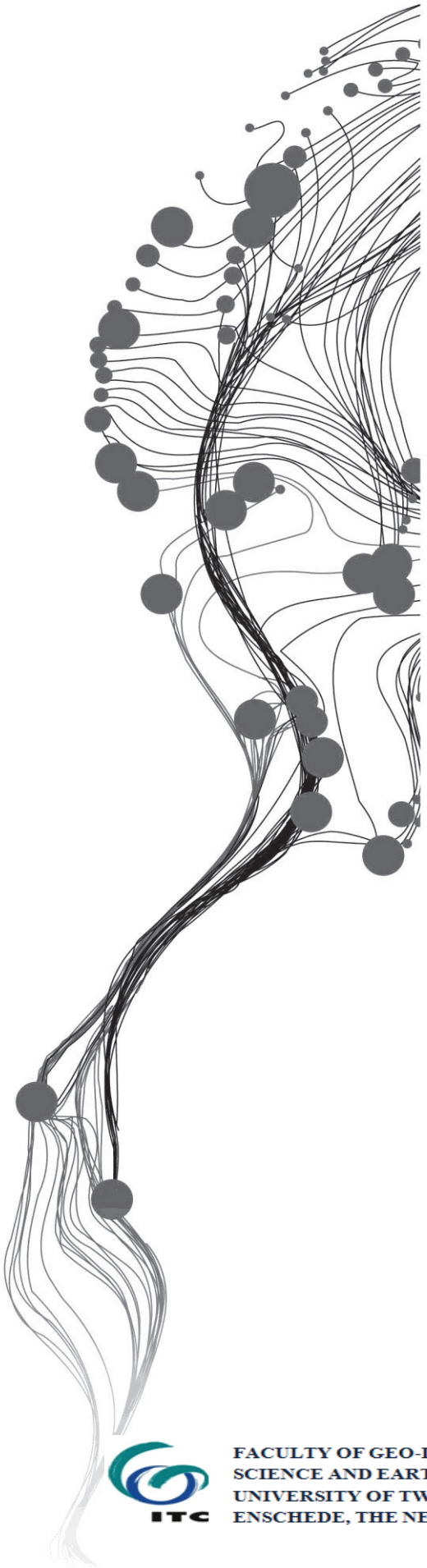
Automated 3D feature extraction for simple geometry buildings using images for GIS data collection

Ravisha Joshi
March, 2014

ITC SUPERVISOR
Dr. Markus Gerke

IIRS SUPERVISORS
Er. Ashutosh K. Jha



Automated 3D feature extraction for simple geometry buildings using images for GIS data collection

Ravisha Joshi
Enschede, the Netherlands [March, 2014]

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

THESIS ASSESSMENT BOARD:

Chairperson	: Prof. Dr. Ir. A. Stein
ITC Professor	: Prof. Dr. Ir. M.G. Vosselman
External Examiner	: Dr. P.K. Garg (IIT, Roorkee)
IIRS Supervisor	: Er. Ashutosh K. Jha
ITC Supervisor	: Dr. Markus Gerke

OBSERVERS:

ITC Observer	: Dr. N.A.S. Hamm
IIRS Observer	: Dr. S. K. Srivastav

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION, UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Government of India

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, The Netherlands. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Dedicated to my parents….

ABSTRACT

Three dimensional models have been widely used for various purposes ranging from urban planning to 3D gaming applications. Reconstruction of three dimensional models has been achieved through various datasets like aerial/terrestrial laser scanning and aerial/ terrestrial images. A lot of work has been done in this field applying various approaches. Most of it focuses either on estimation of primitives or uses prior information about the structure. If we focus on the case of Image-based modeling, reconstructed outputs are largely in the form of mesh models.

In this research thesis, simple geometric model of a building is created using many overlapping images. Digital camera is used to capture several images of the building to be reconstructed. A point cloud is reconstructed by applying Structure from Motion (SfM). The reconstructed point cloud thus obtained is in an arbitrary coordinate system and is required to be transformed to Global coordinate system. This is achieved by applying 3D similarity transform. On the transformed point cloud, RANSAC-based plane segmentation is implemented for the detection of dominant planes. This approach was chosen over other segmentation approaches because of its robustness to outliers and simplicity. Since the data may consist of many outliers, these are removed using statistical filter. The identified dominant planes represent the building wall. However, if ground has sufficient texture, a plane corresponding to ground will also get detected. The intersection of ground plane and all the other planes that are perpendicular to ground plane, are used to estimate two dimensional boundary of the building. The obtained two dimensional boundaries are then extruded to an estimated height of the building. The model is tested against two datasets the accuracy of which is discussed. The models created are a close approximation of the actual structure. However, given the poor accuracy of the digital camera GPS, the positional accuracy of the model does get affected.

Keywords: *Image-based Modeling, Three dimensional modeling, Structure from motion, Plane segmentation, 3D Similarity transform.*

ACKNOWLEDGEMENTS

It is indeed God's grace and divine providence to gain exposure to such a wonderful course as MSc, IIRS-ITC JEP. It would be highly selfish not to acknowledge all the people who have supported me during my research work. It's a herculean task to list out all the people who had helped me but will try my best.

On the completion of my MSc thesis, I owe my deepest gratitude to my IIRS supervisor Er. Ashutosh K. Jha for his continuous support, guidance, motivation and extraordinary scientific perception. Thank you sir, for your precious time, support and motivation throughout my academics at IIRS. You have helped me whenever I approached you, even when my findings were average and the doubts petty. Thank you again sir for motivating and uplifting my spirits when the goings were tough.

I would like to thank Dr. M. Gerke, my ITC supervisor, for his invaluable guidance and suggestions at every stage of this research work. It is he, who had given such an innovative and interesting research topic and taught a lot about computer vision and digital photogrammetry field.

I am also grateful to Dr. Y.V.N. Krishna Murthy, the Director, IIRS for providing excellent research environment and infrastructure to carry out our work. I would like to show my extreme gratitude to Mr. P.L.N. Raju, Group Head, RSGG, IIRS for his constant support and providing critical inputs for making this MSc program an invaluable experience. Would also like to acknowledge Dr. S. K. Srivastava, Head, Geoinformatics Department, IIRS for his guidance and lending a patient ear to our problems. Am also thankful to Dr. Nicholas Hamm, Assistant Professor, ITC for being there for us at all times and for providing valuable advice and especially for making our stay at ITC comfortable and memorable. A special note of thanks to you all for giving your valuable time and patient hearing to our problems, both technical and personal.

I also am honoured at having received lectures from Prof. Dr. Ir. M.G. (George) Vosselman, Dr. K. Khoshelham and Dr. Ir. S.J. Oude Elberink, ITC on 3D Geo-Information. My understanding of the subject owes gratitude to the lectures.

Last, but not the least, I offer my appreciation to my parents, brother, fiancé and all my batch mates from IIRS and ITC for their infinite support. Learning is a continuous and never ending process but times shared with you have shaped my attitude and the memories are etched into my personality, for that am ever grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Three dimensional models have been widely used for several applications such as urban planning, photo tourism [1] and cultural heritage documentation [2]. In recent years, aerial and terrestrial laser scanning has been found to produce most accurate three dimensional models. However, the cost of acquisition could be a limiting factor for some applications. Another limitation of aerial laser scanning is the unavailability of facade details. It mainly provides roof information and has very less or no data corresponding to the vertical walls. Although, it can be combined with terrestrial laser scanning for facade details [4], but, this further increases the cost of acquisition. Instead, Image based modeling could be exploited as a low cost alternative to laser scanning, especially for applications that do not require high accuracy and want to save on acquisition cost.

Image based modeling is a process by which information from two or more images is extracted to create a three dimensional model for an object. Image based model provides flexibility in terms of different viewing angles and positions while being very economic. Color and texture information is also captured in the data. Images can be acquired using complex cameras or sensors applied in photogrammetry or using consumer cameras. With advancement in technology, even the consumer cameras are capable of capturing images at high resolution. Some cameras have additional feature of Geo-tagging the images. Although the accuracy of the GPS camera is quite low (approximately 10 metres), yet they can be satisfactorily used for applications with low accuracy requirement.

Softwares like Autodesk 123D catch and Image modeler are capable of creating three dimensional models automatically. Autodesk 123D catch automatically orients the images in arbitrary coordinate system and creates a mesh model. User interaction may be required for stitching the images if the software is not able to identify common features in the input images and, to specify the scale. It does not support geometric modeling. Image modeler supports geometric modelling but the buildings have to be modeled manually. User interaction increases the time and effort. An automated system is better suited when rapid reconstruction of a large area is to be achieved quickly. The models created automatically might be less accurate but could be helpful in disaster response systems to assess the situation quickly.

Three dimensional reconstructions are commonly visualised as mesh models. These models may have jagged boundaries in case of sparse point cloud, resulting in a model that is difficult to edit. In Laser scanning, three dimensional models have been derived using the plane-based segmentation approach along with prior knowledge of the scene [5]. Similar approach can also be implemented for SfM point clouds. Additional properties of color and texture which are not available in laser scanning may also be exploited that can increase the accuracy of the results.

Instead of computing the model from the planes extracted, two dimensional lines computed by intersection of the planes can be a possible solution for reconstructing three dimensional models. These intersection lines can help extract the two dimensional boundary of urban structures based on which three dimensional models can be reconstructed.

1.2. Motivation and Problem Statement

Image based modeling has been a popular topic of research [6, 7]. It can reconstruct significantly accurate models at low cost [8] as compared to laser scanning. The initial cost of equipment itself is quite high in case of laser scanning, whereas, image-based modeling requires only a good quality camera. Image-based modelling is not restricted to terrestrial images. Aerial images, satellite images, videos and airborne LIDAR have also been used in some researches. The main drawback with aerial and satellite based images is that they largely capture roof information with a very limited view of vertical walls of the buildings. On the contrary, ground based images at high resolution can effectively recover facade information.

3D point cloud from images can be extracted through Structure from Motion (SfM) and subsequent dense image matching. In traditional photogrammetric approach, 3D location and camera pose were required as a priori information. SfM solves this problem by using key points in the images for estimating scene geometry, camera position and orientation. The key points are the common features invariant to scale and illumination that could be identified in the image pairs by applying a method called SIFT [9]. However, SfM results in a point cloud in an arbitrary coordinate system with no scale information. Hence, transformation to absolute coordinate system is achieved by 3D similarity transform using the Ground Control Points (GCP).

Most often the three dimensional model, from point cloud, are produced as polygonal mesh model also referred as Poisson Surface Reconstruction which are complex and difficult to store, index or render efficiently [10]. They do not follow the architectural constraints of the scene. Therefore, if simplified geometry is derived from the mesh models, we risk increasing the already existent error on further processing. Moreover, mesh models tend to look unpleasant when input point cloud has high noise level [11]. Another approach explored by Furukawa *et al.* [3] is by segmentation of point cloud obtained through SfM to derive a front facing model. The intersection of planes was used to calculate the dominant lines in the scene. These lines were used in depth map to implement structural constraints. This approach was primarily aimed at reconstruction of a single depth map, however, it does not discuss about defining the complete boundary of a structure. Our approach exploits these dominant lines to obtain the boundary of buildings on ground and derive a box-like model by extruding the two dimensional building boundary to an estimated height of the building.

1.3. Research Identification

1.3.1. Research Objective

My research topic focuses on reconstruction of three dimensional building models from geotagged images captured using low cost consumer cameras. The main objective of this study is to automatically identify the two dimensional boundary of buildings, extrude it to the estimated height of the building resulting in a three dimensional geometric model of the building and further use the GPS coordinates from images to transform it to global coordinate system and assess the quality of the model.

Sub-Objectives:

- To use the GPS location from geo tag in images to transform it to global coordinate system, and assess the quality of the model.
- To automatically identify the two dimensional boundary of buildings from the point cloud generated through SfM using the plane-based segmentation approach.
- To extrude the building plane to the estimated height of the building resulting in a 3D geometric model of the building.

1.3.2. Research questions

1. What considerations should be taken into account for acquiring images?
2. What is the effect on modeling due to error in point cloud as they are reconstructed from images by SfM software?
3. How is the accuracy of the model affected when Helmert transformation (7 – parameter transformation) is applied for transforming it to global coordinate system using GPS coordinates with low accuracy?
4. Which segmentation approach is suitable for planer point cloud segmentation?
5. How are occlusions and absence of data handled?
6. Under what conditions this modeling approach would be successful or fail?

1.4. Innovation Aimed at

Innovation of this research is to develop a low cost automatic method to derive three dimensional models from Geo-tagged images by using structure from motion and dense image matching techniques. Plane based segmentation method or surface extraction method is explored to extract the two dimensional boundary information. The two dimensional line retrieved by intersection of the planes are used to define the boundary of buildings.

1.5. Thesis Structure

The research work is organized as follows:

Chapter 1: Introduction, this section presents general overview about the research work. It describes the basic idea of topic, motivation, problem statement, research objectives, and research questions.

Chapter 2: Literature Review, this chapter deals with theoretical background of the study and literature review. It also explains various components of computer vision and digital photogrammetry.

Chapter 3: Methodology, this chapter describes the complete workflow of the study and description in details, about data used, hardware and software tools used.

Chapter 4: Results and Discussion, this chapter describe the experiments on the selected data, achieved results, its discussion and analysis.

Chapter 5: Conclusion and Recommendation, this section describes the answer of the research questions in concluded form and recommendations for further study.

2. LITERATURE REVIEW

2.1. Introduction

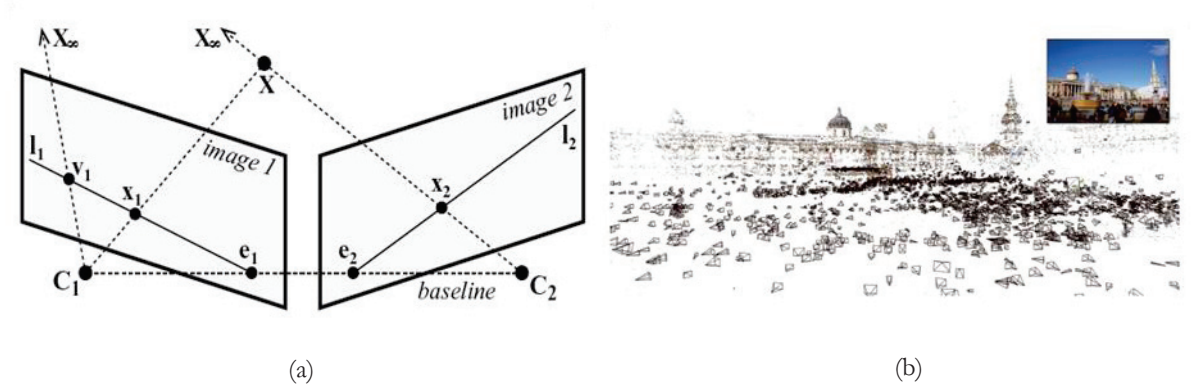
Three dimensional modeling has been described by Remondino and El-Hakim [6] as a process that begins with data collection and finally results in three dimensional models capable of being visualized and analyzed interactively on a computer and hence facilitating user friendliness to various operations. Three dimensional models are being actively used in various fields like urban planning, emergency response systems and cultural heritage documentation, to name a handful few. A large variety of data from different sources is utilized for the reconstructing these models such as aerial and terrestrial laser scanning, stereo image pairs, range images and set of overlapping images. Our research on "Automated 3D feature extraction for simple geometry buildings using images for GIS data collection" focuses on the automation of the modeling process using point clouds obtained through structure from motion to reconstruct building geometry. This has been a topic of research for very long in both computer vision and photogrammetry but, there is still a long way to go in making the process completely free from user interaction. Some existing approaches are reviewed below.

2.2. Point cloud reconstruction

Point Cloud data is used as input in our research to reconstruct the building geometry. This Point cloud is obtained through a sequence of images. In this section we discuss the process of generating point cloud from sequence of photographs.

A sparse three dimensional scene structure is derived from a sequence of overlapping images through Structure from Motion (SfM). Westoby et al.[8] have described in their paper implementation of SfM in geosciences applications for generating DEM from overlapping images. Traditionally, 3D location and camera pose were required as a priori information for scene reconstruction. SfM solves this problem by automatically estimating scene geometry, camera position and orientation. This is achieved by indentifying common features across image pairs. These features, also called key points, are detected by applying robust feature-point detection algorithms, like Scale Invariant Feature Transform (SIFT)[9] and Speeded Up Robust Features (SURF)[12]. Both algorithms can detect features without being affected by variation in scale, rotation, translation and illumination. The detected features are matched in image pairs using Approximate Nearest Neighbours (ANN) algorithm and outliers are removed by Random Sample Consensus (RANSAC). More about RANSAC is discussed in the later part of this section. Using these detected features in image pairs; image or camera orientation is recovered by applying epipolar geometry (Fig 2.1(a)). This relationship between the two views is represented in matrix form, referred as Fundamental matrix. Fundamental matrix can be computed by 8 point correspondence algorithm (linear solution) or 7-point correspondence algorithm (non-linear solution). Using fundamental matrix we can compute relative projection (rotation and translation) matrix for each camera pose. The location of common features after

applying the projection matrix is triangulated which results in a sparse 3D scene reconstruction. The described process could be scaled to generate sparse cloud for multiple photographs, as shown in the figure 2.1(b). This is referred as Structure from Motion as 3D scene structure is created from images taken by a camera in motion [13, 14]. The 3D data reconstructed through SfM is in arbitrary coordinate system with no scale information. Transformation to absolute coordinate system is achieved by 3D similarity transform using Ground Control Points (GCP). 3D similarity transformation requires seven parameters which consist of three rotation angles, three translations and one scale parameter.



(a) (b)

Figure 2.1: (a) Epipolar geometry in a nutshell [13], (b) Sparse point cloud generated from several thousand unordered photographs [1].

Fathi and Brilakis[15] have proposed a method to represent the geometry of the infrastructure using two video cameras. The point cloud is derived using two video streams captured simultaneously by two calibrated cameras. Speeded Up Robust Features (SURF)[12] algorithm is used to detect and match common features between stereo frames. Epipolar geometry constraints are utilized for finding good feature matches, and thus reducing the geometric error. Outliers from image matching are removed using RANSAC algorithm in which fundamental matrix is considered as the mathematical model. To increase the efficiency of RANSAC, Euclidean distance is used as constraint for selecting the matched pair of points. Triangulation is carried out on the matched features to generate the point cloud. Sparse point cloud obtained from both the cameras is then registered using the quaternion motion estimation method to estimate the camera motion. Reconstruction of point cloud is an incremental process as image frames are added one by one for camera pose estimation. Small errors can get introduced in the data due to errors in camera calibration, pixel size, resolution of the image or due to environmental factors (ambient light) and object properties (surface reflectivity) [14]. These small errors get accumulated and result in significant amounts of error after each processing step. To minimize this error, global optimization method such as Sparse Bundle Adjustment (SBA) is essential.

In the above discussion, feature matching was carried out using SIFT and SURF. Both of these approaches use RANSAC for outlier removal. RANSAC is an iterative process for estimating

parameters of a mathematical model such as plane or cylinder in a dataset which may contain outliers. Random Sample Consensus was first discussed by Fischler and Bolles[16] in 1981. Their paper describes one of the many applications of RANSAC i.e., smoothing data consisting of large amount of gross errors. Gross errors are mainly due to classification errors. Instead of working on the complete data at once, RANSAC selects a smallest possible initial data set and adds consistent data to the initial dataset wherever possible. There are three important parameters of RANSAC: error tolerance (to determine compatibility of point with the model), number of iterations and threshold t (number of compatible points for terminating the iterations). Schnabel et al. [17] have discussed another application of RANSAC for detecting various shapes in the point data set with high noise level. In their method, a primitive shape is identified in every iteration of the algorithm, among these the highest scoring candidate is found applying lazy score evaluation scheme. Probability of this highest scoring primitive shape is evaluated as size of shape (in number of points) over total number of shapes detected. Primitive shape is accepted only if its probability is high enough which signifies that no better shape was overlooked during sampling process. Once the candidate shape is accepted, points corresponding to this candidate are removed from the set of input point cloud. The process is repeated until all points have been removed or when it is not possible to extract any more shapes.

Point cloud generated by structure from motion using images of videos is very sparse, and does not always give a clear idea about the structure of the object. A dense reconstruction is therefore, required for the same. A dense point cloud reconstruction can be achieved by implementing Patch-based Multi-view Stereo (PMVS2)[18] algorithm, in which information is extracted from all pixels of the input image. PMVS identifies patches in the scene structure, back-projects it onto the images and expands the patches to nearby pixels to obtain a dense set of patches. These patches are then filtered to remove incorrect matches[19]. Other approaches for dense reconstruction include semi global matching and window-based matching algorithm. Semi global matching has been discussed by Hirschmüller[20]. It uses Mutual Information for matching individual pixels. However, pixel-wise matching can result in erroneous matches; therefore, smoothness constraint is also used. It penalizes the change in disparity of neighbouring pixels. Semi global matching is faster, more robust and minimizes both cost and constraints. Variation in image characteristics such as change in illumination, vignetting effect, etc, and properties of reflecting surface like non-lambertian surfaces can cause radiometric differences and increase the cost of matching. Mutual information based matching is capable of handling such radiometric differences reducing the matching cost. Goesele et al[21] has discussed Window based dense matching. Their approach first generates depth maps for each view individually. These depth maps are then merged to reconstruct a single mesh representation.

2.3. Automated Three dimensional Modeling

Poullis[22] presented a framework for automatically reconstructing three dimensional models from LIDAR point cloud. The process has been divided into three major steps: unsupervised clustering of the point cloud, boundary extraction of the roof surfaces and extrusion of these boundaries to obtain 3D polygonal models. A common problem faced in reconstruction of large

areas is processing of large amount of data altogether at once. Therefore, as a pre-processing step, point cloud is further divided into subcubes. These subcubes are processed parallely and independently from the other subcubes. This is performed by structuring the data as octree. Bounding cube is computed for complete point cloud and subdivided into memory manageable cubes. Points are then assigned to these cubes. Each cube can be further subdivided if maximum limit is attained. Surface is detected using P2C clustering algorithm which exploits the geometrical properties of the point cloud. It can be divided into two parts: extracting patches from point cloud and then extracting surface from these patches. Points are clustered into patches based on local height variation and normal variation of each point with respect to eight neighbouring points. In other words, Points exhibiting similar changes in height and normal with respect to the neighbouring points are grouped together. This is especially useful in extracting slant linear surfaces and uniform non linear surfaces. Patches exhibiting similar geometric properties are clustered together, by comparing normal distribution of adjacent patches. Finally, boundary of the resulting surfaces are extracted using the contour finding algorithm [23]. This is accomplished by calculating the orientation of the surface. Due to noise present in the data there are more orientations than actually possible. Energy minimization through graph cutting is used to extract the dominant orientation. Orientations of the boundaries are computed and iteratively refined to result in dominant orientations for each surface boundary. Since the data is divided and processed in parallel computation significantly less time is taken as compared to sequential processing.

In a paper by Furukawa et al.[3], Manhattan world assumption [24] is employed, i.e., all the surfaces are assumed to be aligned with X, Y and Z axes or in other words piecewise-axis-aligned-planer. This approach can help overcome the issue of matching features in surfaces that lack texture as is the case with many urban structures painted in single color. A dense point cloud is generated from sequence of images using freely available multiview stereo software. To begin with, calibrated photographs are given as input to freely available patch-based MVS software[18]. Output obtained is a set of oriented points consisting of 3D location and surface normal information along with photometric consistency score and a set of visible images. The output, however, is unreliable in case for surfaces with less texture. Further dominant axes are extracted making use of normals computed by PMVS. The resultant axes were found to be within 2 degrees of perpendicular to each other even in the presence of possible errors in camera intrinsic and given that urban structure may not have perfect orthogonal planes. In further processing, set of candidate planes are generated. These planes are used to recover the depth map for each image by assigning most suited plane hypothesis to each pixel in the image. This approach also utilizes the dominant lines, found at the intersection of two planes, for implementing structural constraints on the depth map. These dominant lines are computed using edge filters and used as cue for intersection of two surfaces. Author has also compared the final output, i.e., the depth map with mesh models as shown in the figure below. The depth maps give a clean reconstruction of the scene even in the absence of texture on the surface. On the other hand, mesh models, generated by Poisson Surface Reconstruction software [25], do not respect the architectural structure of the scene as they fill holes with curved surfaces where planer surfaces might have existed. In figure 2.2 we can see the difference in output produced.

(a) (b) (c)

Figure 2.2: Comparison between the described approach and the mesh model generated by Poisson surface reconstruction. (a) PMVS + Poisson Surface Reconstruction, (b) single depth map, and (c) single depth map with texture. [3]

The above approach produces a simple planer model that is easier to render, store and transmit. It performs well in the presence of planer surfaces but for non-planer surfaces it might produce incorrect results. Gallup et al. [26] handled both planer and non-planer surfaces in their modeling approach. A set of images, camera poses and depth maps are given as input. These are generated by Structure from Motion. Random Sample Consensus (RANSAC) is used to obtain planer hypotheses for each image. Multiple planer hypotheses are created and placed in the memory to be accessible to all the images. Planer surface spanning several images generates different plane hypothesis. All these planes are linked and fused together to give a single planar estimate. Subsequently, pixel-wise labelling is performed to label pixels of each image as planer, non-planer or discard on the basis of planer hypotheses, resulting in planar surfaces. A Planar classifier, that has been trained using training data based on image color and texture, is used for the same. Training data consisted of image segments that were labelled by the author as either planar or non-planar. Each image is processed individually without being affected by other images making the process scalable. This algorithm works well with textureless and specular surfaces.

2.4. Point Cloud Segmentation

Most of the modeling approaches use segmentation of point cloud into planes as an integral part of the methodology. In the following paragraphs, commonly employed segmentation approaches are discussed.

The paper by Dorninger and Nothegger [27] defines a highly robust method for modeling buildings from large, unstructured three dimensional point cloud using the segmentation approach. The point cloud is obtained by Image matching and Aerial laser scanning that result in high density data. In this approach, primitives are used to model the building and the task of segmentation and extraction of primitives is carried out simultaneously. The segments are selected or rejected based on constraints such as disjointing of every segment, connectivity to one another or distance from a threshold. Planes are detected using method based on Fast

Minimum Covariance Determinant (FMCD) approach. The author advocates the use of planes instead of point cloud as it reduces the time complexity of the algorithm. Clustering of point cloud is done in feature space to determine the seed clusters. Then region growing is performed by comparing the normal distance between the points from seed cluster plane against a threshold distance. If the distance is less than the threshold, the points are assigned to the plane. Subsequently, the connected component analysis is carried in object space and planes exhibiting similarity are merged considering both object and feature space. Since in the data collected was of high density and the flight height was also low, significant amount of points of vertical walls were captured as well. This aided in accurately estimate the position of wall in case of roof over hangs. In the absence of point data for walls, these are estimated based on boundary of roof.

Rabbani et al. [28] discussed in their paper Region growing approach using smoothness constraint for segmenting planes. This approach is considered suitable for plane fitting and not very useful for higher order surfaces, which are prone to errors in case of noisy data. The author starts with estimating the normal for each point by fitting plane to the surrounding points. For this purpose either K Nearest neighbour (KNN) or Fixed Distance Neighbours (FDN) can be used. Both of these approaches give similar results. Depending upon the density of point cloud, KNN varies the Area of interest, while FDN varies number of points. In KNN, a bigger area of interest is chosen in case of low density and vice-versa for high point density. Similarly more points are chosen in FDN in areas of low density or in the presence of high noise in the data. The second phase of the approach is Region growing, which uses the normals calculated before. Region growing is based on the closeness of the points with each another and with the fact that points in a segment should be smooth. Smoothness constraint can be ensured by checking that angle between normals within a segment do not vary more than an acceptable threshold. This approach was tested in an industrial environment where a lot of planer and curved surfaces exist. This approach also focuses on avoiding over segmentation or under segmentation of point cloud by setting appropriate parameters.

Hough transform has been used innumerable times for successful detection of lines and circles in previous researches. Most of them focused on two-dimensional dataset. Borrmann et al. [29] evaluated variants of Hough Transform with respect to their applicability in robotics. In Standard Hough Transform (SHT), each point is transformed to Hough space and score for each cell (plane) is incremented if the point lies on the plane. This incrementing of cell is referred as voting. In the end, cells with maximum votes represent the plane which comprises of maximum points on it. However, computational cost is directly proportional to size of point cloud in SHT as Hough Transform is performed on all points. Larger point clouds would incur higher computational cost. Probabilistic Hough Transform (PHT) selects a small part of the point cloud to reduce this computational cost. Number of points selected depends on the problem and should be optimally chosen. Another variation to this Adaptive PHT allows selecting larger subset of point cloud as compared to PHT and monitors the planes after each voting process. As soon as stable planer structures are detected, voting process is concluded. Randomized Hough Transform (RHT) differs in selecting of points from point cloud. In RHT, three points are randomly selected for defining a plane. These points should be close enough to

be selected. Closeness of other points to this detected plane is checked. If it is more than a given threshold the plane is accepted otherwise again three random points are chosen and the process is repeated. It is shown in the results that Random Hough Transform outperforms other variants of Hough Transform with respect to runtime, satisfactorily detecting planes and also performs better than Region growing and Hierarchical fitting of primitives in detecting larger planes.

Lari and Habib[30] discuss in their paper a hybrid approach for extracting linear cylindrical features from laser scanning data. Their hybrid approach combines techniques based on spatial and parameter domain. In the spatial domain based techniques, feature detection is dependent on the size of the neighbourhood. Region growing is one such technique the results of which may vary based on the seed points selected. Whereas, parameter-based techniques can be time and memory consuming for structures as simple as cylinder which has five dimensions resulting in 5-D attribute space for feature extraction. Existing hybrid approaches classify features in parameter domain and model them by applying least square fitting in spatial domain. This often yields unreliable results in detecting features aligned in same direction but with different radius values. The approach discussed by Lari and Habib[30] differs in detecting the features first in spatial domain and then extracting them in parameter domain. The implementation is divided in three parts. First part is classification of laser scanned points using Principal Component analysis (PCA). This classification is based on geometric properties of the points. A spherical neighbourhood is considered for the same. PCA results in Eigen values which are used for selecting representation models for each feature. The Eigen vector gives the approximate orientation of a feature and normalized Eigen values help classifying points into linear/cylindrical features. Larger Eigen values represent linear neighbourhood. In the second part, an iterative line and cylinder fitting is used to estimate geometric attributes and define appropriate representation model for each detected feature. Using the representation model, position and direction parameter is computed. Position parameter is the point of intersection of the features and the plane to which it is not parallel. And direction parameter is the direction of cylindrical axis of the feature. Finally, features are segmented in parameter domain which is reduced to low dimensional positional and directional subspace. Cluster extraction is carried out sequentially first in directional attribute subspace and then in positional attribute subspace. These segments represent the linear cylindrical features. This approach gave errors only in case of low point density and for points found on edges of planer features.

2.5. Literature Review Summary

Three dimensional modeling has been a popular research topic in both computer vision and photogrammetry. Some approaches are completely based on user interaction and others are semi or fully automated approaches. Interactive modeling produces quite accurate models but it is cumbersome with large amount of data. Although automatic modeling scales well with large data but it requires setting up a large number of parameters. Estimating these parameters can be time consuming and may require user inputs. Another aspect of three dimensional modeling is the type of data being used. Aerial and terrestrial laser scanners are usually equipped with GNSS and INS to produce point data with relatively high accuracy. On the other hand, MVS data is in arbitrary coordinate system and its positional accuracy depends on the accuracy of the GCP

points. It consists of high noise, outliers or holes due to lack of texture, occlusions, brightness change etc. These differences in LIDAR and MVS data may not allow applying the same methodology to both these datasets. Manhattan world assumption or estimating planer primitives have been used in some methods to overcome the problem of high noise in MVS data. Assuming the buildings to be planar reduces the problem to identification of planes in the point data. These approaches reconstruct three dimensional models by estimating planer primitives representing each surface.

3. METHODOLOGY

This chapter is divided into two sections; the first describes the data used and the study area. The second section describes in detail the methodology adopted in order to achieve the research objective.

3.1. Dataset and Software used

The data used in this project is a set of overlapping photographs taken from Sony Digital Still Camera DSC-HX10V camera. Camera is GPS enabled, i.e., the photographs have the location of exposure station logged as Exchangeable Image File (EXIF) data. Additionally, GPS data is compared with a planimetric map with accuracy of approximately 50cm. VisualSFM, freely available software, is used for creating point cloud data from the images. Table 3.1 lists all the softwares used in the project.

Table 3.1: Software and Instruments Used.

NO.	Software/Packages	Use
1.	VisualSFM	Point Cloud reconstruction and calculation of camera orientation
2.	XnView	Image pre-processing
3.	Visual Studio 2010	Implementing the methodology.
4.	Point Cloud Library	Point cloud handling and processing
5.	Sony Digital Camera(DSC-HX10V)	Capturing Images
6.	KML	Storage and Visualization of generated models.

3.2. Methodology

In this section, the adopted methodology is described in detail.

It starts with image acquisition using a digital still camera. Uncalibrated images are used to reconstruct a three dimensional model as these images are the most easily available images and requires minimum knowledge about the camera parameters for acquisition. Sparse Point cloud is generated applying the Structure from Motion approach, using freely available software

VisualSFM. A dense reconstruction is obtained through the PMVS binaries implemented through VisualSFM. This dense reconstruction is used as an input for our three dimensional modelling approach. The point cloud is in an arbitrary coordinate system, therefore, transformation parameters are computed and a 3D similarity transform is applied to transform the point cloud to global coordinate system. This data is then segmented into clusters representing building or ground plane. Intersection lines between building and ground planes are computed which define the two dimensional boundary of the building. This 2D boundary is then extruded to an estimated height resulting in a two dimensional boundary of the building. The output is stored as KML file which is visualized in Google Earth. Figure 3.1 shows the outline of the proposed methodology. Following it is a detailed description of each implementation step.

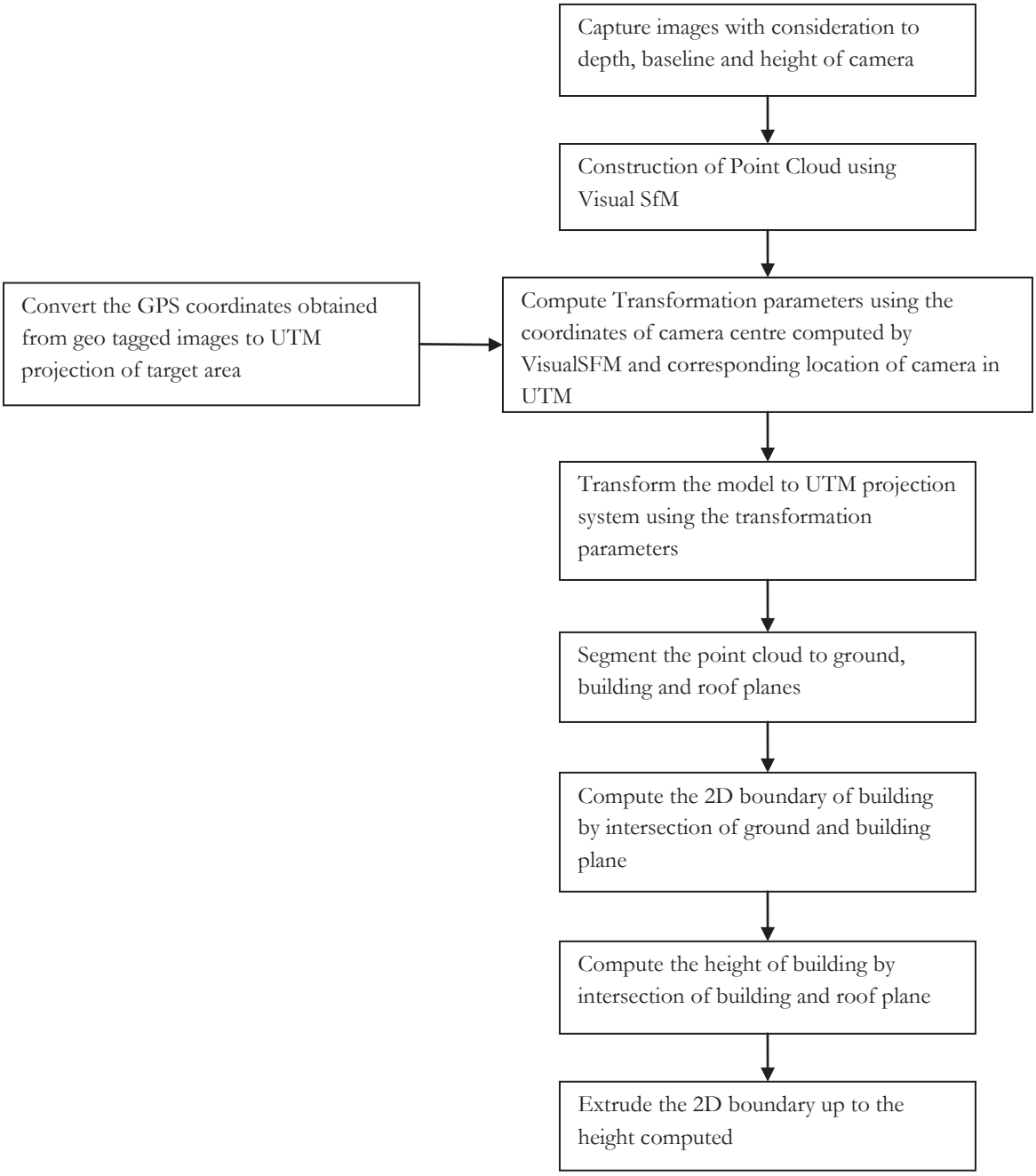


Figure 3.1: Methodology Workflow Diagram

3.2.1. Planning for Data Acquisition

The first part of the project is data planning and acquisition which includes taking photographs of buildings along with their GPS location. ISPRS close range photogrammetry report[31] provides a brief guidance for terrestrial photography. The following are few suggestions stated in the report. If possible, we should visit the site or review the object and take photographs prior to actual data acquisition to get a better understanding of the scene. Field work should be planned and the camera settings configured according to the working conditions such as weather,

visibility, sun or shadows, equipment, assistance and safety regulations. The output depends on the quality of the imagery; therefore photographic skill should be developed for consistently taking sharp images. Some points that should be considered for acquiring photos are as follows:

- Sharpest aperture setting should be used for the lens (often f/8).
- Lens should be fixed to infinity focus.
- Fastest Shutter speed should be used.
- In case of low light condition, ISO should be increased as necessary.
- Approximately 80% overlap should be ensured to obtain good quality output.

Although our methodology does not call for camera calibration requirements, yet the images should be sharp and with significant overlap such that a feature is visible in a minimum of three photographs, preferably more than three. Photographs should cover maximum possible faces of the building. Better approximations of the building shape can be achieved by photographing more faces of the building. The photographs should preferably be of high resolution as they provide better results. If required these high resolution images can be down-sampled to lower resolution but the vice versa is not possible. Another requirement is the GPS location of the exposure stations. This will be captured by the camera which Geo-tags the images. The GPS location is recorded in the EXIF tags along with other information about the camera and the image, such as focal length, model and make of camera, etc. In addition to this, GPS location should be captured with a more accurate GPS devise such as Differential GPS and tape measurements of the building should be taken to compare the accuracy of the model at a later stage.

3.2.2. Point Cloud Reconstruction

Once we have all the images, we reconstruct the point cloud using free software, VisualSFM[32]. Falkingham [33]has provided a brief overview about working with VisualSFM. Images resolution is selected depending upon the available memory for processing. If very high resolution images are used, memory consumption will be more which causes slow processing of images. Most suitable image dimension is 3200 resulting in sufficient information extraction without taking a lot of time. Once all the images are added in the software, GPU-accelerated feature matching is performed that is based on SIFT[34]. Next processing step in VisualSFM is to reconstruct a sparse point cloud, which is followed by dense cloud reconstruction using the PMVS binary files[18]. VisualSFM outputs other information as well along with the dense cloud data, such as, coordinates of camera centre in arbitrary system, corresponding GPS location, focal length, principal point, etc for each image. If all the images have sufficiently large overlap and features are matched effectively, then VisualSFM results in single model. Otherwise, it may result in multiple models.

3.2.3. 3D Similarity Transform

The dense point cloud obtained as above is in an arbitrary coordinate system without scale information. However to compare the measurements of the model in real world, we need to convert this model to global coordinate system. This can be achieved by 3D similarity transformation. The GPS locations obtained from EXIF tags of each image are in Geographic Coordinate System (GCS) WGS84 (latitude, longitude and ellipsoidal height). Since this is a spherical coordinate system, we need to convert it to Cartesian coordinate system. We will transform the data to Universal Transverse Mercator (UTM). This is a two dimensional coordinate system with height same as GCS, i.e. ellipsoidal height. The following parameters are used for converting the latitude longitude values to UTM projection.

Equatorial Radius (meters), a = 6,378,137,

Polar Radius (meters), b = 6,356,752.3142,

Scale along central meridian of zone, k = 0.9996.

Once all the GPS locations of camera centres are converted to UTM projection, we use 3D similarity transformation to transform the model from one coordinate system to another. In similarity transformation, scale parameter is same in all the directions and shape of the model is preserved. 3D similarity transformation uses seven parameters that can be subdivided as three rotation parameters and three translation parameters along x, y and z direction, and one scale parameter. The algorithm requires at least three point correspondences. If we have more than three correspondences, then a least square adjustment is used to reduce the errors. The relation between the two sets is shown by the Bursa-Wolf formula for 3D Helmert transformation in equation 3.1.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{(2)} = \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix} + S * R(\alpha_1, \alpha_2, \alpha_3) * \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{(1)} \quad (3.1)$$

Where:

$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{(1)}$: represent camera centres in arbitrary coordinate system computed by MVS software;

$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{(2)}$: represent camera coordinates in global coordinate system captured in EXIF tag of images;

$\begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix}$: are the three translation parameters along x-,y- and z-axis respectively,

$(\alpha_1, \alpha_2, \alpha_3)$: are the three rotation angles about x-, y- and z- axis respectively.

S is the scale factor, and R is the rotation matrix, that is the product of three rotation matrices as shown given in equation 3.2 and 3.3.

$$R_{3\times3} = R(\alpha_1, \alpha_2, \alpha_3) = R(\alpha_3) \cdot R(\alpha_2) \cdot R(\alpha_1) \tag{3.2}$$

$$= \begin{bmatrix} \cos\alpha_3 & \sin\alpha_3 & 0 \\ -\sin\alpha_3 & \cos\alpha_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\alpha_2 & 0 & -\sin\alpha_2 \\ 0 & 1 & 0 \\ \sin\alpha_2 & 0 & \cos\alpha_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_1 & \sin\alpha_1 \\ 0 & -\sin\alpha_1 & \cos\alpha_1 \end{bmatrix} \tag{3.3}$$

A geo-referenced point cloud is generated by applying the estimated parameters on the complete point data. Although, camera GPS accuracy is quite poor, approximately 10 metres, yet the model will not be affected in terms of shape as 3D similarity transformation is a rigid transformation. The residual error and RMSE are calculated for the transformation by back – transforming the points to validate the transformation parameters.

The geo-referenced point data is further used for plane segmentation.

3.2.4. RANSAC-based segmentation

Most of the urban structures exhibit a common property of being planar and orthogonal to the ground. This assumption is called Manhattan world assumption [24]. Each wall of a building and ground can be represented by planes. The data will also consist of points representing other features like trees, cars and structures that are not part of the building. Additionally, error in measurements introduces outliers in the data. These points are usually sparsely distributed. This property can be exploited to remove such points by applying Statistical filter. A neighbourhood of size k points is selected for each point and sum of distances between each point and its neighbouring point is calculated. Assuming Gaussian distribution of points, mean and standard deviation are calculated. Points falling outside the first standard deviation are considered as outlier and trimmed out from the dataset.

This dataset is further used for plane detection using RANSAC. RANSAC is robust to outliers and is simple to implement. RANSAC randomly selects minimum required sample points to estimate a model. Since we want to detect planes, the minimum number of points required is three. Points that are within a threshold distance from the estimated model are counted. The process of estimating plane and counting points is repeated for a fixed number of iterations, n. The plane with maximum number of points closer to it is selected as the best candidate. The points that are within the threshold distance to the best candidate are termed as inliers. These points are removed from the complete point dataset. RANSAC is again applied on the remaining points to extract remaining dominant planes. The process is repeated until no further planes could be estimated from the remaining points or if the remaining points are less than a threshold number. This process is briefly shown in the Figure 3.2. Applying RANSAC iteratively results in a set of planes defined in Hessian Normal Form as,

$$ax + by + cz + d = 0 \tag{3.4}$$

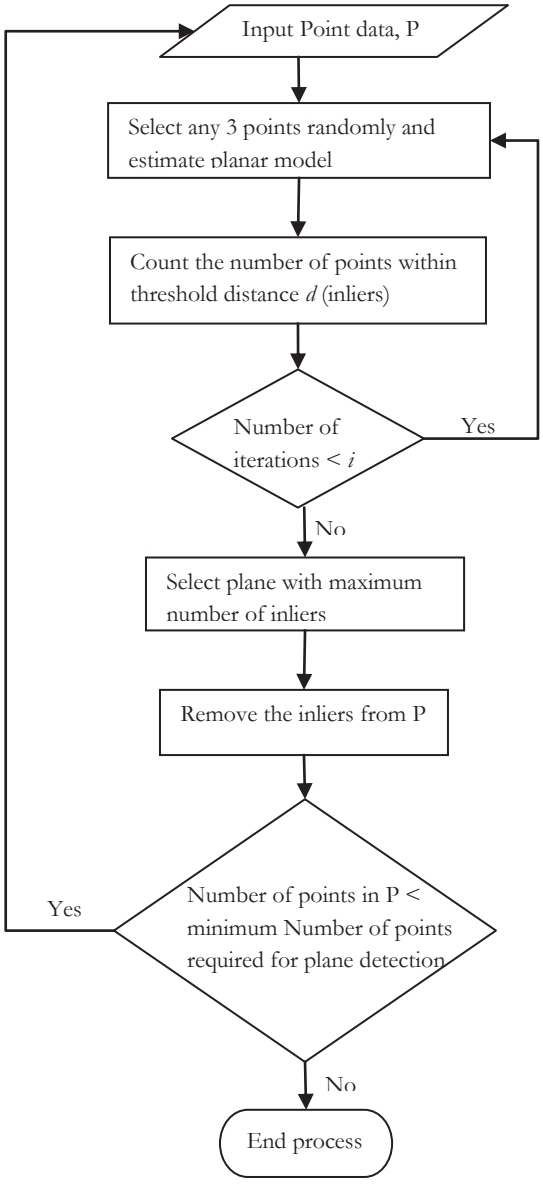


Figure 3.2: RANSAC flowchart

Planes corresponding to building walls will have dense neighbouring points. Whereas, points other than those on the building surface, might have been reconstructed due to presence of trees and other smaller objects in the scene. Points on such features are usually sparse. These points can be removed by applying statistical filter. A neighbourhood of size k points is selected for each point and sum of distances between each point and neighbouring point is calculated. Assuming Gaussian distribution of points, mean and standard deviation is computed. Points falling outside the first standard deviation are considered as outlier and trimmed out from the dataset.

Now we have a set of planes, which most likely represent the building walls. Since the points are in UTM projection system, X and Y value gives the position of the point on the earth surface and Z value gives height of the point. Therefore we can assume X-Y plane to represent the ground surface. Height of the ground can be assumed by the average of a considerable number of minimum Z-values in the point cloud. If the ground has sufficient texture, a plane corresponding to ground might also exist. Hence, in order to find the actual ground plane, angle between this assumed ground plane and all existing planes is calculated as shown in Figure 3.3. A plane is considered as actual ground plane if it makes angle smaller than 20 degrees. 20 degrees is chosen as the actual ground will not be perfectly parallel to the X-Y plane. If no such plane exists, the X-Y plane is assumed as the ground plane. Next we find the planes corresponding to walls of the building which should be orthogonal to the ground plane. Given the presence of noise and outliers we will allow a variation of 10 degrees in the angle between ground and wall.

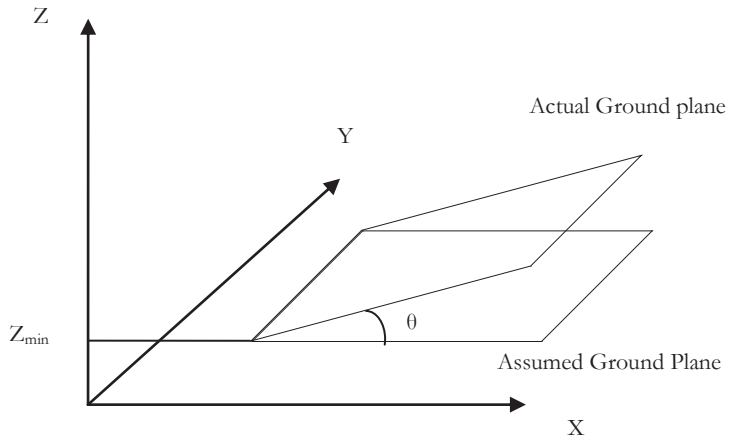


Figure 3.3: Angle between assumed and ground plane

3.2.5. 2D boundary estimation

Once we have classified the building and ground planes, we will define the 2D boundary of the building as on ground. The lines formed by the intersection of ground plane and building walls give the 2D boundary of building. Line equations are computed by applying method based on Lagrange multipliers[35]. However, these line equations define infinite lines. In to extract the line segment forming the building boundary, we need to know the end points of the line segment or building corners. Building corners are the point of intersection between the two adjacent lines. Since there are many lines possible, we should have the information about walls that are adjacent to each other.

The information about adjacent walls can be extracted by estimating the density of points along the edges of the walls with respect to other walls. In other words, adjacent planes will have a significant number of points on the vertical edges that are quite close to the other plane as shown in figure 3.4. Therefore, we count the number of points of one planer segment that are within a minimum possible distance to the plane of another planer segment. It is based on

calculation of point to plane distance which is the dot product of point coordinates and plane normal. Distance from all the points of each planer segment to another plane is calculated. For every point that is within the specified distance, count of points is increased by 1. If the count of points is large enough then the planes are considered adjacent. The process is repeated for all possible combinations of planes. The outputs of this step are pairs of adjacent planes. Another constraint applied here is that the planes should have an angle not less than 80 degrees. This is in accordance with the Manhattan world assumption that urban structures usually have a block like structure where walls are perpendicular to each other.

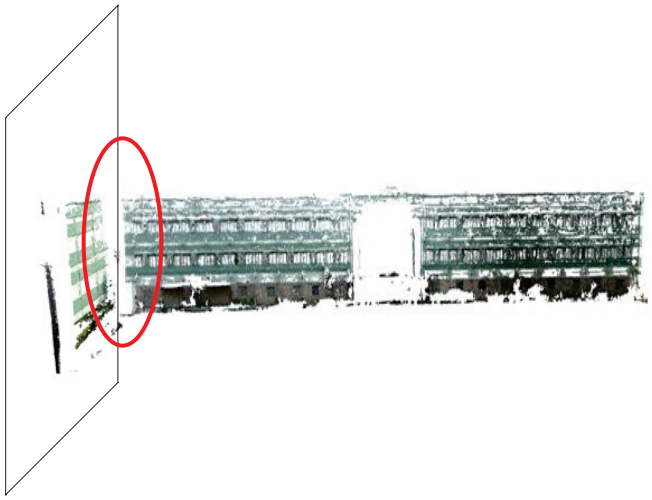


Figure 3.4: Adjacent planes. Red circle shows that the points on the edge one wall are very close to the plane of adjacent wall.

Using plane adjacency information, building corners are calculated. Point of intersection is calculated between boundary lines of adjacent planes. If a plane has two corresponding adjacent planes, then that building wall is completely defined. If however, a plane has only one corresponding adjacent plane, then at least one side of building was not captured in images and hence was not reconstructed in the point cloud. For this wall we have only one corner defined and we need to find the other corner. If we project all the points of this plane on the ground plane, then the farthest point from the defined corner will be the other corner of the wall on ground as shown in Figure 3.5. Projecting the point cloud on ground plane removes the height component of point cloud and the problem of finding a point in three dimensional space reduces to finding a point in two dimensional space. Each point cloud segment that has only one wall adjacent to it is projected on the ground plane. One corner of the planer segment that is obtained from line intersection is used as pivot and the point farthest from pivot point becomes the other corner. But this point could be an outlier. Therefore to reduce the probability of point being an outlier, the farthest point should have some significant number of neighbouring points. This will ensure that the point was part of the wall.

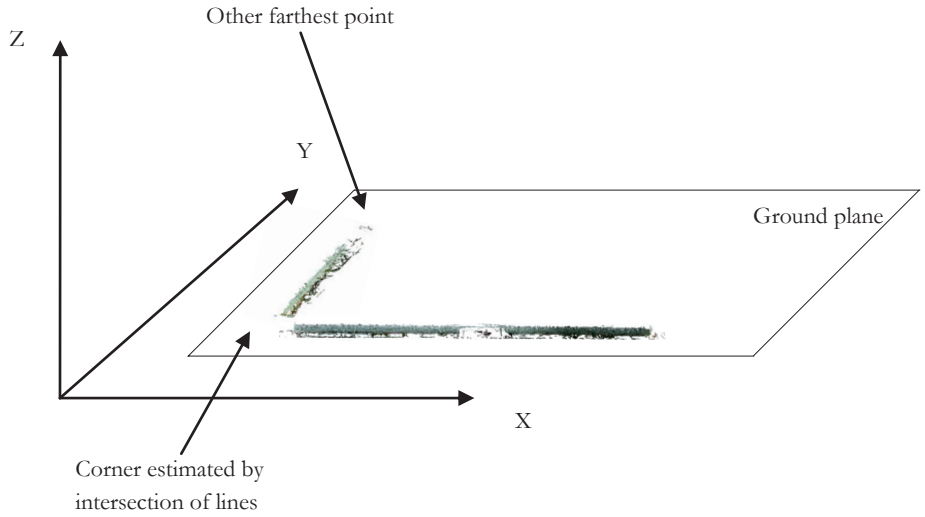


Figure 3.5: Projected Planes. The figure shows the point cloud of walls projected on ground plane. Plane having only one adjacent wall and the farthest point from the estimated corner of the wall.

Now we have the approximate corners of the building on the ground. These points define the 2D boundary of the building. Our final implementation step is to create a 3D dimensional model of building by extruding the 2D boundary of building to the estimated height of the building. The height of a building could be estimated from the height of the individual segments. The height of each individual segment is estimated by calculating mean of at least 100 maximum Z-values in the point cloud. This is done so as to get a better approximation of height. Finally we create the KML output file using the corner coordinates and extruding it to estimated height. The results are discussed in the next section.

4. RESULTS AND DISCUSSION

This section discusses the results obtained through the methodology explained in the previous chapter. In first part of the section, results are discussed and in the second part, accuracy assessment of the models is carried out.

4.1. Point Cloud Reconstruction

Main building of IIRS was chosen as the test scene. Images of three sides of the building were taken. Similarly, the images of another building (gym) in IIRS was used to verify if the process works with a different dataset. A total number of 228 images of main building and 176 images of gym were used. Large number of images ensured larger overlap for a better model reconstruction. Point cloud was reconstructed using VisualSFM as shown in figure 4.1 and figure 4.2. The figures show sparse and dense point cloud reconstruction. Model reconstructed for gym building is shown in Appendix-1.

(a) (b)

Figure 4.1: Sparse reconstruction obtained using Visual SFM. It also shows the camera locations. (a) front view,

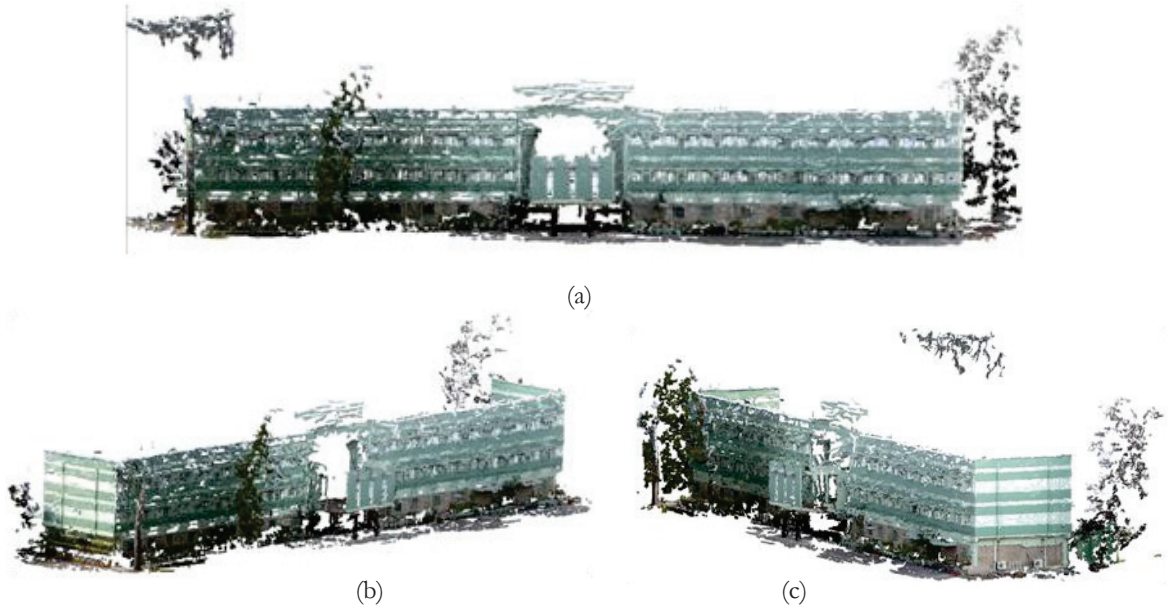


Figure 4.2: Dense reconstruction obtained using Visual SFM. (a) front view (b) & (c) side view on both side of building.

4.2. 3D Similarity Transformation

3D similarity transformation requires seven parameters, i.e., three rotation and three translation parameters along X-, Y- AND Z-direction, and one scale parameter. Camera centres in arbitrary coordinate system and the corresponding GPS locations captured in EXIF tags were provided as input for computing the seven parameters. The following table shows the parameters computed:

Table 4.1 : Values of seven transformation parameters

Parameter	**Value**
Rotation along x-axis, r_X	1.73451
Rotation along y-axis, r_Y	0.196378
Rotation along z-axis, r_Z	3.6842
Translation along x-axis, T_X	215821
Translation along y-axis, T_Y	3.36027e+6
Translation along z-axis, T_Z	693.684
Scale, S	0.11565

RMSE error was calculated by transforming the camera centre coordinates that are in arbitrary coordinate system to global coordinate system and then calculating the residual by subtracting the transformed coordinates and the coordinates obtained from Camera GPS. Approximate RMSE error in easting, northing and height values is 2.089 metres, 4.399 metres and 1.507 metres respectively. The error in transformation is quite high. This will further affect the positional accuracy. Although there is not significant influence on shape estimation as 3D similarity transformation is rigid transformation, i.e., it does not change the shape of the model.

4.3. RANSAC-based Segmentation

RANSAC based plane estimation requires us to set two parameters, first is the distance threshold and second one is the minimum number of points on which RANSAC can be applied. The distance threshold instructs the RANSAC to look for points that are within this specified distance from the estimated model. This parameter should neither have large or very small value, as both will result in incorrect results. In case of very large values, the number of planes identified will be less than expected and, in case of small values; the number of planes will be more. The value of this parameter for test scene was chosen as 0.55. The second parameter specifies the minimum number of points on which RANSAC is applied. This parameter is required to restrict the search to identifying only dominant planes. It will help ignore all the unnecessary planes that might exist. The value of this parameter for test scene was set as 10% of size of complete cloud. The planes detected in the cloud are shown in figure 4.3.

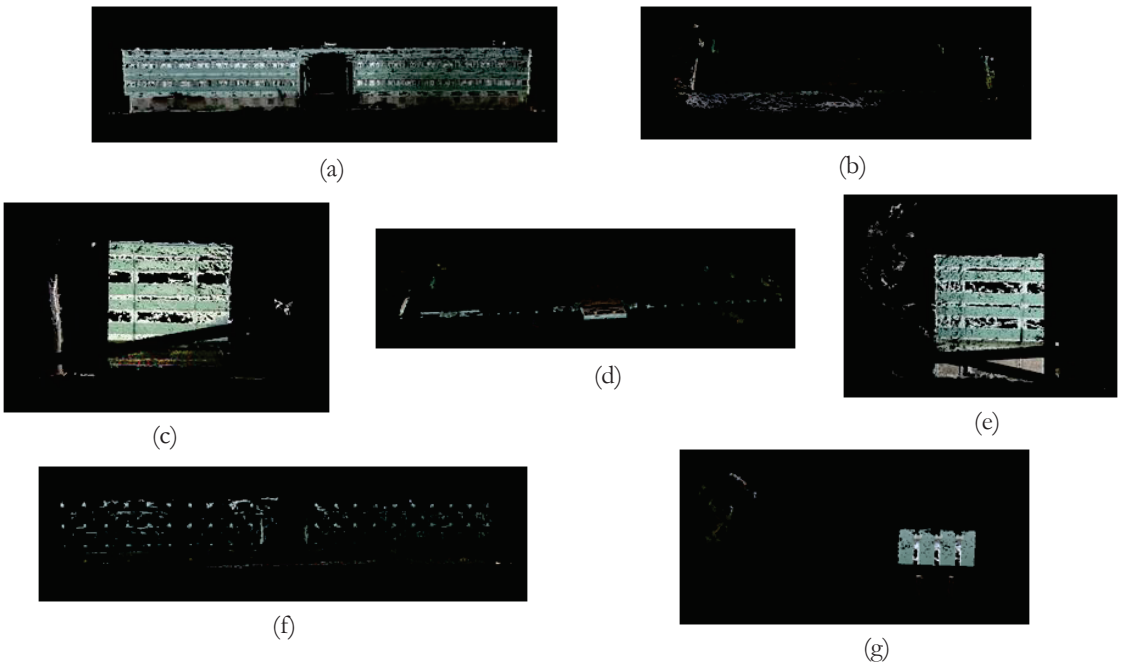


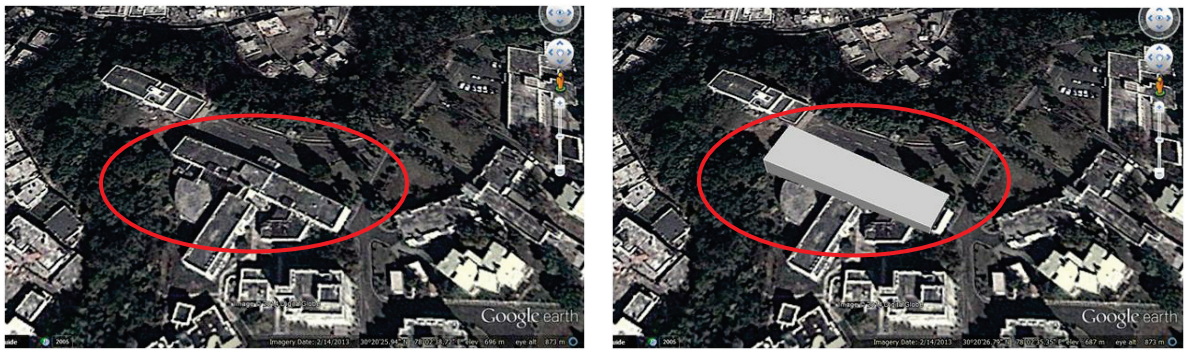
Figure 4.3: Plane segments obtained from RANSAC, (a), (c), (d), (e), (f), and (g) are the planes from building walls and, (b) is the plane segment for ground.

The same methodology was also applied for the gym building dataset. All the parameters were kept same except the distance threshold. It was changed to 0.30. There is no standard way to

determine the value of this parameter. Results using different values were compared and the one giving the best planer models was used.

4.4. Three Dimensional Modeling

The final output obtained is the corner points of the building that define the boundary of the building. The output is exported as a KML file that forms the building model using the corner points, ground height and height of the building. Corner points define the boundary and are extruded to the height of building. The output is shown in Figure 4.4. From the figure we can see that the position and shape of the model closely fits the actual structure in Google Earth. However, dimensions of the model and the actual dimensions of the building do not match. Comparison between the dimensions of the model and building are shown in table 4.2. The distance between the points does not consider the height, i.e., Z value is not considered for calculating the distance.



(a) (b)

Figure 4.4: Three dimensional box-like model of main building of IIRS viewed in Google Earth. (a) The actual location of main building marked in red circle, (b) the three dimensional model imported as KML file.

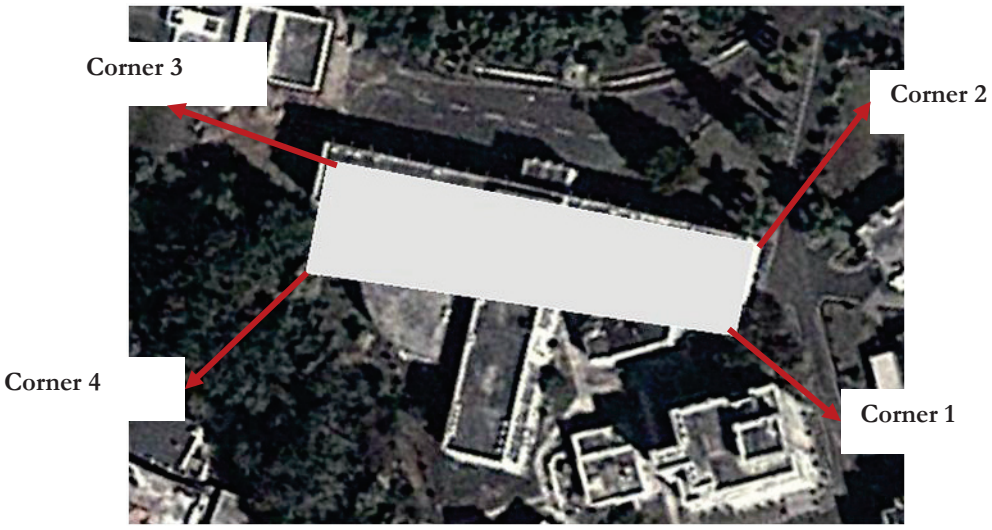


Figure 4.5: Distribution of corners of the building

Table 4.2: Comparison between dimensions of building and model

Dimension	Survey measurements (metres)	Model measurements (metres)	Difference
Length	72	73.076	1.076
Side wall 1	9.8	15.52	5.72
Side wall 2	10.1	18.84	8.74
Height	11.7	18.6	6.9

Table 4.3: Comparison between coordinates of the corner points

Coordinates	Survey measurements (metres)		Model measurements (metres)		Distance (metres)
	Easting	Northing	Easting	Northing	
Corner 1	215852.515	3360225.997	215847.7344	3360217.75	9.5324
Corner 2	215858.441	3360235.81	215855.7188	3360231	5.5269
Corner 3	215798.135	3360274.436	215793.0156	3360266.25	9.6549
Corner 4	215792.084	3360264.786	215783.625	3360250	17.0346

4.5. Accuracy Assessment

Theoretical Accuracy:

Following are the camera and building structure parameters:

- Focal length, = 4.3mm
- Height of the building, = 11.5m
- Physical sensor size = 7.76mm
- Pixel size, sx = 7.76/3200=2.425μm
- Double, the pixel size of the camera, Spx = 4.85μm
- Baseline length, b = 0.8m

Following are the calculations from [14]:

- Spx ≈ 2 * sx ≈ 2 * pixelsize = 4.85μm
- Scale, Mb = H/c ≈ 2675
- sH = (H/b) * Mb * spx = (11.5/0.8) * 2675 * 4.85 = 18.64 cm
- sX = mb * sx = 6.49 mm

Depth accuracy, sH = 18.64 cm

Parallel accuracy, sX = 6.49 mm

As shown in Figure 4.4, the model appears similar in shape to the actual structure. Additionally, it is positioned very close to the actual position of the building in Google Earth. Table 4.2 shows the measurements taken along the building wall and the reconstructed model. It also shows the difference in measurements. Figure 4.5 shows the corners of the building that were estimated in the process. GPS locations of these corners estimated from our approach and as obtained from the planimetric map of accuracy approximately 50 cm are compared in table 4.3. The table also highlights the distance between the coordinates. . The variation in the length of wall having two adjacent walls in the input point cloud, i.e., wall between corner 2 and corner 3 is small as compared to the variation in length for walls with only one adjacent wall, i.e., wall between corner 1 and corner 2 and wall between corner 3 and corner 4. This large difference in length is due to the presence of outliers in the data that is explained in the following paragraphs. The shift in the position of the model is within the accuracy of the camera GPS. Similar, comparisons were also made for gym building and the results are shown in Appendix-2.

Large difference in length of side-wall can be explained as follows. The two side-walls have only one wall adjacent to it in input point cloud, i.e., wall between corner 2 and 3, and corner point associated with this adjacent wall was calculated using line intersection. However, due to absence of adjacent wall along the other edge in the point dataset, an assumption was made for calculating other corner of the wall. The assumption made was that the other corner will be the point on the wall (or point cloud) that is farthest from the already calculated corner of the wall. This point cloud was projected on the ground plane to ignore height of the wall. Since the data consists of outliers, the farthest point may not necessarily be part of the wall, causing an increase in the length of the wall.

The difference in height is due to method applied for estimation of ground and building height. Ground and building heights were estimated by taking average of minimum and maximum of z-values respectively. However, the test scene also consisted of trees, few of which resulted in dense point cloud reconstruction. These points could not be removed using statistical filter as it removes points having sparse neighbourhood. This caused the overall increase in height of the model. However taking the average of the maximum possible number of values did reduce the influence due to the presence of the trees in the scene, however at the cost of increased processing time.

4.6. Limitation of the process

The modeling process gives satisfactory results if at least 3 faces of a building are covered. However, if only 2 faces are captured then the resultant shape may not match the actual structure as the result will be a triangular block.

The output of the modeling process depends on the input point cloud. If the input data consists of sparsely distributed points, then RANSAC will not be able to detect the correct planes, resulting in an incorrect or no result at all. Sparse point data is usually resulted from texture less surfaces or if there was less overlap in image pairs.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

Image-based modeling is widely used for reconstruction of three dimensional models from two dimensional images. In this research thesis, a new approach to reconstruction of 3D models was described and implemented. Instead of estimation of primitives for planer surfaces, intersection of planes was used as the basis for defining two dimensional boundary of building. Further the boundary was extruded to the estimated height of building. No prior information about the structure was used. Although the positional and shape accuracy was not very high, yet final models satisfactorily resembled the actual structure as viewed in Google Earth. Positional accuracy suffered because of the inaccuracy in the GPS locations captured by the digital cameras. If in future, cameras are equipped with better location estimation, accuracy of the process will also increase. Processing was fast and is beneficial for applications which require quick results. This methodology is suitable for applications that do not require high accuracy and have low data acquisition budget. The process is independent of data but does require an approximate estimation of distance between points for providing the value of distance threshold required in RANSAC-based plane segmentation. Moreover, at least three faces of the structure are required to be visible in the data for good results. Shape accuracy will also increase if all the faces of the structure are covered in images.

5.1.1. Answers of Research Questions

1) *What considerations should be taken into account for acquiring images?*

The images should have maximum possible overlap to reconstruct a good point cloud. Approximately 80-90% overlap is required. A good overlap covers up for the textureless surfaces. While acquiring images consideration to environmental conditions such as direction of sun, etc, should be given. Images should be sharp.

2) *What is the effect on modeling due to error in point cloud as they are reconstructed from images by SfM software?*

Point cloud reconstructed using MVS software consists of high amount of outliers. The points corresponding to a wall do not exactly lie on a plane. This might cause detection of wrong planes thus resulting in incorrect models. Another problem with outliers is caused while height estimation of the building. High amount of outliers will increase or decrease the height of the building from the actual measurement.

3) *How is the accuracy of the model affected when Helmert transformation (7 – parameter transformation) is applied for transforming it to global coordinate system using GPS coordinates with low accuracy?*

Seven-parameter transformation is a rigid transformation producing no effect to the shape of the model even if the accuracy of GPS device is low. However, the positional accuracy of model is

affected as even a slight error gets propagated and increases with every processing step. Error in transformation parameters is less when the number of images is more as compared to when number of images is less. While computing the transformation parameters the error in least square estimation step reduces with increase in the number of camera centres.

4) Which segmentation approach is suitable for planer point cloud segmentation?

RANSAC-based segmentation was chosen in this research thesis because of its simplicity and robustness. Even in the presence of outliers it estimates planes with good accuracy. However, the result is dependent on the choice of sample points selected for estimating the plane model. Greater number of iterations allowed for selection of different combinations of random sample points increasing the probability of getting better planes. But increase in the number of iterations will also increase the processing time.

5) How are occlusions and absence of data handled?

Points that represent the trees and other small objects are removed using the Statistical filter. Partial occlusion does not affect plane detection as it is based on distance of points from the estimated plane model and not point to point distance. However, if a part of building is completely occluded, for example by a fence, slight errors in modeling may get introduced due as there would be no ground plane obtained and the ground plane will have to be assumed. Similarly, absence of data also does not affect the model.

6) Under what conditions this modeling approach would be successful or fail?

The modeling approach will give good results in the following conditions:

- All the walls of the building are visible in the point cloud data.
- All the building walls have high point density, which makes detection of planes easier and more accurate.
- Outliers are less in the data. This enhances the height estimation accuracy.

The modeling approach will fail in the following conditions:

- If the number of walls visible in point data is less than three.
- If the point density is very less, i.e. points are sparsely distributed. Plane detection becomes difficult under this condition.
- High error in GPS coordinates.
- High level of outliers.

5.2. Recommendations

- In the present work, no prior information about the structure or scene was used. However, height information about the building and ground could be used to obtain more accurate models.

- In some cases, it might not be possible to obtain images of the all sides of building. In such a situation, shapefiles can be used to estimate the complete boundary.

- Combining the camera with more accurate GPS device will increase the accuracy of the model in terms of its position.

- This methodology is independent of data, therefore, can be used with crowd sourcing application where people provide images that are geo-tagged.

REFERENCES

1. Furukawa, Y., et al. *Manhattan-world stereo*. in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*. 2009.

2. Snavely, N., et al., *Scene Reconstruction and Visualization From Community Photo Collections*. Proceedings of the IEEE, 2010. **98**(8): p. 1370-1390.

3. El-Hakim, S.F., et al., *Detailed 3D reconstruction of large-scale heritage sites with integrated techniques*. Computer Graphics and Applications, IEEE, 2004. **24**(3): p. 21-29.

4. Frueh, C. and A. Zakhor. *Constructing 3D city models by merging ground-based and airborne views*. in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. 2003.

5. Pu, S. and G. Vosselman, *Knowledge based reconstruction of building models from terrestrial laser scanning data*. ISPRS Journal of Photogrammetry and Remote Sensing, 2009. **64**(6): p. 575-584.

6. Remondino, F. and S. El-Hakim, *Image-based 3D Modelling: A Review*. The Photogrammetric Record, 2006. **21**(115): p. 269-291.

7. Haala, N. and M. Kada, *An update on automatic 3D building reconstruction*. ISPRS Journal of Photogrammetry and Remote Sensing, 2010. **65**(6): p. 570-580.

8. Westoby, M.J., et al., *'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications*. Geomorphology, 2012. **179**(0): p. 300-314.

9. Lowe, D., *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, 2004. **60**(2): p. 91-110.

10. Chauve, A.L., P. Labatut, and J.P. Pons. *Robust piecewise-planar 3D reconstruction and completion from large-scale unstructured point data*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010.

11. Schindler, F., W. Worstner, and J.M. Frahm. *Classification and reconstruction of surfaces from point clouds of man-made objects*. in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 2011.

12. Bay, H., et al., *Speeded-Up Robust Features (SURF)*. Computer Vision and Image Understanding, 2008. **110**(3): p. 346-359.

13. Musialski, P., et al., *A Survey of Urban Reconstruction*. Computer Graphics Forum, 2013. **32**(6): p. 146-177.

14. Gerke, M., K. Khoshelham, and B. Alsadik, *3D Geo-Information for Imagery*. 2013, ITC, Enschede, The Netherlands, ITC Lecture Notes Block III, Module 13.

15. Fathi, H. and I. Brilakis, *Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features*. Advanced Engineering Informatics, 2011. **25**(4): p. 760-770.

16. Fischler, M.A. and R.C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Commun. ACM, 1981. **24**(6): p. 381-395.

17. Schnabel, R., R. Wahl, and R. Klein, *Efficient RANSAC for Point-Cloud Shape Detection.* Computer Graphics Forum, 2007. **26**(2): p. 214-226.

18. Furukawa, Y. and J. Ponce, *PMVS*. 2008.

19. Furukawa, Y. and J. Ponce, *Accurate, Dense, and Robust Multiview Stereopsis.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(8): p. 1362-1376.

20. Hirschmuller, H., *Stereo Processing by Semiglobal Matching and Mutual Information.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008. **30**(2): p. 328-341.

21. Goesele, M., B. Curless, and S.M. Seitz. *Multi-View Stereo Revisited.* in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. 2006.

22. Charalambos, P., *A Framework for Automatic Modeling from Point Cloud Data.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. **35**(11): p. 2563-2575.

23. Suzuki, S. and K. be, *Topological structural analysis of digitized binary images by border following.* Computer Vision, Graphics, and Image Processing, 1985. **30**(1): p. 32-46.

24. Coughlan, J.M. and A.L. Yuille. *Manhattan World: compass direction from a single image by Bayesian inference*. in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. 1999.

25. Kazhdan, M., M. Bolitho, and H. Hoppe. *Poisson surface reconstruction*. 2006.

26. Gallup, D., J.M. Frahm, and M. Pollefeys. *Piecewise planar and non-planar stereo for urban scene reconstruction*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010.

27. Dorninger, P. and C. Nothegger, *3D segmentation of unstructured point clouds for building modelling.* International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2007. **35**(3/W49A): p. 191-196.

28. Rabbani, T., F. van Den Heuvel, and G. Vosselmann, *Segmentation of point clouds using smoothness constraint.* International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2006. **36**(5): p. 248-253.

29. Borrmann, D., et al., *The 3D Hough Transform for plane detection in point clouds: A review and a new accumulator design.* 3D Research, 2011. **2**(2): p. 1-13.

30. Lari, Z. and A. Habib, *A NOVEL HYBRID APPROACH FOR THE EXTRACTION OF LINEAR/CYLINDRICAL FEATURES FROM LASER SCANNING DATA.* 2013.

31. ISPRS. *Tips for the effective use of close range digital photogrammetry for the Earth sciences.* ISPRS - Commission V - Close-Range Sensing: Analysis and Applications Working Group V / 6 - Close range morphological measurement for the earth sciences, 2008-2012 2010.

32. Wu, C. *VisualSFM : A Visual Structure from Motion System*. 2/19/2013 [cited 2013 20-8-2013]; Available from: http://homes.cs.washington.edu/~ccwu/vsfm/.

33. Falkingham, P. *Generating a Photogrammetric model using VisualSFM, and post-processing with Meshlab*. 2014/02/09/05:26:51; Available from: http://www.academia.edu/3649828/Generating_a_Photogrammetric_model_using_VisualSFM_and_post-processing_with_Meshlab

34. Lowe, D. *SIFT*. 01-02-2014]; Available from: http://www.cs.ubc.ca/~lowe/keypoints/.

35. Krumm, J., *Intersection of Two Planes*.

Appendix 1: Another dataset used and the parameters provided.

The photographs were taken of gym building in IIRS campus.

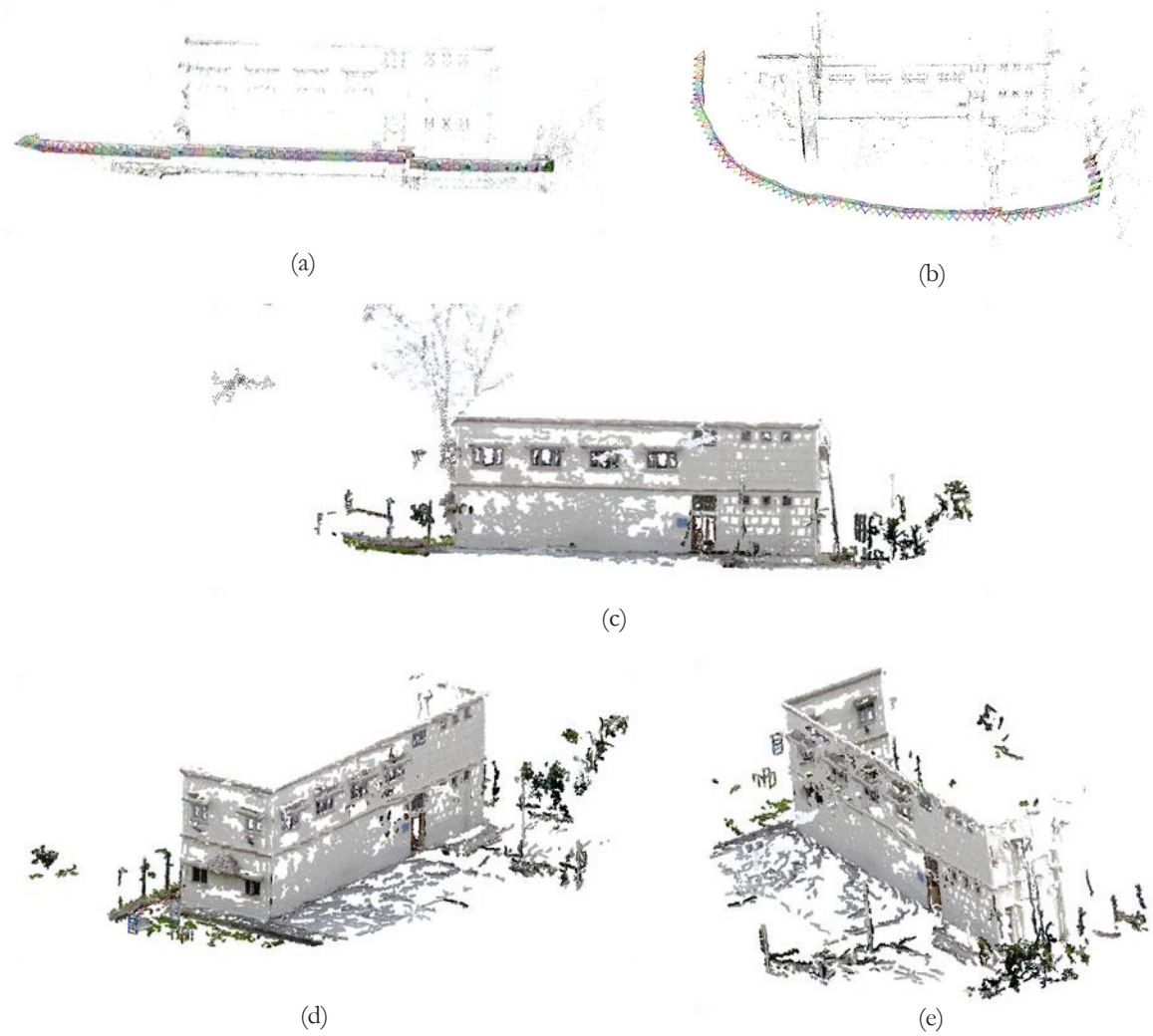
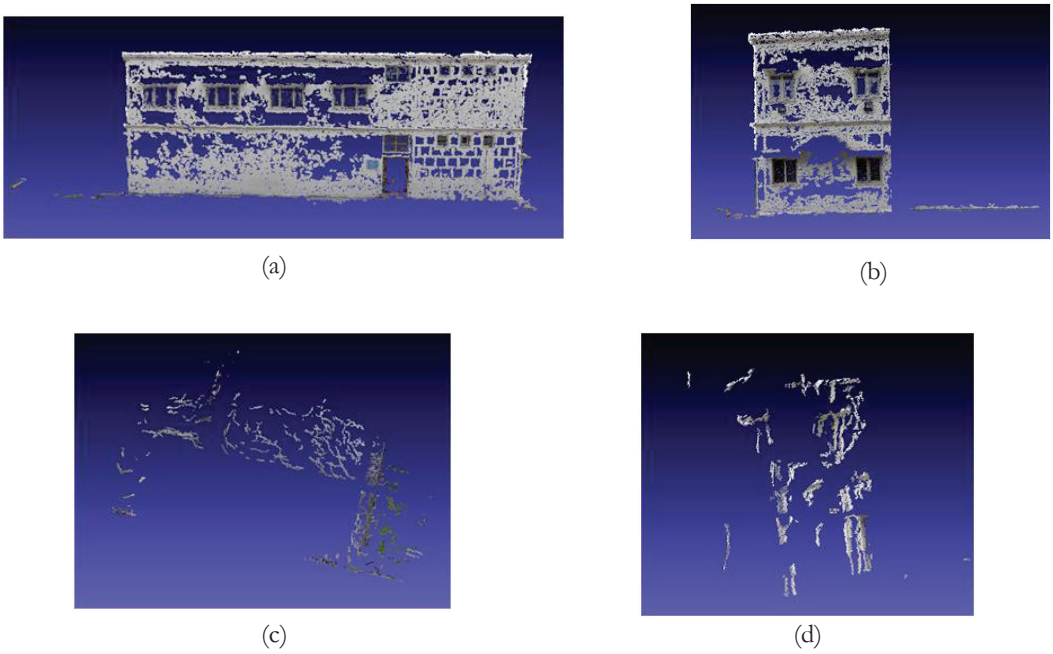


Figure 1: Gym building, (a) and (b) are sparse point cloud reconstruction; (c), (d) and (e) are dense point cloud reconstruction.

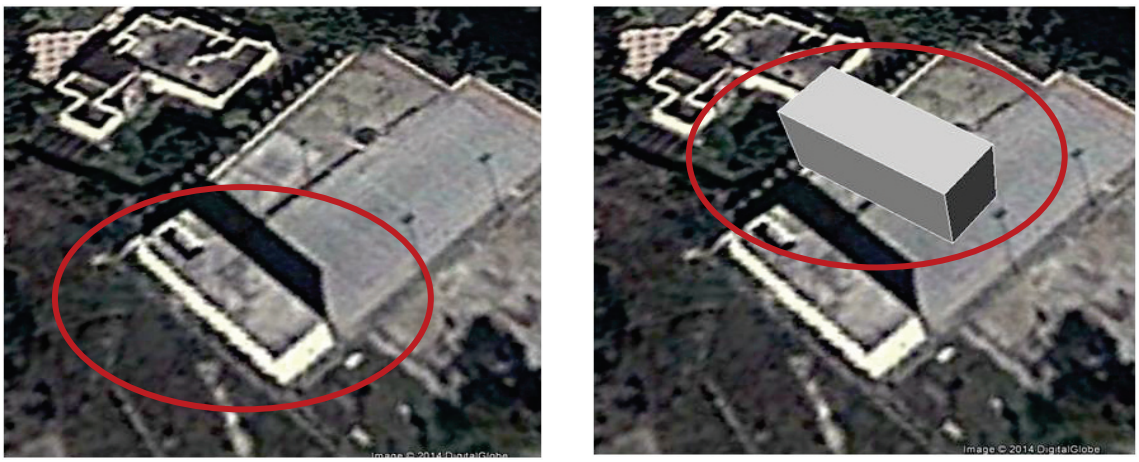
Final model obtained.



(a) (b)

(c) (d)

Figure 2: RANSAC-based segmentation; (a), (b), (c) and (d) are the planes extracted using the same methodology.



(a) (b)

Figure 3: Three dimensional box-like model of main building of IIRS viewed in Google Earth. (a) the actual location of main building marked in red circle, (b) the three dimensional model imported as KML file.

All the parameters were unchanged except RANSAC distance threshold. Value for RANSAC distance threshold was given as 0.30.

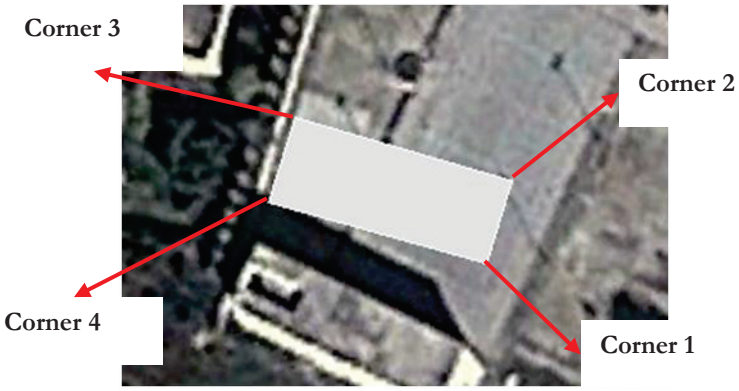


Figure 3: Distribution of corner points as used in table 2.

Appendix 2: Comparison of dimensions and coordinates of gym building.

Table 1: Comparison between dimensions of building and model

Dimension	Survey measurements (metres)	Model measurements (metres)	Difference
Length	22	20.96	-1.04
Side wall 1	7.10	7.770	0.67
Side wall 2	7	8.04	1.04
Height	9.00	9.72	0.72

Table 2: Comparison between coordinates of the corner points

Coordinates	Survey measurements (metres)		Model measurements (metres)		Distance(m)
	Easting	Northing	Easting	Northing	
Corner 1	216033.704	3360270.525	216035.6875	3360283.25	12.87866
Corner 2	216036.598	3360276.853	216037.7188	3360290.75	13.94212
Corner 3	216016.175	3360286.159	216017.9531	3360295.75	9.754435
Corner 4	216013.425	3360279.772	216015.7969	3360288	8.563047