

EFFICIENT FLIGHT PLANNING FOR BUILDING FAÇADE 3D RECONSTRUCTION

H.K. Palanirajan^{a*}, B. Alsadik^b, F. Nex^a, S. Oude Elberink^a

^a Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands -
h.k.palanirajan@student.utwente.nl, f.nex@utwente.nl, s.j.oudeelberink@utwente.nl

^b Center of Applied Remote Sensing and GIS, University of Zakho, Kurdistan Region of Iraq- b.alsadik@visitor.uoz.edu.krd

Commission I, ICWG I/II

KEY WORDS: flight plan, UAV, optimization, façade, LoD3 generation, camera network design

ABSTRACT:

Three-dimensional (3D) building model is gaining more scientific attention in recent times due to its application in various fields such as vehicle autonomous navigation, urban planning, heritage building documentation, gaming visualisation and tourism. The quality of the Level of Detail (LoD) of building models relies on the high-resolution data sets obtained for the building. As an alternative to laser scanners, Unmanned Aerial Vehicles (UAV) are efficient in collecting good quality images and generate reliable LoD3 of buildings (i.e. to model both roof and facades of a building) at comparatively lower cost and time. However, the complete collection of images on building facades is usually performed by manual flights along the different façade to assure a homogenous image coverage with the same resolution on each element: no offline autonomous procedure to define the main façade planes and acquire complete image sets independent of UAV platform have been developed yet. This paper proposes a novel methodology to generate the flight plan in correspondence of building facades. The Digital Surface Model (DSM) obtained from an initial nadir flight is used as an input to identify the target building and plan the image acquisition around it. The optimised flight plan ensures complete coverage of the building with a minimum number of images. The coordinates as well as the attitude of each planned image can be finally loaded on a UAV to perform the flight. In order to validate the proposed methodology, some tests performed on synthetic buildings of growing complexity and very different shapes are presented.

1. INTRODUCTION

Since the introduction of digital maps, the levels of detail incorporated in it have consistently increased. 3D city models have become a hot topic of research in the last decade as it finds application in autonomous vehicle navigation, urban planning, heritage building documentation, gaming visualisation, civil engineering, disaster management, rooftop solar potential estimation, tourism, augmented reality. This field has been a common area of interest among Computer visions community and Photogrammetric experts for a long time, but a common standard framework was not specified in this field (Forlani, Roncella, & Nardinocchi, 2015). Finally, Open Geospatial Consortium (OGC, 2012) came up with the Level of Detail (LoD) scheme ranging from 0 to 4, which is widely followed in 3D city models. In recent times, the reconstruction of building structures with façade information became a fascinating field of study for survey services (Cefalu, Haala, Schmohl, Neumann, & Genz, 2017). 3D reconstruction of building models has always depended on the point cloud data using ALS. However, ALS cannot be useful to model the façade information (LoD3) of buildings due to the high altitude at which the data is collected. Besides, the survey undertaken using ALS seems unpractical because of the expensive instruments and staffing a pilot to fly a crewed aeroplane.

To reconstruct facades, Terrestrial and Mobile Laser scanners (TLS & MLS) can be useful. But they are expensive and tedious to carry around on field. Meanwhile, Unmanned Aerial Vehicles (UAVs) become handy in this situation as it can be flown at required proximity to the object of interest such as a

building. The images acquired can be used to generate reliable LoD3 of buildings at comparatively lower cost and time. The fact that sensors carried by UAV can be customised based on the needs of mission provides a strategic advantage over other platforms of remote sensing. The low cost and non-intrusive nature of UAV have made it a viable alternative solution for researchers to monitor and document building sites. At present advances are made on lightweight LiDAR systems to be mounted on UAVs, but they are still expensive. However, by selecting appropriate hardware and flight parameters using photogrammetry, same quality of depth information can be obtained using commercial drones that are ten times cheaper.

The quality of the 3D reconstruction heavily depends on the completeness and accuracy of dense point cloud generated (Mostegel, 2011). The point cloud comparison between Close Range Photogrammetry (CRP) and TLS showed that the difference in accuracy is considerably small (Harwin & Lucieer, 2012; Incekara & Seker, 2018; James & Quinton, 2014; Murtiyoso & Grussenmeyer, 2017). This is true when images are captured with ambient illumination conditions, and the surface is textured. CRP fails considerably in smooth surfaces due to lack of identifiable tie points and high reflectance/refractance.

The reliability of 3D models developed from UAV images depends on the manual operations, the expertise of a pilot and in some cases grid-based automation. In the latter, the region is captured in equally spaced viewpoints in the form of grid which doesn't consider the shape of the target object. In most cases, the flying is manual or in regular grids when the surface is

* Corresponding author

regular. And the process is performed piece by piece rather than the whole building in one go. Therefore, expertise of pilot plays major role in determining the homogeneity of the data for 3D reconstruction. This approach can result in gaps in the data coverage leading to inconsistency in the model (Martin, Rojas, Franke & Hedengren, 2016). Therefore, homogenous image coverage with same resolution on each element of façade need to be ensured. To acquire complete image sets without any discontinuity in the surface: an autonomous procedure is required to define the principal façade planes and completeness of the data. On the other hand, collecting redundant data over the same region increases the total flight time as well as the computation time during the image matching process.

One of the classic problems in robotics is to identify optimal viewpoints for reliable 3D modelling, i.e. identifying next-best-view (NBV) in reference to the previous image that will have sufficient overlap. Technically, the number of such views can be infinite (Chen, Li, & Kwok, 2011). To solve this NBV problem, many heuristic approaches relying on dense point clouds are available. However, they are not highly suitable for complex scenes like an urban area where actual 3D structures are unknown (Huang, Zou, Vaughan, & Tan, 2017). Hence, it becomes necessary to determine a minimum number of images that are sufficient to reconstruct the complete 3D model of facades at a predefined resolution to assure maximum coverage and overlap in each viewpoint. These are the optimal viewpoints which constitute the optimal flight plan necessary to realise coherent data.

Thus, the motivation of the proposed research was to develop an algorithm that can produce an optimal flight plan for acquiring images to generate a reliable and complete point cloud for 3D building model (LOD3). Procedures like Next-Best-View (NBV) are computationally expensive and require an onboard processing unit for determining the new viewpoint which is not a feasible commercial solution. These procedures collect an enormous amount of redundant data. On the other hand, the proposed method was independent of the drone typology making it a flexible solution for wider range of cheaper drones. For an unknown area, the rough DSM generated from nadir images was used as the initial input. From the DSM, the target building was identified, and the footprint was extracted. Based on the building footprint, the algorithm computed the location of viewpoints and the camera sensor orientation for the flight mission. The optimal flight planning was done based on user requirements of Ground Sampling Distance (GSD), overlap percentage, Base to Depth ratio to assure theoretical accuracy of the dataset. The initial camera network was planned to be redundant with very short baselines. The planned images were analysed, and redundant cameras which do not significantly improve the accuracy and completeness were filtered. By the iterative process of visibility testing and filtering based on coverage, homogeneity of each point in the scene (building facade) was ensured. This algorithm helped to reduce the number of viewpoints without compromising the quality of dataset thus saving time and cost during data acquisition phase. The output of the algorithm is the position (X, Y, Z) and attitude of the camera (Ω , Φ , K) for all the images to be acquired.

2. METHODOLOGY

The steps involved in the process are: Boundary extraction, Façade definition, Dense camera network design, Visibility analysis, Coverage based filtering.

2.1 Building extraction from DSM

The process aimed to identify the building footprint to be used in the camera network design for traversing along the boundary of the building. The initial step was to identify the target building in the DSM. To segment the target building from DSM, a seed pixel on the roof surface of the building was selected by the user in an interactive approach. With the height data of seed pixel as mean value of the segment, the neighbouring pixels were checked to find the closest value to the mean. The pixel was added to the region only if it satisfied the height threshold criteria. Initially this threshold was set to 5m which is the height of two floors of the building approximately. This was done to identify the different parts of the building with varying heights such as Church towers as a single segment. In this way, the entire building surface pixel was added to the region until there was no pixel found to be closer to the mean of the region i.e. due to a steep fall in the height due to ground next to a building pixel. This region-growing segmentation was performed with the code developed by (Kroon, 2008).

From the ortho-mosaic of the region, Green Red Vegetation Index (GRVI) was computed to exclude trees close to the building in the segmentation output. GRVI is one of the spectral indices used for identifying vegetation in RGB images. It is the ratio of difference between Green and Red bands to the sum of the same bands. The reflectance of canopy is greater in β_{Green} than β_{Red} . Therefore, $GRVI > 0$ can be identified as vegetation and $GRVI \leq 0$ can be interpreted as non-vegetation (Motohka, Nasahara, Oguma, & Tsuchida, 2010). Based on this threshold, a binary image was obtained. Image subtraction was carried out between the binary image from the building segmentation and the GRVI output. This step resulted in the elimination of trees if present in the building segment. At this stage, there can be jagged edges due to the removal of some vegetation pixel. This was corrected by performing morphological closing. Morphological closing is the sequence of dilation and erosion process to improve the continuity of boundary pixels by filling in small gaps and removing the outliers in the edges. The operations are performed with flat structuring element of 3x3 window. The contour of the building segment was traced using Moore-Neighbor tracing algorithm with Jacob's stopping criterion (C Gonzalez, E Woods, & R Masters, 2009; Ghunaim, 2000). The boundary pixels needed further processing to retrieve a regular polygonal building shape. Therefore, the outlier vertices were removed by relative distances of the vertices in the boundary ($< 2m$). Regularising the boundaries helped to overcome the difficulties in segmentation process as seen in figure 1.

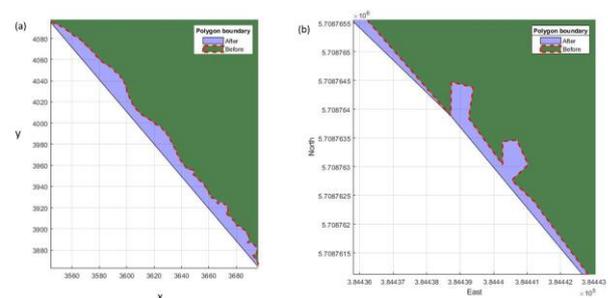


Figure 1 (a) Boundary after outlier removal (b) Final boundary after regularisation

2.2 Façade definition

The façades of the building must be defined in terms of length and orientation for interpretation. This helps to plan the exterior orientation of camera network to capture sufficient images by traversing along these façades. The corners of the façade need to be clearly defined in order to plan convergent camera network for connecting the façades. The boundary vertices obtained from the previous step represents the façade outline. Since the boundaries were filtered with threshold of $<2\text{m}$, a façade of minimum length of 2m were able to be identified for planning the camera network. The building corners need to be identified to plan the connecting cameras. In real world, not all the façades make right angles to each other. Therefore, the true corner locations need to be detected and classified into Exterior and Interior based on the angles made between the façades. This classification helps to plan according to the type of corners. The façades making angles between 30° and 160° are classified as interior and the façades making angles between 190° and 350° are classified as exterior corners which are shown in figure 2.

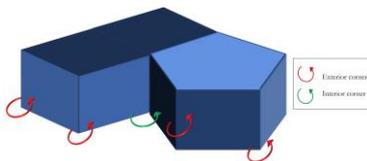


Figure 2 Interior (green) and Exterior (red) corners of a building

In order to ensure that every feature on the surface of façade is captured (i.e. check the visibility) a 3D grid of each façade was generated. The height of the building is extracted from the DSM by getting the Z value of the building segment. The lowest point of the building was obtained from the neighbouring pixels outside of the building segment corresponding to the ground surface. The minimum Z value of these pixels was assumed to be the ground surface, over which the building stood.

Assuming the walls to be vertical, the boundary vertices were projected to the ground level. The points were computed for the surface of façades in a regular grid interval. Depending on the required point cloud density for the façade, the parameter was adjusted to produce desired number of points per m^2 .

2.3 Initial dense camera network

The initial camera network is designed to be redundant on purpose, not to lose coverage in any region on the façade. The cameras are then filtered based on coverage in the next step. This kind of reduction from maximum observation scheme in photogrammetry is suitable to arrive at the optimal number of images to reconstruct the object (Agarwal et al., 2011; Alsadik, 2015). In general, high overlap percentage was considered for the initial dense camera network. From this information, the dense camera network along the façades was computed. If there was any camera, too close to the façade, they were removed for safety of UAV. This dense network plans special connecting cameras between the façades of the building. To ensure coverage of every point on the surface, to produce 3D model the basic photogrammetric parameters need to be considered. The algorithm allows the user to set the required criteria such as Ground Sampling Distance (GSD) and overlap percentage for a 3D model. Based on this criteria, other photogrammetric parameters for the camera placement are determined from the traditional aerial photogrammetric block described by (Wolf & Dewitt, 2013).

In order to ensure complete coverage, different approaches were adopted for various conditions while traversing along the façades. All the cameras in the network are planned along the façade in a sequence in anticlockwise direction, using the orientation of the façades. This helps to reduce the time while flying the drone on field as it will require less manoeuvres. The cameras were placed in a line parallel to the surface of façade at intervals determined by the Base to Depth (B/D) ratio. Several research on stereopair images had shown the importance of B/D ratio in reducing the error in measurement of the model (Hasegawa, Matsuo, Koarai, Watanabe, & Masaharu, 2000; Hu, Gao, Li, & Li, 2016; Jing, Yong-she, Jian-rong, & Jun-feng, 2008). For a better coverage and reliable accuracy B/D ratio for initial camera placement was considered from $0.1 - 0.4$. The nadir cameras of each façade were connected by converging cameras in the corners to avoid wide baselines. These cameras make an arc along the corner of the building, providing oblique images observing points on both the façades. For initial dense camera network, cameras are placed at every 10° transition. However, for the interior corners, the convergent transition was not possible as the façades make it difficult to fly close by. In such cases, the last and first cameras of the corresponding façades were tilted to look towards the interior corners as shown in figure 3. If that did not ensure better coverage, multiple images were taken at different directions capturing the inner corner from the same location.

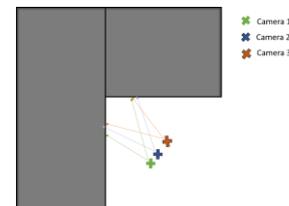


Figure 3 Cameras of neighbouring façades (green, red) tilted to look at the inner corner

2.4 Visibility analysis

The visibility check was performed to confirm if all the points in building façades were covered in at least 3 images. However, points measured in more than 4 images yielded less improvement in terms of accuracy as the previous research shows (Remondino et al., 2006). This was done by ray intersection method. Each façade point was tested with the angle made by the ray from the camera centre to the object point. From the Frame width and Frame height of the camera, Horizontal field of view (HFOV) and Vertical field of view (VFOV) were determined using the focal length. A matrix was created with camera id along the rows and façade point id along the columns. The cells were updated with value '1', if the point was visible in the particular camera. In this way, the number of images viewing a specific point was computed by adding along each column. This data was used for filtering the redundant camera in the next stage. Each camera based on the façade to which it was mapped, starts checking the points in that particular façade. Sometimes the points in neighbouring façades might be visible, if the façade normals make angles less than 90° . In such cases, the angle between camera looking direction and the normal of the façade was determined. This angle should be less than $\pm 30^\circ$ from the façade normal for the point to be verified by adjacent cameras. This helps to avoid taking into account the points that are viewed at steep oblique angles which will not be very useful for 3D reconstruction. In this way, object in the line of sight was detected as obstacle blocking the view of

other points in the neighbouring façade. This method helped to check the actual number of cameras viewing a point in the entire façade surface of the building.

2.5 Filter based on coverage

An ideal number of images sufficient to produce the same quality of 3D model as that of dense network need to be acquired to save time. This kind of minimal camera network problem was studied by (Alsadik, 2015). In that research, the author proposes two methods of filtering the redundant cameras: Filtering for coverage, filtering for accuracy. In this study, the filtering of cameras based on coverage was adapted to the UAV case according to Alsadik's work. The points visible in more than 3 images were iteratively identified as overcovered. The corresponding redundant cameras looking at the same point were back-traced from the matrix. Based on the number of points covered by each camera, the redundant cameras are sorted in ascending order. The camera covering least number of points was identified for removing from the camera network. Before removing a camera location, the algorithm checks whether all the points were still visible in at least 3 images. An additional check is performed considering the effect on B/D ratio given by the removal of this camera location. If the removal of a camera creates a huge baseline beyond the threshold specified, the camera is retained.

2.6 Optimization of camera network

The minimal camera network can be optimised to reduce the total error in the object points. The constraints are specified to limit the changes to minimal camera network. The objective is to obtain the optimal covariance matrix for the object coordinates which will maximise the accuracy for the given camera network. In general, the function constraints within an acceptable B/D ratio for accurate ray intersection. The 6 degrees of freedom (XYZ $\Omega\Phi\kappa$) for the camera are limited to ± 5 m space and $\pm 20^\circ$ for the angles. The optimization solution was adopted from the work of Alsadik (2015).

3. RESULTS

In this chapter, the results of each step involved in the study are detailed. One building obtained from the DSM and one synthetic building to assess the algorithm potential in particular complex conditions have been considered.

3.1 Building extraction

To study the presence of nearby trees occluding the building and posing difficulty to recover the true shape, building with trees in its neighbourhood shown in the figure 1 (a) was chosen.

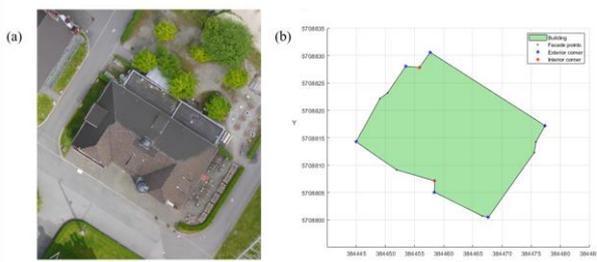


Figure 4 (a) Target building (b) Regularised boundary

The seed point on these buildings were selected for the region growing segmentation. The segmentation for building with no

trees in the vicinity was retrieved almost close to its original shape. Whereas, the building in figure 4(a) was segmented along with the trees as they overlap some parts of the building.

From the orthoimage, GRVI map was prepared. GRVI image obtained at the end of this process was binarized based on threshold > 0 , yielded the binary map of trees in the neighbourhood. It was observed that all the green vegetation including grass are identified as vegetation clearly. In the building, it is visibly identifiable that trees overshadow the boundary. The difference between the building segment and the GRVI image resulted in partially recovering the boundary of the building. These building segments traced by Moore's neighbour tracing algorithm due to the presence of trees, do not represent the true shape of the building. Hence the boundary was regularised to remove outliers and intermediate points in façade and a rough outline was obtained as shown in figure 4 (b).

3.2 Façade definition

From the regularised boundary vertices, the orientation of the façades was computed. Based on the threshold angles between the façades, corners of the building were identified as exterior and interior. The building had 3 interior corners of which 2 were identified. However, the third interior corner was not identified because it was less than 1.5 m from the main building façade itself. The height of the building is computed by finding the difference between the maximum value in the building segment and the value of neighbouring pixel to the boundary which is assumed to be the ground surface. Facades were then generated considering the height value and the building footprint. Points were generated on the surface at equal intervals of 1m for faster computations. Figure 5 shows the Building facade point cloud of generated using the boundary vertices and height of the building.

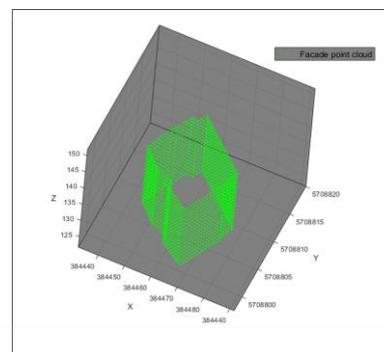


Figure 5 Building facade point cloud

3.3 Dense camera network

For camera view planning, the parameters of Panasonic DMC-GX1 model was used. These interior orientation parameters can be changed according to the cameras used. The initial overlap and sidelap were chosen as 90% and 40% with a GSD of 2cm for the dense camera network. This determines how close the UAV must fly near the object to achieve this resolution and adds a constraint to limit the maximum distance away from the object it can fly. Considering the accuracy of GNSS positioning in UAVs, a minimum distance of 10m from the object was used. Given the minimum distance, the minimum flying height for the UAV was 4.6m. This was to ensure that the foot of the building is covered.

The number of cameras planned, depended on the surface area of the façade to be covered i.e., perimeter * height of the building. The depth at which images to be captured were 20m and 40m away for real and synthetic buildings respectively. In this case, the planned GSD for building would be 0.5cm and 1cm to ensure a sub-centimetre resolution. These distances are uniform throughout the image capturing to get a homogenous resolution of building facades. Based on this, GSD, scale, baseline distances were computed. Assuming the expected collimation accuracy to be 1 pixel and maximum parallax as half of the frame width, the theoretical accuracy of the camera network of building was calculated. From the equations of Forstner (1998), the computed horizontal and vertical accuracy are 0.029m and 0.016m for real building; 0.04m and 0.06m for synthetic building respectively. For the dense camera network with 80% forward overlap at a predefined scale given by ratio between focal length and depth, the baseline distances were computed to be 4.1m and 7.15m respectively. It was observed that, the estimated accuracy for the computed dense camera network is within the acceptable range of few centimetres. However, it has to be noted cautiously, these values may vary in real field conditions. The initial dense camera network for real and synthetic building is shown in figure 6a and 6b respectively.

3.4 Visibility analysis

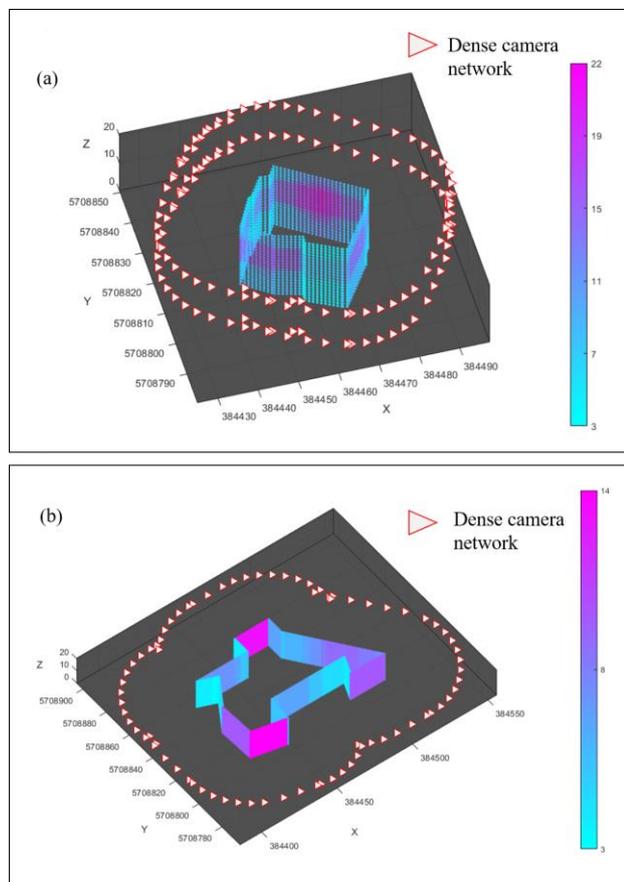


Figure 6 Number of images in which each point is visible (a) real (b) synthetic building in dense camera network

The visibility of each point of the building façade in the dense camera network was verified. This type of visibility check was quick for sparse point cloud of approximately 10,000 points. The algorithm was able to compute the results for visibility

analysis in less than a minute in this case. Figure 6 shows the visibility of each point in number of cameras for real and synthetic buildings. The dense camera network has a large number of cameras to capture the corners. Since the façade points are covered in these oblique images apart from the nadir images that were planned for each section of façade, visibility is as high as 22 and 14 images for some points in real and synthetic buildings respectively. The results of the visibility analysis were used to detect over-covered facade points and the corresponding cameras were removed in the next phase.

In order to study the efficiency of visibility analysis, a building with curved façade was used. From the figure 7 it was evident that, even though sufficient cameras were present, the check did not happen for the neighbouring points as they were interpreted as separate facades.

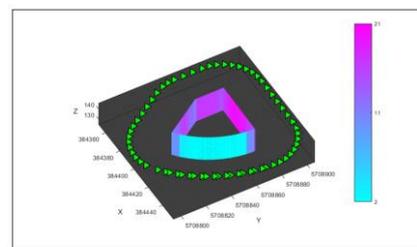


Figure 7 Visibility check for a curved façade

3.5 Coverage based camera filtering

From the visibility check, the algorithm was able to ascertain the parts of building which were over-covered by cameras. Then, the dense camera network was filtered based on the approach. The façade points that were covered in more than 3 cameras were identified and their corresponding cameras were listed as redundant cameras. From this list, the camera that was covering least number of points was selected and checked for B/D threshold of 0.7 for nadir cameras and 0.6 for the cameras at exterior corner. If the threshold is satisfied even after removing the camera, it was removed. In this way, redundant cameras are filtered iteratively until no further cameras can be removed. In most of the cases, the cameras were removed in one iteration successfully.

For the real and synthetic building, the initial dense camera network consisted of 139 and 89 cameras. This was reduced to 74 and 50 cameras respectively. The filtering of cameras for coverage, reduced the number of points captured in total by a camera network. However, the average number of cameras observing a point is as high as 6 and 5 for the buildings which is more than the required 4 images per point for 3D reconstruction.

In the minimal camera network, the highest number of images viewing at a point was 14 and 9 for the buildings. This is comparatively lesser than that of dense camera network. From the figure 4 and 5 it can be noted that the highest coverage of points was along the overlap of 2 strips in building, especially in the centre of the flat façade without any inner corners. This was particularly because of the additional converging cameras that were placed on both the corners of a façade, looking at it in oblique angles which allows to cover more points than the nadir cameras of the facades itself. However, constraints on the looking direction for these cameras i.e. $< \pm 30^\circ$ to the façade normal was sufficient to ensure the quality of point measured.

The algorithm ensured the reduction of dense camera network by 46% and 43% in these cases. This significantly saves a lot of processing time in the later stage. The minimal camera network with coverage in number of images is shown in figure 8.

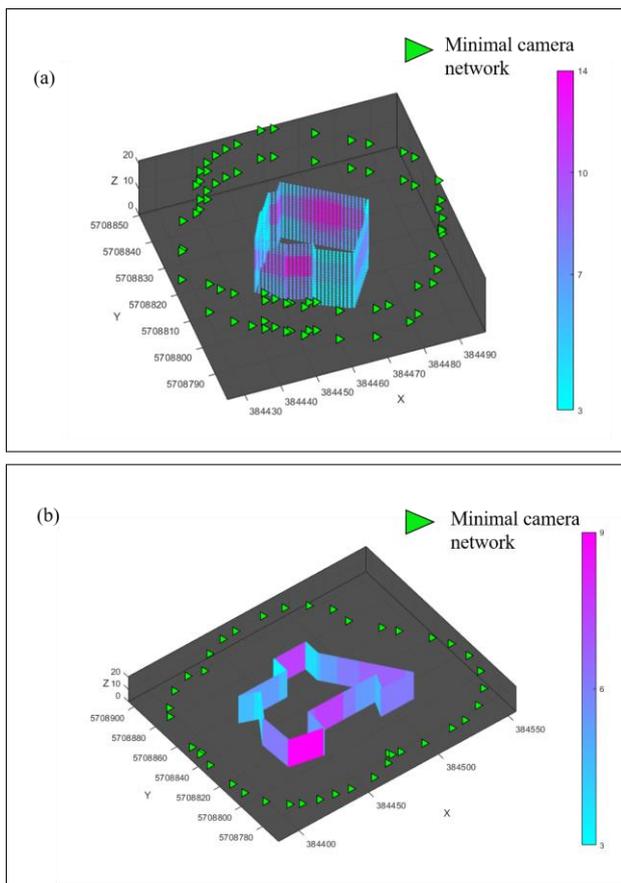


Figure 8 Minimal camera network (a) real (b) synthetic building

3.6 Optimization of camera network

The camera optimization is performed for each façade independently. A 3D model façade is shown in figure 9 for a reality alike simulation. In this figure, 4 GCP coded targets are placed on the corners of the facade while 5 checkpoints are placed on the facade and uniformly distributed. These GCPs are placed to orient the 3D model into the world local coordinate system while checkpoints are used for evaluating the output accuracy represented by the RMSE in their coordinates. In order to compare the improvement given by the optimisation, the original network is analyzed by creating synthetic images in the exact orientations mentioned earlier and then oriented using the Metashape software (“Agisoft,” 2019). The orientation is verified by evaluating the RMSE of the GCPs and checkpoints coordinates. Total RMSE of control points and checkpoints for minimal camera network were 1.1mm and 4.7mm initially.

Then the optimisation is applied fixing the angular orientations for simplicity while optimizing the camera positions to move within ± 5 meters from their initial values. The optimization is converged and stopped when the step size is smaller than a given threshold. Total RMSE of control points and checkpoints after optimization were 0.3mm and 2.9mm respectively. The changes in the configuration of the camera network is shown in figure 10.

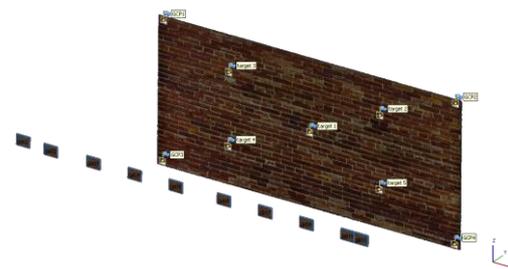


Figure 9 Distribution of GCPs and checkpoints and the initial minimal camera network

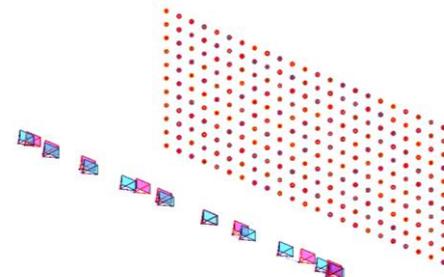


Figure 10 Original cameras (cyan) and optimal cameras (magenta) for fixed rotation optimisation

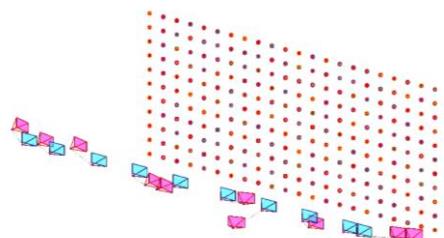


Figure 11 Optimization results including rotations. The original cameras (cyan) and optimal cameras (magenta)

In a second test, we run the optimization for the camera positions to move within ± 5 meters from their initial values while $\pm 20^\circ$ for the phi angles. The corresponding total RMSE of control points and checkpoints after optimization were 0.4mm and 1.0mm respectively. The optimal camera network is shown in figure 11.

4. DISCUSSION

The proposed flight planning method relies on delineating the building boundary from the rough DSM obtained from nadir flight. Mean based region growing algorithm strongly depended on the parameters used which limited from getting the building as one segment. In such cases, different segments of the building can be merged to get the overall footprint of the building.

Using GRVI for removing the tall trees, that hampered the precise boundary extraction worked well for the building in this case. However, some trees in the other region of DSM were not clearly identified because of presence of flowers that had different spectral signature. This misinterpretation can be easily incorporated into the building outline leading to irregular outline of the building. While regularising the boundary, some

intermediate points in the façade still existed and were considered an individual façade. If not removed, these points will lead to difficulties in visibility testing, as the test was implemented to check only the points in direct neighbouring façades. In most of the cases, boundary outlines were able to be retrieved with reasonable shapes. In case of this building, during regularisation of the outline, a small section of building was compromised due to the presence of trees. This can be solved by adapting a better generalising strategy when the vertices of boundary are too close to each other such as fitting a line with low residuals in the place of sequence of nearby points due to trees. In this way, some structures close to the trees can be retrieved.

When the traced boundary (x, y) coordinates were used to create a polygon using predefined function `polyshape` in Matlab, some of the unwanted vertices which do not add any definitive value to the shape were automatically removed. This helped reduce the vertices of the traced boundary. However, non-linear irregular boundaries due to trees and the cavities need to be rectified to get a good approximation of the building shape. It was observed that a small portion of building near the trees, was not able to be recovered as part of façade.

The building façade was assumed to be flat with no protrusion such as balcony or hanging roof. In this way points for the façade was generated based on the outline extracted from the DSM. This can lead to self-occlusion of some parts of the building based on the flying height of the UAV. Sometimes, minor inaccuracies due to variation in depth at these locations is possible. However, flying at a certain distance, these difference in depth will be negligible. The detection of corners of the building depends directly on the angles made between the meeting façades. However, in exceptional case, if the façades of buildings do not meet at corners in the angles specified, corner detection have to be altered for better detection. Since, corner detection is necessary for planning converging camera network. This step is critical to ensure the smooth connectivity of façades.

The accuracy of the 3D model depends on multiple factors such as calibration of the instruments, precise positioning, image matching on featureless surfaces, etc. The overlap percentage used in the design can be increased, if the building façade has minimal features which can act as tie points. This can ensure more homologous points in every image pair that will reduce error in matching. Varying the depth at which an object is observed directly impacted the theoretical accuracy of synthetic buildings flying at 40m away compared to flying at 20m distance.

In one of the test cases, the total cameras in the dense camera network configuration for two buildings with similar perimeter and height were 118 and 139. The variation in the number of cameras can be attributed to the number of corners of the building. More corners in a building added additional cameras in the network.

The mapping of cameras to corresponding façades helped to perform visibility analysis of points faster. For 10,000 points, the algorithm was able to deliver results in less than a minute. However, when tested for curved façade, the algorithm did not perform well since the curved façade was defined with points closer to each other, each of which was treated as individual façade by the program which was designed to check only the direct neighbouring facade. Even though sufficient camera coverage was there, visibility check could not be performed

from all the cameras as each camera is mapped only to a small segment of the curved façade. The limitation on visibility check for curved surface is shown in figure 7.

The coverage-based camera filtering provided better results for regular polygons with only exterior corners. When there were many interior corners, it was difficult to reduce the number of cameras as the number of cameras viewing at inner corners are rather limited and have significant impact if removed.

The optimisation improved the accuracy of project in both the cases. However, adjusting the camera looking angles along with position provided better accuracy compared to fixed rotation optimisation. Since optimisation is computationally intense step it is better performed when the accuracy of project is estimated to be considerably low.

5. CONCLUSION AND FUTURE DEVELOPMENTS

In this research, the algorithm was developed to compute the positions and orientation of cameras for UAV to capture a building for 3D model. The camera view planning, visibility analysis and the coverage-based filtering method was time efficient and could produce results in less than a minute. A minimal camera network for 3D modelling of a building with an average visibility of more than 3 images and estimated accuracy could be produced using the algorithm. The tests were performed on real as well as simulated buildings to check the efficiency of the algorithm. The performance of algorithm was better for well-defined building footprints of complex geometric shapes without obstacles in the scene. The camera planning and visibility analysis can be still improved by considering the obstacles in the scene which causes occlusion of facade.

The view planning is suited for obtaining homogenous resolution for the façades assuming the nadir images can be obtained over the building that will provide similar resolution for the roof. For gable roofs, an efficient planning is needed to get homogenous data on the roof similar to that of façade. By adapting a similar converging camera network in connecting the roofs to façades, problems in the connecting edges can be addressed. From the DSM, the slope of the gable roofs can be calculated and cameras can be placed to face the normal of these planes.

In visibility testing, the presence of obstacles such as trees in vicinity was not considered. This problem can be addressed either by planning cameras around the trees trying to capture the occluded façades in oblique images. Since the position of trees are known from the GRVI output, this can be used to check if a tree is affecting the line of sight from the camera to the façade point, by similar box ray intersection method. This can help to improve the reliability of visibility testing.

Filtering of redundant images proved to reduce the processing time up to 60% as demonstrated by (Alsadik, Gerke, Vosselman, Daham, & Jasim, 2014) using commercial and open source software such as photomodeler, Metashape, VSfM for orientation. The developed algorithm was able to reduce the number of cameras by up to 46%. This will help to collect quality data with less effort in terms of data collection and processing.

Optimisation is performed for individual façades and the converging networks are not considered. The optimisation can be improvised to multi-strip case for UAV. The method needs

to be tested on field to check the efficiency of the algorithm in producing building models and may be developed as an interactive application for non-professional users.

REFERENCES

- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., & Szeliski, R. (2011). Building Rome in a Day. *Commun. ACM*, 54(10), 105–112. <https://doi.org/10.1145/2001269.2001293>
- Agisoft. (2019). Retrieved March 5, 2019, from <https://www.agisoft.com/>
- Alsadik, B. (2015). *Guided close range photogrammetry for 3D modeling of cultural heritage sites*. <https://doi.org/10.3990/1.9789036537933>
- Alsadik, B., Gerke, M., Vosselman, G., Daham, A., & Jasim, L. (2014). Minimal camera networks for 3D image based modeling of cultural heritage objects. *Sensors (Switzerland)*, 14(4), 5785–5804. <https://doi.org/10.3390/s140405785>
- C Gonzalez, R., E Woods, R., & R Masters, B. (2009). *Digital Image Processing, Third Edition. Journal of biomedical optics* (Vol. 14). <https://doi.org/10.1117/1.3115362>
- Cefalu, A., Haala, N., Schmohl, S., Neumann, I., & Genz, T. (2017). a Mobile Multi-Sensor Platform for Building Reconstruction Integrating Terrestrial and Autonomous Uav-Based Close Range Data Acquisition. *International Conference On Unmanned Aerial Vehicles In Geomatics*, 42–2(W6), 63–70. <https://doi.org/10.5194/isprs-archives-XLII-2-W6-63-2017>
- Chen, S., Li, Y., & Kwok, N. M. (2011). *Active vision in robotic systems: A survey of recent developments. International Journal of Robotics Research* (Vol. 30). <https://doi.org/10.1177/0278364911410755>
- Forlani, G., Roncella, R., & Nardinocchi, C. (2015). Where is photogrammetry heading to? State of the art and trends. *Rendiconti Lincei*, 26(S1), 85–96. <https://doi.org/10.1007/s12210-015-0381-x>
- Forstner, W. (1998). On the Theoretical Accuracy of Multi Image Matching, Restoration and Triangulation.
- Ghuneim, A. G. (2000). Contour Tracing. Retrieved March 16, 2019, from http://www.imageprocessingplace.com/downloads_V3/root_downloads/tutorials/contour_tracing_Abeer_George_Ghuneim/index.html
- Harwin, S., & Lucieer, A. (2012). Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from Unmanned Aerial Vehicle (UAV) imagery. *Remote Sensing*, 4(6), 1573–1599. <https://doi.org/10.3390/rs4061573>
- Hasegawa, H., Matsuo, K., Koarai, M., Watanabe, N., & Masaharu, H. (2000). *DEM Accuracy and the Base to Height (B/H) Ratio of Stereo Images. ISPRS Archives. XIXth ISPRS Congress Technical Commission IV. Vol. XXXIII-B4* (Vol. XXXIII).
- Hu, F., Gao, X. M., Li, G. Y., & Li, M. (2016). DEM extraction from worldview-3 stereo-images and accuracy evaluation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2016–Janua(July)*, 327–332. <https://doi.org/10.5194/isprsarchives-XLI-B1-327-2016>
- Huang, R., Zou, D., Vaughan, R., & Tan, P. (2017). Active Image-based Modeling with a Toy Drone. Retrieved from <http://arxiv.org/abs/1705.01010>
- Incekara, A. H., & Seker, D. Z. (2018). Comparative Analyses of the Point Cloud Produced by Using Close-Range Photogrammetry and Terrestrial Laser Scanning for Rock Surface. *Journal of the Indian Society of Remote Sensing*, 1–11. <https://doi.org/10.1007/s12524-018-0805-z>
- James, M. R., & Quinton, J. N. (2014). Ultra-rapid topographic surveying for complex environments: The hand-held mobile laser scanner (HMLS). *Earth Surface Processes and Landforms*, 39(1), 138–142. <https://doi.org/10.1002/esp.3489>
- Jing, L.; Yong-she, S.; Jian-rong, W.; Jun-feng, Y. The research and design of the base-height ratio for the three linear array camera of satellite photogrammetry. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2008, XXXVII Pt B1, 757–760.
- Kroon, D.-J. (2008). Region Growing. Retrieved from <https://nl.mathworks.com/matlabcentral/fileexchange/19084-region-growing?focused=5098324&tab=function>
- Martin, R. A., Rojas, I., Franke, K., & Hedengren, J. D. (2016). Evolutionary view planning for optimized UAV terrain modeling in a simulated environment. *Remote Sensing*, 8(1), 26–50. <https://doi.org/10.3390/rs8010026>
- Mostegel, C. (2011). Supplementary Material: UAV-based Autonomous Image Acquisition with Multi-View Stereo Quality Assurance by Confidence Prediction, 1–7.
- Motohka, T., Nasahara, K. N., Oguma, H., & Tsuchida, S. (2010). Applicability of Green-Red Vegetation Index for remote sensing of vegetation phenology. *Remote Sensing*, 2(10), 2369–2387. <https://doi.org/10.3390/rs2102369>
- Murtiyoso, A., & Grussenmeyer, P. (2017). Documentation of heritage buildings using close-range UAV images: dense matching issues, comparison and case studies. *The Photogrammetric Record*, 32(159), 206–229. <https://doi.org/10.1111/phor.12197>
- OGC. (2012). OGC CityGML Standard. *OGC Member Approved International Standard*, 1–344. <https://doi.org/OGC-12-019>
- Remondino, F., & El-hakim, S. (2006). Image - based 3D Modelling: A Review, (December 2017). <https://doi.org/10.1111/j.1477-9730.2006.00383.x>
- Wolf, P. R., & Dewitt, B. A. (2013). *Elements of photogrammetry with applications in GIS* (4th ed.). McGraw-Hill Education.