2016 14th International Conference on Control, Automation, Robotics & Vision
Phuket, Thailand, 13-15th November 2016 (ICARCV 2016)

View Planning for 3D Shape Reconstruction of Buildings with Unmanned Aerial Vehicles

Wei Jing*, Joseph Polden†‡, Pey Yuen Tao†‡,Wei Lin†‡, Kenji Shimada*
*Carnegie Mellon University, Pittsburgh, PA, 15213, USA
wj@andrew.cmu.edu, shimada@cmu.edu
†Mechatronics Group, SIMTech, 71 Nanyang Drive, 638075, Singapore
{poldenjw, pytao,wlin}@simtech.a-star.edu.sg
‡SIMTech-NUS Joint Laboratory on Industrial Robotics, NUS, 117575, Singapore

***Abstract*—This paper presents a novel view planning method to generate suitable viewpoints for the reconstruction of the 3D shape of buildings, based on publicly available 2D map data. The proposed method first makes use of 2D map data, along with estimated height information, to generate a rough 3D model of the target building. Randomized sampling procedures are then employed to generate a set of initial candidate viewpoints for the reconstruction process. The most suitable viewpoints are selected from the candidate viewpoint set by first formulating a modified Set Covering Problem (SCP) which considers image registration constraints, as well as uncertainties present in the rough 3D model. A neighborhood greedy search algorithm is proposed to solve this SCP problem and select a series of individual viewpoints deemed most suitable for the 3D reconstruction task. The paper concludes with both computational and real-world field tests to demonstrate the overall effectiveness of the proposed method.**

Keywords: View Planning, UAV, Optimization, 3D Reconstruction

I. Introduction

Unmanned Aerial Vehicles (UAVs) have undergone rapid development in recent years. Most commercially available UAVs are now equipped with both high resolution camera and GPS systems, which makes them suitable for use in 3D shape reconstruction tasks, particularly for large scale outdoor scenes [1][2]. For these reconstruction tasks, the first step is to identify a set of camera viewpoints which meet the requirements of the application; the process of finding these viewpoints is called view planning [3][4][5][6]. Ideally, the view planning process involves finding the smallest possible number of individual viewpoints which best cover the target object.

Significant amounts of work in view planning spanning many different applications have been published in recent years[1][6][7]. From these works, the approaches for view planning prolems generally fall into two categories[4]: model-based approaches and non-model-based approaches. These definitions depend on the availability of a geometric model of the target object prior to the planning process. If the geometric model of the target object is known prior to the view planning process, model-based approaches can be applied; otherwise a non-model-based approach must be used.

For model-based view planning, "generate-and-test" approaches, summarized in [4], are popular due to their simplicity and effectiveness; early work with these methods focused mainly on view planning for small objects such as mechanical parts. These methods generate a large set of initial candidate viewpoints, from which a subset of suitable viewpoints is selected through combinatorial optimization [4][8]. Candidate viewpoints are typically generated by offsetting from the target object's surface [7], or on the surface of the sphere which encapsulates it [9]. In more recent years, model-based view planning has been incorporated into UAV applications where large target objects, such as buildings or outdoor environments, need to be inspected or reconstructed via aerial photography. Recent examples are presented in [1] and [10]; however, these examples assume that detailed 3D polygonal models of the target objects are already available prior to the planning procedure.

On the other hand, researchers have also proposed a number of non-model-based view planning methods for 3D reconstruction tasks [11][12][13]. Most of the approaches use Next-Best-View (NBV) methods to plan viewpoints online without prior geometric information of the target object. In general, these NBV methods iteratively identify the next best viewpoint online based on a cost-planning function and information from previously planned viewpoints. They also make use of partial geometric information of the target object, reconstructed from planned viewpoints, to plan future sensor placements. However, these methods are mainly applied to small scale reconstruction problems and may suffer a high computational cost for large scale or detailed tasks.

In this paper, we present a novel view planning method to tackle the planning problem for 3D building reconstruction when the 2D map data of the target building is available. The proposed method uses the preliminary map data to assist in finding a series of suitable viewpoints for 3D reconstruction tasks for large scale outdoor building structures. The main contributions presented in this paper are:

- A novel view planning framework that identifies the viewpoints for 3D building reconstruction with publicly available 2D map data and estimated building height information.
- A formulation of the modified Set Covering Problem (SCP), which considers image registration constraints as well as uncertainties present in the generated rough model.

978-1-5090-3549-6/16/$31.00 ©2016 IEEE

- An algorithm to solve the formulated optimization problem by performing a greedy neighbourhood search.

II. Problem Formulation and Proposed Framework

A. Problem Formulation

Unlike model-based and non-model-based view planning methods presented in introduction, the view planning application discussed in this paper is different in that only a limited amount of geometric information of the target object is available beforehand. With this in mind, we can describe the view planning problem presented in this paper as a "semi-model-based" view planning problem. This prior information comes in the form of:

- 2D longitude and latitude information of each perimeter vertex of the target buildings,
- a rough estimation of building height based on the number of stories in the building,
- the specifications of the UAV's on-board camera sensor, and
- the requirements from the reconstruction application (e.g. the maximum viewing angle, the coverage frequency, etc.).

The view planning problem discussed in this paper is now defined as finding the minimum number of viewpoints required for the 3D shape reconstruction of outdoor buildings, based on prior information.

The view planning method presented in this paper involves two major steps, as outlined in Fig. 1. Firstly, a rough model of the target building is generated from freely available online 2D map data. The longitude and latitude coordinates of the 2D footprint are mapped to Cartesian coordinates in Euclidean space, and a rough 3D model of the target building is then completed by extruding the footprint in Cartesian coordinates by the building's estimated height. This rough 3D model of the target building is used as an input to the second phase. A "generate-and-test" view planning approach is used to find suitable viewpoints whilst adhering to the requirements of the application and sensor specifications.

B. Preparation of Target Building Model

The first step of the proposed method is to prepare a rough polygonal model of the target building. This is done with 2D map data of the building and its estimated height, as well as a subdivision process on the surface of the generated model.

1) Generate 3D Model with 2D Map Data and Estimated Height: The rough 3D polygonal model is generated in three steps using 2D map data and an estimated building height. First, the GPS coordinate information of each perimeter vertex of the target building are extracted from an online map. Then, the coordinates of each vertex is transformed from longitude and latitude to Cartesian coordinates using the Haversine formula [14] shown in Eq. (1). The conversion is performed by selecting an arbitrary vertex, v_o, as the origin. x/y positions in Cartesian coordinates for the remaining vertices are then calculated as the distance along the longitude/latitude directions between themselves and v_o using Eq. (1). In the third step, using an estimated height for each vertex, the 3D rough model is generated via extrusion, as shown in Fig. 2.

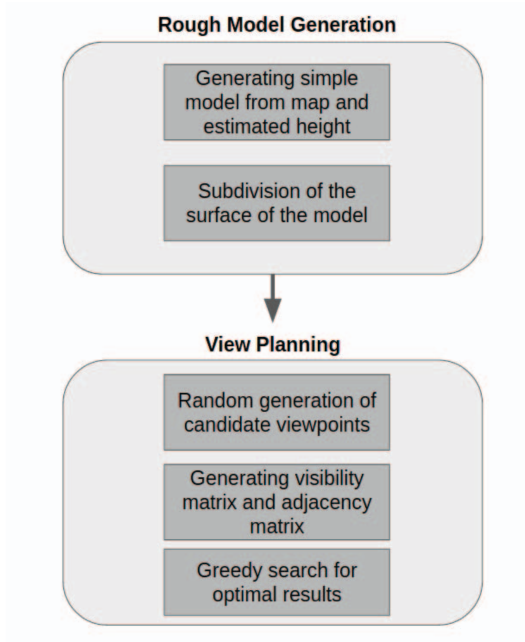


Fig. 1: The outline of proposed method

$$
\begin{aligned}
d =& 2r \arcsin\Big[\sin^2\Big(\frac{\alpha_2-\alpha_1}{2}\Big) \\
&+ \cos(\gamma_1)\cos(\gamma_2)\sin^2\Big(\frac{\gamma_2-\gamma_1}{2}\Big)\Big]^{\frac{1}{2}}
\end{aligned} \tag{1}
$$

where r is the radius of the earth. α_1, α_2 and γ_1, γ_2 represent the longitude and latitude information of two given points.

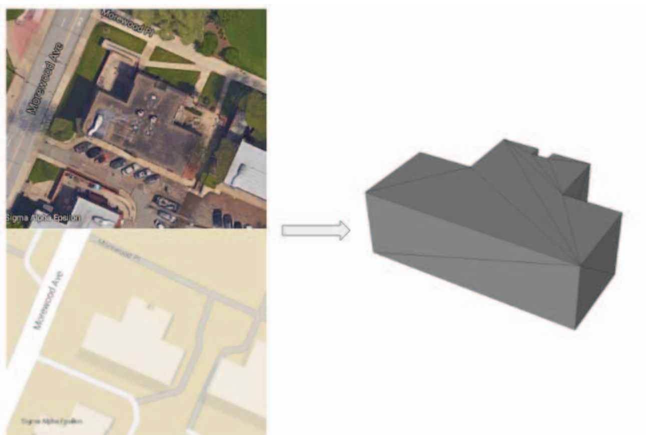
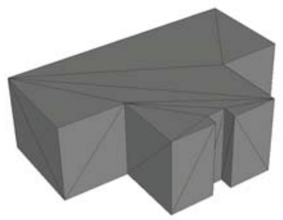
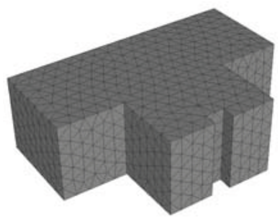


Fig. 2: 2D map to 3D model; the two pictures on the left are the satellite image and 2D map data; the picture on the right are the 3D polygonal model generated from map data.

2) Subdivision of the Model Surface: At this stage of the view planning process, the 3D polygonal model generated in the previous step is not suitable due to the existence of large variations in patch size and also low quality properties of its shape (e.g. sharp, narrow triangles present in the surface mesh). To overcome this, the Bubble Mesh [15] method is used to subdivide the model's surface patches into a well-shaped triangular mesh that preserves the original shape well, and also maintains patch uniformity. The results of this subdivision process on an example target building model are shown in Fig. 3.



(a) Original model with 34 triangles



(b) The model with 1187 triangles after subdivision using Bubble Mesh[15]

Fig. 3: Subdivision of the target building

C. View Planning Based on Rough Model

The view planning method proposed in this paper is adapted from the "generate-and-test" framework described in previous literature [4]. First, a large sample of candidate viewpoints is generated around the target building model. To solve the view planning problem, a smaller subset of viewpoints is selected from this candidate set via combinatorial optimization, based on a series of problem-specific requirements.

1) Generate Candidate Viewpoints: In the candidate viewpoints generation step, a set of redundant viewpoints is generated for use in the later selection process. In this paper, candidate viewpoints are randomly generated around the target building within a distance smaller than the maximum viewing range of the UAV's camera. The candidate viewpoints generation method is adopted from previous work[16]. Once the generation of the candidate set is completed, visibility testing is performed between each surface patch of the target building model and each member of this candidate viewpoint set. Additionally, each pair of viewpoints in the candidate set is evaluated in terms of image registration feasibility constraints.

2) Generate Visibility Matrix and Adjacency Matrix: Before the viewpoint selection process can commence, a preposssessing step which introduces two matrices is performed. The visibility matrix A_{vm} is a 2D binary matrix which provides visibility information between each surface patch of the target model and viewpoints belonging to the candidate set. The adjacency matrix A_{am} is a 2D binary matrix that is used to measure connectivity information between viewpoints. The purpose of A_{am} is to provide a measure which will ensure the success of the final image registration process.

The visibility matrix is used as the basis for the SCP to find the most suitable set of viewpoints to solve the view planning problem. Visibility between a given viewpoint and a surface patch is evaluated by the following conditions:

- The surface patch must be located within the camera's Field-Of-View (FOV) (checked using a perspective projective camera model[17] to project surface patches onto the image plane of the camera) and Field-Of-Depth (FOD).
- The viewing angle must be less than a specified maximum.
- There must be a clear line of sight, i.e no physical blockages or occlusions between the viewpoint and surface patch (this occlusion condition is checked via a fast ray-triangle intersection algorithm[18]).

The resultant visibility matrix A_{vm} is a $N \times M$ matrix, where M and N represent the number of viewpoints and number of surface patches, respectively. The Boolean information encoded in the matrix represents the visibility between a given surface patch and viewpoint pair.

The adjacency matrix encodes connectivity information between viewpoints. Connectivity is evaluated as the feasibility of an image registration process between two viewpoints. In the $M \times M$ adjacency matrix A_{am}, two viewpoints are considered connected if the following two conditions are satisfied:

- The distance between two viewpoints must be smaller than D_{th}.
- There is at least λ_{min} overlapped viewing area shared by the two viewpoints.

Once the visibility matrix and the adjacency matrix are generated, the selection process via combinatorial optimization can commence.

3) Modified Set Covering Problem with Constraints: In order to address uncertainties present in the rough building model, and issues relating to the image registration process, the view planning problem discussed in this paper is formulated as a modified SCP with additional constraints. In this circumstance, full coverage is not required. Instead, a certain coverage ratio, δ_d, is introduced, along with an additional constraint to ensure the resultant viewpoint set is fully connected. This ensures that the image registration process is performed to a suitably high quality. The SCP in this paper is formulated to find the least number of viewpoints that cover every surface patch at least γ times:

$$
\begin{aligned}
& \min \sum_{i=1}^{n} x_i, \quad \text{where } x_i \in \{0,1\} \\
\text{s.t.} \quad & \sum (A_{vm} \cdot \mathbf{x} \geq \gamma) \geq \delta_d \\
& \text{There exists a path between } (x_i, x_j) \forall x_i > 0, x_j > 0,
\end{aligned}
$$

where x_i corresponds to the i^{th} viewpoint; $x_i = 1$ denotes that the viewpoint is selected, whilst $x_i = 0$ infers that it is not. γ represents the required coverage frequency (the minimum number of times each surface patch must be captured from different viewpoints) and δ_d is the desired coverage ratio. The first constraint presented in the formulation is to enforce δ_d of the surface of the target building is covered at least γ times.

The second constraint ensures the resultant viewpoint set is fully connected and the 3D multi-view reconstruction process will then generate a full single model of the target object, as opposed to several separate partial models.

The combinatorial optimization problem formulated in this paper is solved with the neighbourhood greedy search algorithm proposed in Alg. 1. The method continuously searches for the viewpoint with maximum coverage in the neighbourhood of the solution set. The neighbourhood is defined as the viewpoints connected to the current solution set. If such a viewpoint is found, it is added to the solution set. This process continues until the specified coverage requirements are met.

Algorithm 1 Neighborhood greedy search algorithm for modified SCP with constraints

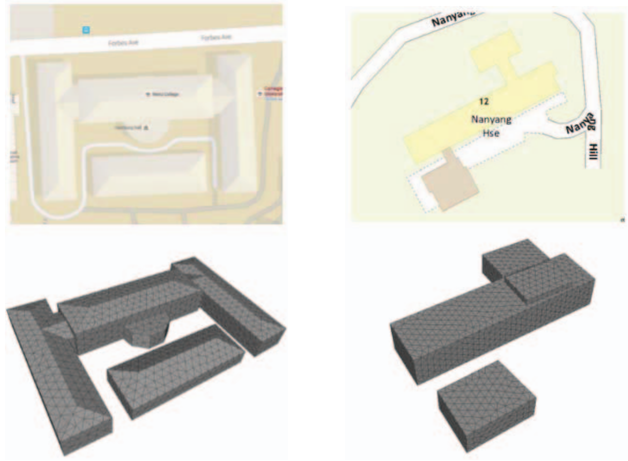
Input: The set of triangular surface patches, P; the set of candidate viewpoints, V; the visibility matrix, A_{vm}; the adjacency matrix, A_{am}; the required coverage frequency, γ; and the desired coverage ratio, δ_d.

Output: The resultant coverage ratio, δ; the number of resultant viewpoints, n_{res}; and the resultant viewpoints set, V_{res}.

$V_{res} \leftarrow \emptyset$
$N_{numOfTriangles} = \text{sizeof}(P)$
$P_{uncover} \leftarrow \text{findUncovered}(A_{vm}, \gamma)$
$P \leftarrow \text{delete}(P,\ P_{uncover})$
$V_{res} \leftarrow \text{append}(V_{res},\ \text{random}(V))$
while $P \neq \emptyset$ **and** $\delta < \delta_d$ **do**
    $V_{neighbour} \leftarrow \text{Neighborhood}(V_{res}, A_{am})$
    $v \leftarrow \text{maxCover}(V_{neighbour}, P, A_{vm})$
    $V_{res} \leftarrow \text{append}(V_{res}, v)$
    $V \leftarrow \text{delete}(V,\ v)$
    for each $p_j \in P$ **do**
        if $\text{count}(p_j, V_{res}, A_{vm}) > \gamma$ **then**
            $P \leftarrow \text{delete}(P,\ p_j)$
        end if
    end for
    $\delta = \text{sizeof}(P)/N_{numOfTriangles}$
end while
$n_{res} = \text{sizeof}(V_{res})$
return δ, n_{res}, V_{res}

III. Experiment and Discussion

To validate the view planning method proposed in this paper, we have conducted two computational tests and one field test. These tests are performed on two separate target buildings. The first building, the Hamburg Hall, shown in Fig. 4-(a), is located on the Carnegie Mellon University (CMU) campus in USA. The second target building, the Nanyang House, shown in Fig. 4-(b), consists of a large three-storey building with two peripheral two-storey buildings, and is located on the Nanyang Technological University (NTU) campus in Singapore. Computational tests are performed by applying the view planning method outlined in II-A to rough 3D models generated for these two building structures. A field test is then conducted on the Nanyang House structure as well, where a UAV with on-board camera equipment is used to capture a single 2D image at each planed viewpoint. This set of images is then used to reconstruct a more detailed 3D model of the target building.



(a) Hamburg Hall in CMU (b) Nanyang House in NTU

Fig. 4: Target buildings

A. Setup

The UAV used in this experiment is a DJI Phantom 3 [19]. The intrinsic parameters of the on-board camera are derived via standard chessboard calibration methods[20]. Eleven pictures of the 9×7 chessboard are taken at different poses; the resultant calibrated intrinsic matrix $\mathbf{K}$ is given as:

$$K = \begin{bmatrix} 2337.76 & 0 & 1991.83 \\ 0 & 2346.33 & 1466.06 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

The remaining view planning parameters used in this paper are listed in Table. I.

TABLE I: View planning parameters

Viewing Angle	$75°$
FOD (meters)	$(1, 50)$
Required Coverage Frequency	$\gamma = 2, 3, 4$
Required Coverage Ratio	$\delta_d = 99.5\%$
Minimum Overlapping	$\lambda_{min} = 5$ patches
Maximum Connected Distance	10 m
Minimum Viewpoint Height	10 m

B. View Planning

The rough 3D polygonal model (shown in Fig. 4), the camera intrinsic matrix (shown in Eq. (2)), and the parameters in Table. I are used as input for the view planning method outlined in Fig 1. For Hamburg Hall, a total number of $2,858$ candidate viewpoints are generated; and for Nanyang House, a total number of $2,949$ candidate viewpoints are produced from our method. The results with different required coverage frequencies are shown in Table. II. The 3D visualization of the planned viewpoints and the target buildings are shown in Fig. 5 and 6.

TABLE II: Comparison of Results with Different Required Coverage Frequency

	Number of viewpoints required		
	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Hamburg Hall Building	33	52	68
Nanyang House Building	26	41	55

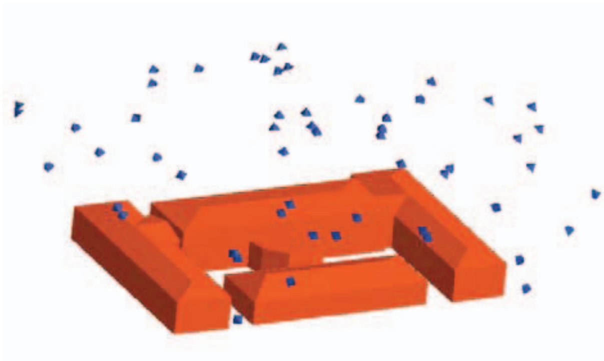


Fig. 5: 3D visualization of the Hamburg Hall building and the planned viewpoints ($\gamma = 3$)

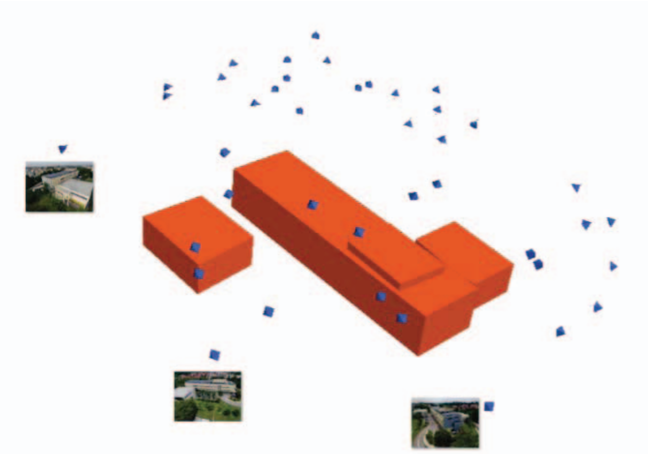


Fig. 6: 3D visualization of the Nanyang House building, the planned viewpoints and examples of pictures taken at some planned viewpoints in a later field experiment ($\gamma = 3$)

C. Field Experiment

A field experiment was conducted on the Nanyang House structure by flying the UAV to capture a single 2D image at each planned viewpoint, as shown in Fig. 6. The resultant collection of images, shown in Fig 7, is then used to reconstruct a 3D model of the target building. The 3D reconstruction is performed using the Structure From Motion (SFM) workflow. In this paper, we use VisualSFM [21][22] for feature finding, feature matching, image registration and camera pose estimation. With the SIFT feature descriptor used in this paper, 39 out of 41 planned viewpoints are successfully registered for 3D multi-view dense reconstruction. Two different free software packages, CMP-MVS [23] and Multi-View Environment

Fig. 7: The pictures taken at planned viewpoints

(MVE) [24], are used for the dense reconstruction and surface reconstruction tasks with these images. The results of these reconstructions are presented in Fig. 8 and 9.

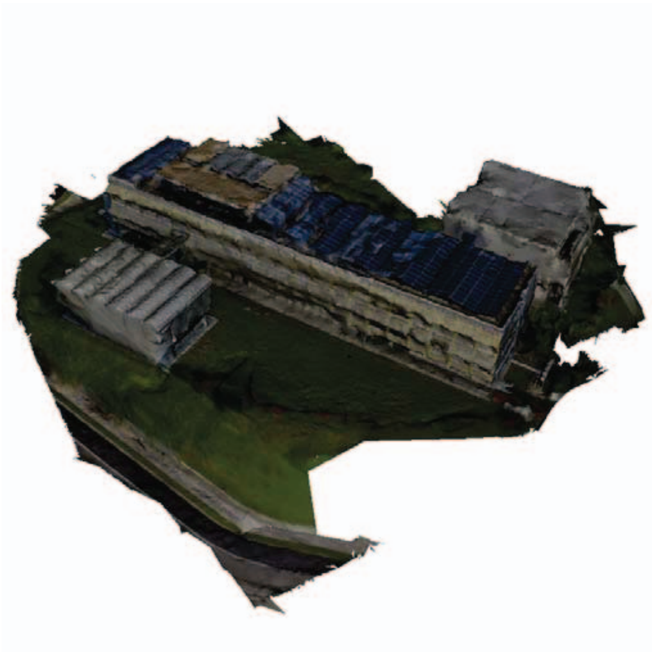


Fig. 8: The 3D reconstruction result by CMP-MVS[23]

D. Discussion

The experiments conducted in this paper indicate that the proposed view planning method is able to identify suitable viewpoints for large-scale outdoor 3D building reconstruction tasks. The 3D models of the target buildings are successfully reconstructed, and show surface details of the target buildings. It is worth noting that the quality of the reconstructed model could be further improved by adjusting certain view planning parameters. For example, a smaller viewing angle, larger required coverage frequency, or smaller maximum FOD

Fig. 9: The 3D reconstruction result by MVE [24]

may lead to reconstruction results with higher surface detail. However, one must acknowledge that the natural downside of this increment in detail is that more individual viewpoints would be generated in the solution, an example of this is shown in Table. II.

IV. Conclusion

In this paper, we have proposed a view planning method that makes use of publicly available 2D map data to identify a set of suitable viewpoints for 3D shape reconstruction tasks of outdoor buildings. The effectiveness of the proposed method in a real-world situation has been demonstrated via a field experiment, where a camera-equipped UAV is used to capture a series of images from a number of individual viewpoints specified by our viewpoint planning method. These images are used to successfully reconstruct a detailed 3D model of the target building. Future work will focus on modeling uncertainties in the UAV's localization, and considering it for view planning in order achieve better and more reliable results.

Acknowledgment

The authors acknowledge the support from the Collaborative Research Project under the SIMTech-NUS Joint Laboratory on Industrial Robotics, Ref: U13-R-024AU and support from the SERC Industrial Robotics Program Work Package 7 and 8, project number: 1225100007 and 122510008.

References

[1] Andreas Bircher, Kostas Alexis, Michael Burri, Philipp Oettershagen, Sammy Omari, Thomas Mantel, and Roland Siegwart. Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics. In *IEEE International Conference on Robotics and Automation*, pages 6423–6430. IEEE, 2015.

[2] Korbinian Schmid, Heiko Hirschmüller, Andreas Dömel, Iris Grixa, Michael Suppa, and Gerd Hirzinger. View planning for multi-view stereo 3d reconstruction using an autonomous multicopter. *Journal of Intelligent & Robotic Systems*, 65(1-4):309–323, 2012.

[3] Konstantinos A Tarabanis, Roger Y Tsai, and Peter K Allen. The mvp sensor planning system for robotic vision tasks. *IEEE Transactions on Robotics and Automation*, 11(1):72–85, 1995.

[4] William Scott, Gerhard Roth, and Jean-François Rivest. View planning for automated 3d object reconstruction inspection. *ACM Computing Surveys*, 35(1), 2003.

[5] Shengyong Chen, Youfu Li, and Ngai Ming Kwok. Active vision in robotic systems: A survey of recent developments. *The International Journal of Robotics Research*, 30(11):1343–1377, 2011.

[6] Brendan Englot and Franz S Hover. Three-dimensional coverage planning for an underwater inspection robot. *The International Journal of Robotics Research*, 32(9-10):1048–1073, 2013.

[7] William R Scott. Model-based view planning. *Machine Vision and Applications*, 20(1):47–69, 2009.

[8] Glenn H Tarbox and Susan N Gottschlich. Planning for complete sensor coverage in inspection. *Computer Vision and Image Understanding*, 61(1):84–111, 1995.

[9] SY Chen and YF Li. Automatic sensor placement for model-based robot vision. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):393–408, 2004.

[10] L Geng, YF Zhang, JJ Wang, Jerry YH Fuh, and SH Teo. Mission planning of autonomous uavs for urban surveillance with evolutionary algorithms. In *IEEE International Conference on Control and Automation*, pages 828–833. IEEE, 2013.

[11] Christian Potthast and Gaurav S Sukhatme. Next best view estimation with eye in hand camera. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[12] J Irving Vasquez-Gomez, L Enrique Sucar, and Rafael Murrieta-Cid. View planning for 3d object reconstruction with a mobile manipulator robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4227–4233. IEEE, 2014.

[13] YF Li and ZG Liu. Information entropy-based viewpoint planning for 3-d object reconstruction. *IEEE Transactions on Robotics*, 21(3):324–337, 2005.

[14] BP Shumaker and RW Sinnott. Astronomical computing: 1. computing under the open sky. 2. virtues of the haversine. *Sky and telescope*, 68:158–159, 1984.

[15] Kenji Shimada and David C Gossard. Bubble mesh: automated triangular meshing of non-manifold geometry by sphere packing. In *Proceedings of the third ACM symposium on Solid modeling and applications*, pages 409–419. ACM, 1995.

[16] Wei Jing, Joseph Polden, Wei Lin, and Kenji Shimada. Sampling-based view planning for 3d visual coverage task with unmanned aerial vehicle. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2016.

[17] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[18] Tomas Möller and Ben Trumbore. Fast, minimum storage ray-triangle intersection. *Journal of graphics tools*, 2(1):21–28, 1997.

[19] DJI. DJI phantom 3 advanced specs. *URL http://www.dji.com/product/phantom-3-adv/info#specs*, 2016.

[20] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[21] Changchan Wu. VisualSFM: A visual structure from motion system. *URL http://ccwu. me/vsfm/*, 2011.

[22] Changchang Wu. SiftGPU: A gpu implementation of scale invariant feature transform. *URL http://cs. unc. edu/˜ ccwu/siftgpu*, 2011.

[23] Michal Jancosek and Tomás Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3121–3128. IEEE, 2011.

[24] Simon Fuhrmann, Fabian Langguth, and Michael Goesele. Mve-a multiview reconstruction environment. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (GCH)*, volume 6, page 8. Citeseer, 2014.