement and building modeling by 3D visual computing without the knowledge of the position and orientation information of the aerial vehicle a priori, wherein an asymmetry correlation corner matching is used. The output of the system will be a water-tight triangle mesh based 3D building model texture mapped with the aerial images. We developed an interactive 3D viewer based on OpenGL and C++ to allow the user to virtually navigate the reconstructed 3D scene with mouse and keyboard.

2 Rooftop Detection by Visual Object Recognition

In this section, we propose an original approach for rooftop detection from aerial images. The rooftops of the buildings are seen as independent objects in the image. Our main contribution is to propose original features that characterize rooftop of the buildings.

As a first step of our algorithm, one starts with an initial over-segmentation of an image segmentation by partitioning an aerial image into multiple homogeneous regions. Here we choose to employ Mean Shift [4] as the image segmentation algorithm, see Fig. 1 for an example. It is worth mentioning that our model is not tied to a specific segmentation algorithm. Any method that could propose a reasonable over-segmentation of the images would fit our needs.

2.1 Feature Extraction

When experimenting with each segmented region, we extract the features of the corresponding circum-rectangular image block and use color descriptors by first transforming the (R, G, B) image into the normalized (r, g, b) space [5] where $r = R/(R+G+B)$, $g = G/(R+G+B)$, $b = B/(R+G+B)$. Using the normalized chromaticity value in object detection has the advantage of being more insensitive to small changes in illumination that are due to shadows or highlights. Each rectangular image block will then be resized to a fixed size (96 pixels in this paper) square block by bi-cubic interpolation, without preserving its aspect ratio. This will partially eliminate the influence of the rooftop shape and size on recognition.

Similar to SIFT descriptor [13], multi-scale orientation histogram features are extracted independently from the r and g channel of the normalized image block and concatenated into one 256*n dimensional descriptor, where n represent the number of scale. That is, we only extract a SIFT like descriptor from the image block at each scale, with the whole image block as a frame and without dominant direction assignment. This will partially eliminate the influence of different rotation of the

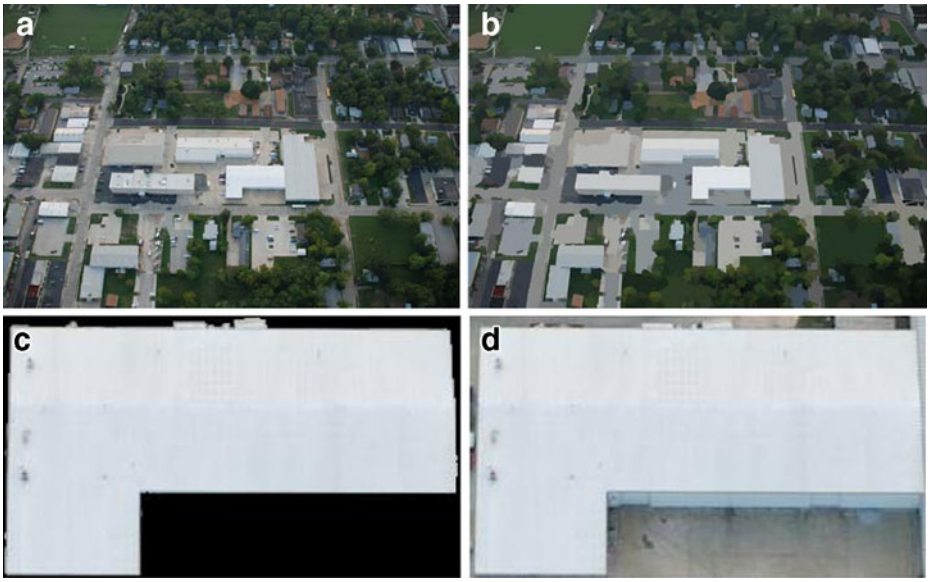

Fig. 1 Sample of aerial image. **a** Original image; **b** over-segmentation of the image by Mean Shift; **c** a segmented image region; **d** circum-rectangular image of the segmented region **c**

rooftop on object recognition. When computing the descriptor at each scale, the σ of the difference-of-Gaussian function is set to be the half of the size of the block.

2.2 Classification

To test the utility of the SIFT like feature for the rooftop recognition, we conducted a standard cross-validated classification procedure on the high-dimensional feature model.

To speed up the computation and improve classification performance, we reduced the dimensionality of the feature output prior to classification. In this paper, we adopt principle components analysis (PCA) to reduce the dimensionality of the feature to 256.

Training and test images were carefully separated to ensure proper cross-validation. Groups of 30 training example image blocks and 30 testing example image blocks of each category were drawn from the full image set. In order to eliminate the influence of those rooftop like regions (such as parking lots) on the training model, we remove them from the training set. Feature vectors and PCA eigenvectors were computed from the training images, and the dimensionality reduced training data were used to train a multi-class support vector machine (SVM) using the Statistical Pattern Recognition Toolbox for Matlab [7]. In this paper, we only need to classify the segmented image region into rooftop region and non-rooftop region.

Following training, absolutely no changes to the feature representation or classifier were made. Each feature vector of the test image was transformed using the PCA projection matrix determined from the training images, and the trained SVM was used to report the predicted category of the test image. See Section 4 for the detailed experimental result.

3 Rooftop Refinement and 3D Building Modeling

In this section, we will utilize the information from 3D reconstruction to remove those non-rooftop regions and construct the 3D building model from image pairs.

Before 3D modeling of the buildings, we need to perform calibration of the intrinsic/extrinsic parameters of the camera as well as the extraction of the ground plane by simple human–computer interaction. The intrinsic parameters were obtained by popular Camera Calibration Toolbox for Matlab. The camera pose and ground-plane estimation between image pairs were conducted semi-automatically. We manually chose some salient corner correspondences in the image pairs and refine them by finding the sub-pixel corners like [2], and then estimate the parameters of camera pose and ground plane by the robust five-point algorithm [16] coupled with random sample consensus (RANSAC) [6]. Figure 2 shows the flow chart of rooftop refinement and 3D building modeling process.

3.1 Dominant Line Contour and Corner Extraction of the Candidate Rooftop Region

As we know, the appearance of the rooftop is usually uniform. Accordingly, the peripheral edges of the rooftop usually locate on or near the boundary of the segmented region (See Fig. 1).

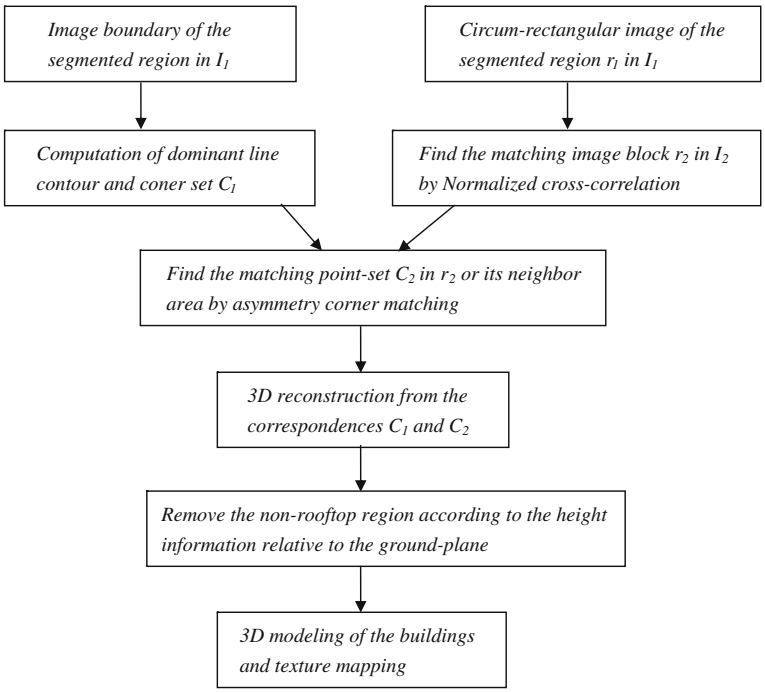

Fig. 2 Flow chart of the rooftop refinement and 3D building modeling

In order to obtain the main peripheral corner points of the rooftop, we extract the dominant line contour of the segmented region as a reference, wherein the binary-segmented area will be used as the input image. Note that, there are often some "spur" areas near the boundary of the segmented region. To prevent the "spur" from disturbing the extraction of dominant line contour, some morphological operations, like opening and closing operation, should be performed on the binary image. Next, an edge linking and line segment fitting algorithm [12] will be conducted on the smoothed contour of the region. After we obtain all the line edges along the contour we can select the dominant line contours among them. Finally, the main corner points of the boundary can be calculated by the intersections of the neighboring line segments. We describe the details in the Algorithm 1. Figure 3 illustrates an example of the pipeline.

Algorithm 1 Extraction of dominant line contour and corner point from segmented region

1. $I_1 \leftarrow$ binary image of the segmented region.
2. Perform open and close operations on I_1 to remove the "spur" area near the contour.
3. $I_2 \leftarrow$contour of I_1
4. $L_1 \leftarrow$line edge set of I_2 by edge linking and line segment fitting
5. $L_2 \leftarrow$Initialization of dominant line contour set
6. **for** each line segment $l_i \in L_1$ **do**
7. **if** the length of l_i is less than a threshold, **then continue**
8. **if** L_2 is empty, **then**
9. add l_i into L_2 and **continue**
10. **end if**
11. $\theta \leftarrow$angle between l_i and the latest line segment in L_2
12. **if** θ is great than a predefined threshold α, or less than $\pi - \alpha$
13. add l_i into L_2 and **continue**
14. **end if**
15. **if** length of l_i is great than that of the latest line segment in L_2
16. replace the latest line segment in L_2 by l_i
17. **end if**
18. **end for**
19. $C_1 \leftarrow$corner points by intersection of the neighboring line segments in L_2.

The output C_1 of Algorithm 1 is the coarse extraction of corners by the boundary of candidate rooftops, see the green crossing labeled in Fig. 3c. Note that since they are estimated from the boundary of the segmented region so might not be the real image corners of the candidate rooftop. We use C_1 as a reference value to detect the real image corners of the candidate rooftop: Firstly, for each line segment that connect two neighboring point in C_1, get the neighborhood image. Secondly, obtain the binary edge image by the Canny edge algorithm [3], and use an edge linking and line segment fitting algorithm [12] to produce a line edge set L. Thirdly, find the proper line segments among L with sufficient length and within close proximity to the reference line segment. There is a trade-off consideration between length and proximity that can be tuned by different weighting strategy. Finally, calculate the real corner points by intersecting the neighboring line segments. See Fig. 3d for an illustration of dominant line segments and corner points.

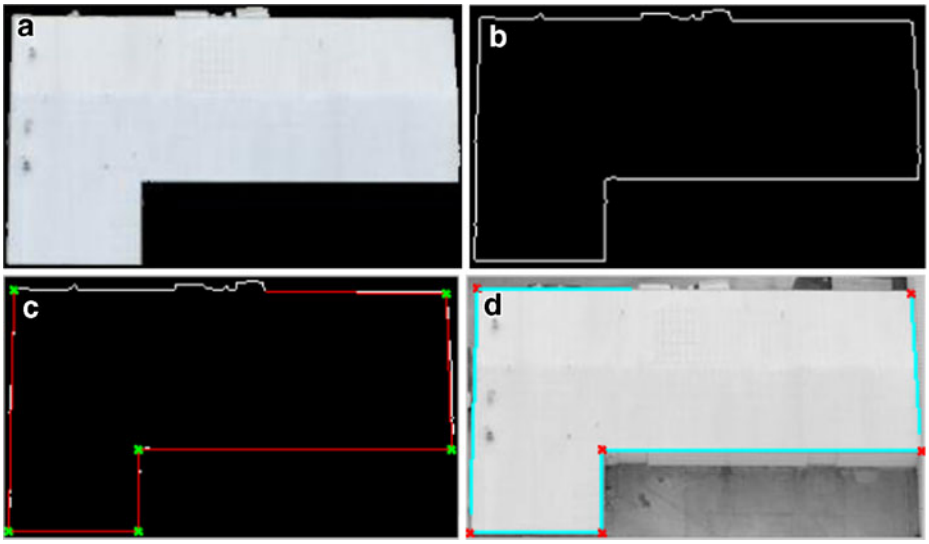

Fig. 3 Dominant line contour and main corner extraction of the candidate rooftop region. **a** segmented region; **b** boundary contour of the segmented region; **c** dominant line contour of the segmented region; **d** dominant line edges and corners of the rooftop from real image edges

Note that, any region whose dominant contours and corners are failed to be extracted will be discarded automatically and will be considered as a non-rooftop region.

3.2 Corner Matching of the Rooftop by Asymmetry Correlation

After obtaining the corner points of the candidate rooftop in Section 3.1, to compute the 3D information of the rooftop, we need to find the corner correspondences in another image. In this paper, we search corner correspondences by employing a hierarchy matching strategy coupled with asymmetry normalized cross-correlation.

In this paper, since the images are captured from the aircraft, which flies mainly in translational motion and stays at relatively stable altitude. So the variance of object rotation and scaling between adjacent image frames is relatively small. Thus we first conduct image matching of the rectangular candidate rooftop blocks between image pairs by normalized cross-correlation. Wherein, in the case of matching two similar image blocks, the epipolar geometry can be used to reject the false matches [8]. After this step, we can approximately obtain the correspondent image area of the candidate rooftop in another image (see Fig. 4 for an example).

After the rooftop region correspondences were found, we can further obtain the corner correspondences by corner matching based on the region correspondence. Traditional correlation based image matching methods [21] are limited to the short baseline case. Zhao et al. [22] proposed a corner matching method based on the rotation and scale invariant normalized cross-correlation, which is suitable for larger camera motion. This approach works well when the neighboring areas of the corner vary smoothly between image pairs. However in real applications, there often exists large variation of image content near the object boundary, which is called the

Fig. 4 **a** and **b** are a pair of rooftop image block correspondences. The appearance of the neighboring area of the image corners with a red square box centered at it is quite different

"covered/uncovered" phenomenon as can be seen in Fig. 4. The method of [22] will fail in this case.

In order to solve the problem of "covered/uncovered" in corner matching, we propose a novel asymmetry correlation strategy in this paper. The basic idea of the proposed approach is to maximize the usage of the stable area near the corner for matching, i.e. the corner point will not locate at the symmetric center of the matching block. Figure 5 shows an illustration of this method. This way, the image window for correlation matching will contain as much informative features as possible, and simultaneously avoiding the covered/uncovered area from interfering with the image matching.

For the sake of simplicity and efficiency, we defined only four types of "L" corner in this paper, and designed four corresponding types of matching window,

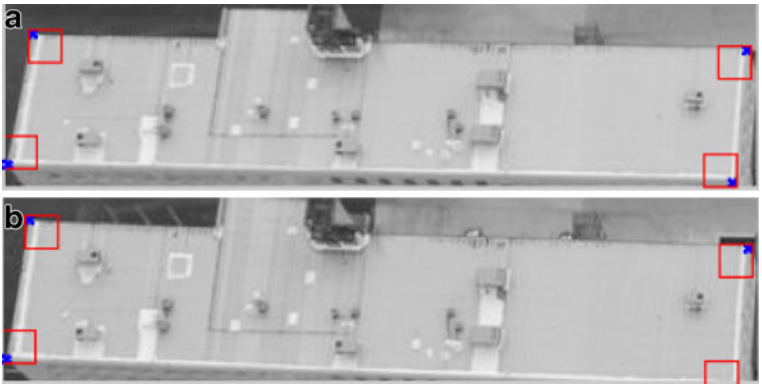

Fig. 5 An illustration of asymmetry correlation strategy for corner matching. **a** is the source image and **b** is the object image. The *blue* "x" represents corner point and *red square box* around it denotes the corresponding matching window

respectively (see Fig. 6). Given a "L" corner point c_1 in candidate rooftop region r_1 in image 1, we use a corresponding correlation matching window of size $(2n + 1) \times (2n + 1)$ centered with some eccentricity to this point, see the right column in Fig. 6. We then select a rectangular search area of size $(2m + 1) \times (2m + 1)$ $(m > n)$ around this point in the corresponding region r_2 in the second image, and perform a correlation operation on a given window between point c_1 in the first image and all possible points c_2 lying within the search area in the second image. Note that, the coordinates of the correlation matching result should take a translation to obtain the real location of the object corner, according to the eccentricity of the matching window. In the case of the corner not belonging to any of the above four types, a correlation matching window with the same size and centered at this point will be used instead.

In our implementation, $n = 8$ for the correlation window, and a translation of 1 or 2 pixels for the corner point away from the closest edge. For the search window, m is set to be about 3 or 4 times of n.

3.3 3D Reconstruction of the Rooftop and Building Rendering

After we obtain the corner correspondences of the candidate rooftop region from an image pairs, the 3D coordination of image corners in camera coordinates system can be calculated by triangulation method [8]. We determine whether a candidate region is a rooftop or not according to the height information of the corners relative to the ground plane. For convenience, we transform all the above 3D coordinates of the corners into the ground plane coordinate system. Given the rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ of the ground plane relative to one of the camera, a 3D point $\boldsymbol{p}$ can be transformed into $\boldsymbol{p}'$ in ground plane coordinate system by the following equation:

$$p' = R^T (p - t) \tag{1}$$

Any regions that are considered to be a rooftop should satisfy both of the following conditions: (1) the average height of the corners is greater than a threshold value; (2)

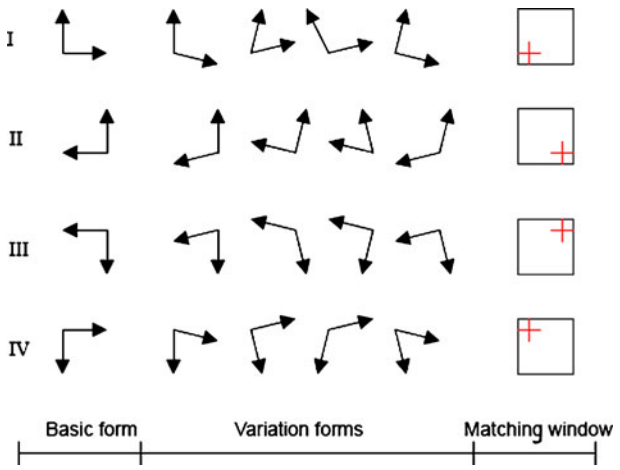

Fig. 6 Four types of "L" corner and part of their variation forms. The corresponding matching windows are listed in the *right column*, wherein the *red* "+" denotes the location of the image corner

at least half of the corners are higher than a specific threshold value. Regions that do not satisfy all of the above two conditions will be treated as non-rooftop regions and will be eliminated from further consideration.

As to the rendering of the building, since there is not enough information of the building's façade in the aerial images, we only perform texture mapping for the rooftop of the buildings. The polygonal rooftop surface is represented by triangular mesh to reduce geometric complexity and to tailor the model to the requirements of 3D computer graphics based visualization systems. To reduce noise it is recommended to first smooth the height information of the rooftop by averaging. Since the aim of this paper is to construct the model of the buildings, the image itself can be used as texture map of the ground plane.

4 Experimental Results

We have tested our system on different aerial images taken over the city of Columbia in Missouri, USA from hot air balloon. We report on experiments performed on an aerial image set collected at about 20 cm resolution over the urban and suburban area.

As a first step of our test, we start with an initial over-segmentation of all the object images by partitioning them into multiple homogeneous regions. Here we choose to employ Mean Shift [4] as the image segmentation algorithm, see Fig. 1 for an example of the segmentation result.

In the test of rooftop detection, we conduct experiments on ten groups of training and test samples. Figure 7 shows some image samples of rooftop block and Fig. 8 shows some image samples of non-rooftop block. When training the rooftop model, we extract the SIFT-like features of each image block at two scale according to the method described in Section 2.1, hereby we get a 512-vector. To speed up the computation and thus improve the classification performance, we reduced the

Fig. 7 Some image samples of rooftop block

Fig. 8 Some image samples of non-rooftop block

dimensionality of the feature vector to 256 by PCA. These dimensionality reduced feature vectors were then used to train a multi-class support vector machine (SVM) using the Statistical Pattern Recognition Toolbox for Matlab [7]. As we only need to classify the segmented image block into rooftop region and non-rooftop region, a two-class model is sufficient here.

The average recognition rate of rooftop regions is 88.8889%, whereas, the average recognition rate of non-rooftop regions is 72.8571%. The relative low recognition rate of non-rooftop region might be due to the fact that, regions such as parking lot may have very similar appearance or structure as the rooftop. Fortunately, these non-rooftop regions can be removed by the subsequent rooftop refinement procedure.

Since we conduct the classification computation only on the segmented image block, the efficiency of rooftop detection will depend on the size of the segmented image block. Since an over-segmentation on the whole scene is adopted, most of the image blocks are very small and can be classified in near real-time.

Our system identified most of the relative higher buildings and modeled them as multiple polygonal buildings. After the parameters of the camera pose and the ground plane were estimated, we are able to detect and build the complete model from image of size 1939×1296, like Fig. 1a, in about 5 minutes on a low-end PC with no user interaction at all.

Figure 9 briefly demonstrates the 3D model of buildings constructed from the scene in Fig. 1a without texture mapping. In order to make a comparison, the calibrated ground plane is also drawn in this figure, which is the quadrangle composed of points *1–2–3–4*.

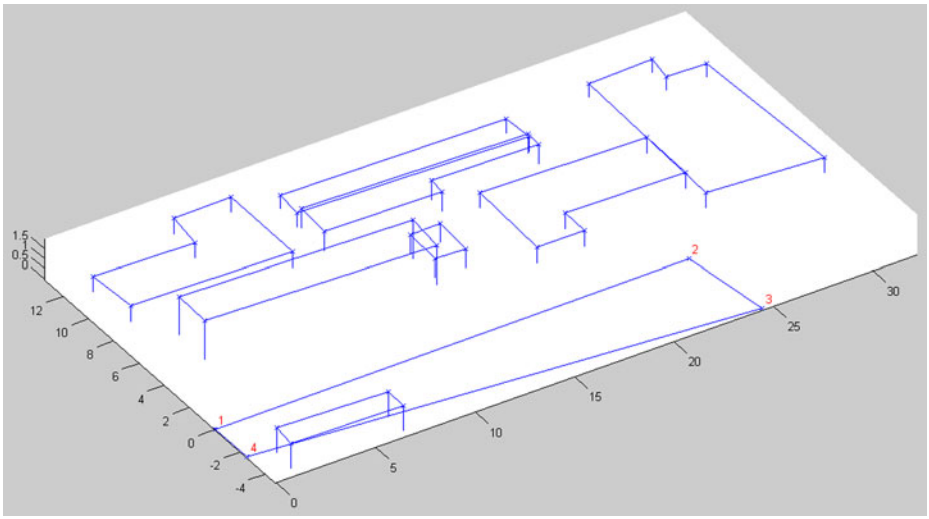

Fig. 9 A brief 3D model of buildings constructed from the scene in Fig. 1a without texture mapping. The quadrangle composed of points *1–2–3–4* is the ground plane

Figure 10 shows the 3D modeling of the full scene of Fig. 1a, with texture mapping based on the second image. The black strip around the 3D building model is to help make the building more salient in the scene. Note that, there is some distortion on the bottom of the ground texture, because this area is non-overlapped with the first image.

Fig. 10 The 3D model of the full scene in Fig. 1a, with texture mapping based on the second image

5 Conclusion

In this paper, a complete system for automatic rooftop detection and 3D building modeling from aerial images was presented. It integrates different components that gradually extract 2D and 3D information necessary to construct visual building model from aerial images. This system has the potential to be developed into a system taking aerial video of urban scene as input. The bridge links aerial video and our system are the object tracking and the selection of key frame. How to robust track the object rooftop region and select appropriate key frame in multiple aerial views will be our future efforts.

Acknowledgements This work is supported in part by the Leonard Wood Institute (#LWI 61031). The first author is also supported by the Young Faculty Research Grant of Shanghai Jiao Tong University and National Natural Science Foundation of China (60805018).

References

1. Baillard, C., Schmid, C., Zisserman, A., Fitzgibbon, A.: Automatic line matching and 3D reconstruction of buildings from multiple views. In: Proc. of ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS, vol. 32, Part 3-2W5, pp. 69–80 (1999)
2. Bouguet, J.Y.: Camera Calibration Toolbox for Matlab (2008). Available at: http://www.vision.caltech.edu/bouguetj/calib_doc/
3. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**, 679–714 (1986)
4. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)
5. Elgammal, A., Harwood, D., Davis, L.: Non-parametric model for background subtraction. In: Proc. of ECCV'00, pp. 751–767 (2000)
6. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
7. Franc, V., Hlavac, V.: Statistical Pattern Recognition Toolbox for Matlab (2008). Available at: http://cmp.felk.cvut.cz/cmp/software/stprtool/
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, UK (2000)
9. Hu, J., You, S., Neumann, U.: Integrating LiDAR, aerial image and ground images for complete urban building modeling. In: Proc. of 3DPVT, pp. 184–191 (2006)
10. Jaynes, C., Riseman, E., Hanson, A.: Recognition and reconstruction of buildings from multiple aerial images. Comput. Vis. Image Underst. **90**(1), 68–98 (2003)
11. Kim, Z.W., Nevatia, R.: Automatic description of complex buildings from multiple images. Comput. Vis. Image Underst. **96**(1), 60–95 (2004)
12. Kovesi, P.: Edge Linking and Line Segment Fitting (2007). Available at: http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **2**(60), 91–110 (2004)
14. Maloof, M.A., Langley, P., Binford, T.O., Nevatia, R., Sage, S.: Improved rooftop detection in aerial images with machine learning. Mach. Learn. **53**(1–2), 157–191 (2003)
15. Mayer, H.: Automatic object extraction from aerial imagery—a survey focusing on buildings. Comput. Vis. Image Underst. **74**(2), 138–149 (1999)
16. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 756–770 (2004)
17. Porway, J., Wang, K., Yao, B., Zhu, S.C.: A hierarchical and contextual model for aerial image understanding. In: Proc. CVPR'08, pp. 1–8. Anchorage, Alaska (2008)
18. Verma, V., Kumar, R., Hsu, S.: 3D building detection and modeling from aerial LIDAR data. In: Proc. CVPR'06, vol. 2, pp. 2213–2220 (2006)

19. Vestri, C., Devernay, F.: Using robust methods for automatic extraction of buildings. In: Proc. CVPR'01, vol. 1, pp. 133–138 (2001)
20. Wei, L., Prinet, V.: Building detection from high-resolution satellite image using probability model, geoscience and remote sensing symposium, In: Proc. of IGARSS'05, pp. 25–29 (2005)
21. Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.T.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif. Intell. J. **78**(1–2), 87–119 (1995)
22. Zhao, F., Huang, Q., Gao, W.: Image matching by normalized cross-correlation. In: Proc. of ICASSP'06, vol. 2, pp. 729–732. Toulouse, France (2006)