INTEGRATION OF COMPUTER VISION ALGORITHMS AND GNSS FOR GEOREFERENCED POINT CLOUD GENERATION FROM VIDEO

Thesis submitted to the Andhra University, Visakhapatnam in partial fulfillment of the requirement for the award of *Master of Technology in Remote Sensing and GIS*

Submitted By:

Mayank Sharma

Supervised By:

Mr. S. Raghavendra

Indian Institute of Remote Sensing, ISRO,

Dept. of Space, Govt. of India Dehradun – 248001

Uttarakhand, India

August, 2014

Disclaimer

This work has been carried out in partial fulfilment of Masters in Technology program in Remote Sensing and Geographic Information System at Indian Institute of Remote Sensing, Dehradun, India. The author is solely responsible for the contents of the thesis.

Dedicated to my Parents…

Certificate

This is to certify that ***Mayank Sharma*** has carried out the dissertation entitled ***"Integration of Computer Vision Algorithms and GNSS for Georeferenced Point Cloud Generation from Video"*** in partial fulfilment of the requirements for the award of ***M. Tech. in Remote Sensing and GIS***. This work has been carried out under the supervision of ***Mr. S. Raghavendra,*** Scientist/Engineer 'SD', PRSD, Indian Institute of Remote Sensing, ISRO, Dehradun, Uttarakhand, India.

……

(Mr. S. Raghavendra)
Project Supervisor

……

(**Ms. Shefali Aggarwal**)
Head, PRSD, IIRS

……

(**Dr. S. K. Saha**)
Dean Academics and Group
Director ERSSG, IIRS

ACKNOWLEDGEMENT

I take this opportunity to express my profound gratitude to my guide Mr. S. Raghavendra, Scientist 'SD' at IIRS, Dehradun who has provided me with tremendous support throughout my thesis, and has offered his expertise, knowledge, patience, and kindness, all of which has assisted me in the completion of this thesis. Without his guidance and help, this thesis may not have been completed. I could not ask for a better supervisor.

I would especially like to express my gratitude to my all-time favourite senior Jayson Jariwala for his continuous and valuable guidance during the project work, without whom this work would not be possible. I am grateful to him for his cooperation and support during my project work. It is he, who taught me vast knowledge of computer vision and digital photogrammetry field.

I would like to thank my roommate Ankur Dixit, for the motivation and encouragement environment provided by him. His positive attitude, caring and cheerful nature supported me to complete this research work. My special thanks to Apoorva Tripathi, B.Tech student, Thapar University, for her full support and cooperation during my field work. I am highly obliged to Col. Sanjay Mohan, from Indian Army, for the valuable information provided by him whenever I needed.

I am also deeply grateful to Dr. S.K. Saha (Dean Academics and Group Director) and Shri P.L.N. Raju (Group Director, RSGG) for their comments and suggestions during the various presentations of the research work. I wish to express my sincere gratitude to Ms. Shefali Agrawal (Course Director and Head, PRSD), for motivating me to choose this topic of research and her guidance during the research work. I would also like to thank Mr. Ashutosh Bhardwaj, Scientist 'SE', IIRS, for his support and guidance. I am grateful to Dr. Y.V.N. Krishna Murthy, the Director, IIRS for providing excellent research environment and infrastructure to carry out this research work.

I would also like to express my gratitude to the Andhra University for affording me this opportunity to carry out this research work. I also wish to thank my fellow students: Anant Nautiyal, Unmesh Khati, Vaibhav Kumar and Vivek Singh for their support and camaraderie during the tenure of this research. I take this opportunity to express my gratitude towards the positive attitude and ever eagerness of the non-teaching staff in offering a helping hand needs a special mention in facilitating my research efforts.

At last, I thank my younger brother Pranay Sharma and my parents for their immense support and blessings that helped me throughout to carry out this research work.

(MAYANK SHARMA)

Declaration

I, ***Mayank Sharma***, hereby declare that this dissertation entitled ***"Integration of Computer Vision Algorithms and GNSS for Georeferenced Point Cloud Generation from Video"*** submitted to Andhra University, Visakhapatnam in partial fulfilment of the requirements for the award of ***M. Tech in Remote Sensing and GIS***, is my own work and that to the best of my knowledge and belief. It is a record of original research carried out by me under the guidance and supervision of ***Mr. S. Raghavendra,*** Scientist/Engineer 'SD', PRSD, Indian Institute of Remote Sensing, Dehradun. It contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: Dehradun Mayank Sharma

Date: Aug, 2014

ABSTRACT

Mobile Mapping is recently the most emerging area for research and development. The research in this field started in early 1990s and since then it has created a revolution in the Remote Sensing and GIS. It is an efficient and flexible approach for rapid GIS data collection. However, the disadvantage of being an expensive technique always acts as an obstacle in its further development. The integration of Computer Vision based image processing techniques along with the digital photogrammetric techniques in the field of Mobile Mapping can overcome this disadvantage. This research work focusses on the development of the technology to generate the georeferenced point cloud of an object by capturing the video of the object along with the GPS coordinates of the trajectory of the camera. This research also focusses on the development of a GPS integrated Mobile Mapping System (MMS) for rapid acquisition of the data in kinematic mode. The development of such a model is going to be highly beneficial in terms of low cost, easy and rapid data collection with less constraints for data storage. Although, the 3D georeferenced point cloud generated through the image based approach is less accurate as compared to the other terrestrial based techniques like laser scanner but it is highly cost efficient.

In this study, the image frames are extracted from the video and processed using the OSM bundler software in which the estimation of the position of exposure station and point cloud generation is done using Structure from Motion (SfM) algorithm. The georeferencing of the point cloud is done by using the collinearity equations for space intersection. Few target points are measured on the test scene by the Total Station and are used as a reference to perform the accuracy assessment of the point cloud generated using the video and images. The images are taken using different grades of conventional camera to compare their accuracy for point cloud generation. The study has shown that the point cloud generated using the image frames extracted from the video is denser in comparison to the point cloud generated using the static images. The georeferencing accuracy obtained in the point cloud generated by the video is 3.46m, 7.06m, 3.04m in Easting, Northing and Height respectively. Also it is observed that the CMOS cameras produces far better results in comparison to the CCD cameras.

Keywords: *Computer Vision Algorithms, Camera Calibration, Terrestrial Laser Scanner, Mobile Mapping, Collinearity equations, Global Positioning System, Android Application.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background

Photogrammetry and Mobile Mapping are very broad terms in themselves. Mobile Mapping and 3D reconstruction is recently the most emerging area for research and development. The research work in the field of Mobile Mapping started since last three decades and there is still much to do, (Li, 1997). The major advancements in the field of Digital Photogrammetric technology and Remote Sensing have proven it the most cost effective technique in comparison to the conventional surveying technique. The integration of computer vision based image processing techniques, digital photogrammetry and Mobile Mapping have created a milestone in the field of 3D GIS. Less time consumption, easy data acquisition, and high flexibility are the certain advantages that are the results of the development in the Mobile Mapping technologies, (Li, 1997) . Also the Videogrammetry i.e. capturing the videos of an object to perform some measurement related tasks is in huge demand these days. It is currently the best way to overcome the limitations of the conventional methods used in terrestrial photogrammetry, (Fathi & Brilakis, 2011).

This research work focusses on the development of the technology to generate the georeferenced 3D point cloud of an object by capturing the video of the object along with the GPS coordinates of the trajectory of the Mapping sensor. "*A point cloud is a set of data points in some coordinate system*", and each point in the point cloud has a well-defined three dimensional position in that coordinate system, ("Point cloud," 2014).This research also focusses on the development of a GPS integrated Mobile Mapping System (MMS) for rapid acquisition of the data in kinematic mode. The development of such a model is going to be highly beneficial in terms of low cost, easy and rapid data collection with less constraints for data storage. This project work attempts to find out the best possible solution for the 3D reconstruction at a very low cost, using the conventional camera for data acquisition and open source software for the data processing. Computer Vision Algorithm in mobile mapping is been largely attempted for 3D reconstructs and trajectory estimation. Apart from it, it is also used in various well known fields like games, image processing, artificial intelligence and augmented reality. Although, the 3D georeferenced point cloud generated through the image based approach is less accurate as compared to the other terrestrial based techniques like laser scanner but it is highly cost efficient, (Yang, Chao, Huang, Lu, & Chen, 2013). Furthermore, the following sections will provide the details about the technologies used in this research work.

1.1.1Mobile Mapping

Mobile Mapping is the most emerging technique for spatial data acquisition in the field of GIS and gaining the high importance because of the advantages like time consumption is less, easy data acquisition and high flexibility, (Li, 1997). As the name signifies, the mobile mapping comprises of a moving or mobile platform, navigation (GPS, INS etc.) and mapping sensors. An aircraft or any other land based vehicle can be used as the mobile platform. To capture the information of the objects to be surveyed, mapping sensors are employed that can be metric or

non-metric cameras, laser scanners or radars. Navigational sensors like Global Positioning Systems (GPS) and Inertial Navigation Systems (INS) are used to provide the position and orientation information of the mapping sensor, (Li, 1997). Figure 1.1 shows a common example of MMS.

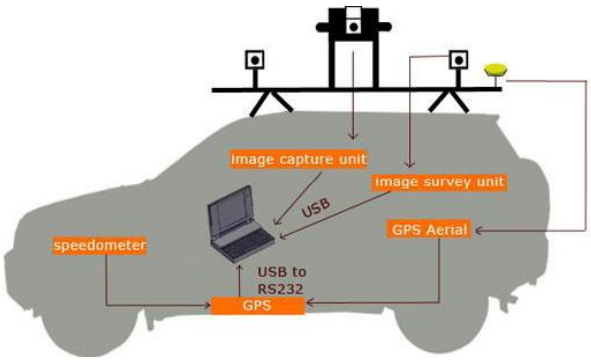


Figure 1.1 Mobile Mapping System Architecture(Source: http://www.easypano.com/images/city8/street-view-car-structure.jpg)

1.1.2 Computer Vision Algorithms

Even after the tremendous development in the technology towards virtual world and 3D visualization, there is no comparison of any technology with human vision. We (human beings) have the capability to perceive the objects around us in 3D very easily and can judge the shape, size, texture and pattern of the objects. Computer vision algorithms can be defined in many ways like Computer Vision is an area that deals with the development of algorithms and other mathematical techniques to reconstruct the three dimensional structure of the objects through the two dimensional images, (Szeliski, 2010), "*Computer Vision is a field that includes methods for acquiring , processing, analyzing and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions*", ("Computer vision," 2014).

Structure from Motion (SfM) is among the most widely used computer vision algorithms for 3D reconstruction and trajectory estimation. The SfM approach is based on the principle to form the relation between the images by applying relative and absolute orientation. The use of the photogrammetric techniques in SfM enable the estimation of the camera parameters, (Yang et al., 2013). The ability of SfM to estimate the camera parameters (interior and exterior orientation parameters) is of main importance, as it can directly replace the high cost Inertial Navigation Sensors (INS). Scale Invariant Feature Transform (SIFT) for feature detection, Approximate Nearest Neighbors (ANN) for feature Matching and Random Sample Consensus (RANSAC) for outlier Removal are also used in SfM.

SIFT algorithm is used for detection and description of the features present in the images, (Lowe, 1999). The extracted features from the images are further stored into the database for the estimation of the camera motion and tracking purpose. "*This approach has been named the Scale Invariant Feature Transform (SIFT), as it transforms image data into scale-invariant coordinates relative to local features*", (Lowe, 2004). ANN or Approximate Nearest Neighbor is used to

enhance the feature matching process between the detected features in the pair of images and thus leads to identify the motion of the camera, (Yang et al., 2013). RANSAC algorithm is used for the removal of outliers and for finding the F-matrix for all the pair of images, (Yang et al., 2013). RANSAC is highly efficient and robust method for outlier removal which works on the basis of model fitting approach. It takes a small set of data and analyses the parameters of the predefined models. If the parameters obtained by the taken data set do not satisfies the predefined model, then data points are treated as outliers, (Fischler & Bolles, 1981).

1.1.3 Navigational Sensors

The navigational sensors play a vital role in Mobile Mapping for accurate positioning and orientation information of the mapping sensors. INS or Inertial Navigation Sensor and *Global Navigation satellite System (GNSS)* are the major navigational sensors implemented in Mobile Mapping systems for orientation estimation and proper positioning of the mapping sensors. INS is a high cost device used for getting the orientation parameters of the mapping sensor. Accelerometers and gyroscopes are used in INS to find the orientation parameters of the mapping sensor. Its accuracy decreases with the increase in time. GNSS is the system of satellites to find out the location or position of any point anywhere in the world. The various global navigation satellite systems that are available these days are NAVSTAR GPS of USA, GLONASS of Russia, BEIDOU of China, and IRNSS of India, ("Global Positioning System," 2014). Among all these mentioned navigation systems, Global Positioning System (GPS) of USA is widely used. It was developed by US Department of Defense for military operations and is available for civilians with very low accuracy, (*GPS Positioning Guide*, 1995). All these systems require high cost receivers and complex processing for accurate positioning and navigational information.

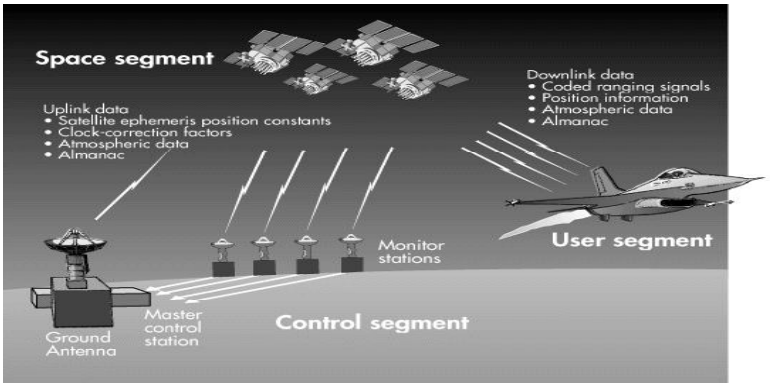


Figure 1.2 GPS Segments(Marry E. Greece, 2000)

These navigation systems are divided into three segments

- Space Segment
- Ground Control Segment
- User Segment

Space segment consists of the group of satellites in the space and are dispersed in such a way that a minimum of 4 satellites are visible from any point on the earth. These satellites are continuously

monitored by the ground control stations, (*GPS Positioning Guide*, 1995). Figure 1.2 shows the 3 segments of GPS.

Ground segment consists of the controlling stations on the earth, which continuously monitor these satellites. These stations are responsible for determining the location of the satellites in space and confirming that they are suitably operational. If a satellite is not working suitably, then, the ground control station marks its status as "unhealthy" and changes it to "healthy" as soon as it becomes suitably operational, (*GPS Positioning Guide*, 1995).

Anyone who handles the GPS signal receivers to find out the locational information comes under the user segment. A large variety of GPS signal receivers is available. Even, the mobile phones are well equipped with the GPS facility these days. The antenna, processing and display unit, power supply, RF section are the major constituents of GPS receivers, (*GPS Positioning Guide*, 1995).

The positioning methods of GPS are mainly single point positioning and relative positioning, (*GPS Positioning Guide*, 1995). Both these methods are distinguishable in terms of working as well as accuracy. In single point positioning, the coordinates of an unknown are determined in a defined projection system by the help of the known positions of at least 4 satellites, however, in relative positioning, the receiver coordinates of the unknown point are determined in reference to the receiver coordinates of a known point. Relative positioning is also known as Differential Positioning or Differential GPS (DGPS), (*GPS Positioning Guide*, 1995). In Relative Positioning, the accuracy is very high as compared to the single point positioning method.

GPS positioning methods are also classified as static and kinematic positioning method. The GPS receiver is kept fixed in the static mode while the receiver is in the state of continuous motion in kinematic mode. The processing of the GPS data can be done either in the real time as well as after the acquisition of whole data is done, (*GPS Positioning Guide*, 1995).

1.1.4 Android

Android-a well-known operating system for touch screen mobiles, tablets PCs is primarily designed on Linux kernel, ("Android (operating system)," 2014). It was developed by the Android Inc. which was founded by Andy Rubin (co-founder of Danger), Rich Miner (co-founder of Wildfire Communications, Inc.), Nick Sears (once VP at T-Mobile), and Chris White (headed design and interface development at WebTV) and later it was supported by the Google, ("Android (operating system)," 2014). It supports a wide range of devices and applications. It has millions of applications that are freely downloadable from Google Play. Its fully customizable feature has attracted a lot of application developers towards open source technology.

1.2 Motivation and Problem Statement

Earlier the 3D reconstruction is mainly done by the aerial photogrammetry and LiDAR technology. The implementation of the terrestrial photogrammetry for 3D reconstruction was quite difficult, (Yang et al., 2013). The evolution of Digital Photogrammetry and its integration with modern image processing techniques has given a new direction to the researchers. It has

resulted in the simplification of the sophisticated photogrammetric techniques, such that complex photogrammetric task can be accomplished quickly and easily. Now it is possible to use the non-metric or conventional cameras for mobile mapping and 3D modelling such that a highly cost effective solution is now available to the users. This all is the result of the advancement in the field of digital photogrammetry and image processing techniques. Also with the help of terrestrial mobile mapping and the close range photogrammetry, the reconstruction of the outer facades of the building is quite easy as it provides much information of the building facades as compare to the aerial photogrammetry,(Tian, Gerke, Vosselman, & Zhu, 2010). Although the Terrestrial Laser Scanners are highly efficient in dense 3D point cloud generation, however, they are very expensive. The development and use of open source software in photogrammetric applications has made it a truly cost effective technique for mobile mapping and 3D modelling.

In most of the earlier Mobile Mapping systems the equipment used are very costly, for example the VISAT (Tao, Chapman, & Chaplin, 2001) system developed by the Department of Geomatics engineering at The University of Calgary, produces good results with high accuracy however uses a range of mapping and navigational sensors (GPS & INS) that are very costly. These technologies require a large and complex set up which makes the installation and operational expenditure very high. Also the Photogrammetric software available commercially are very expensive.

This research work focusses on the development of the cost effective technology for the generation of the georeferenced 3D point cloud of a scene from a video captured through the consumer grade cameras along with the positional information recorded by the GPS. The use of consumer grade cameras and the open source software make it truly a cost effective and beneficial technology. The motivation for this project is as follows:

- *GPS Outage Problem and Accuracy of the Navigation Sensor*: There might be the situations when the GPS signals are not available or having poor GDOP in the intermittent areas of MMS trajectory. So, the actual position and orientation of the points cannot be determined, (Hassan, Ellum, & El-Sheimy, 2006). Mostly in the urban areas due to dense, high and complex structure of buildings, the navigation sensors do not provide high accuracy or possess poor dilution of precision(DOP). In this case, the coordinates of the exposure station cannot be determined accurately. Thus it is difficult to perform the georeferencing of the 3D point cloud. Therefore a methodology needs to be developed to find the coordinates of the exposure station in the intermittent GPS unavailable areas.
- *Image Overlap Problem*: As per the literature available most of the work in close range photogrammetry is carried out by capturing the images from different locations. So, there might be the situation of no or less overlap between the successive images. A video is a sequence of image frames that are moved at a very high rate up to 8 to 25 frames per second or even more. Thus, the video coverage of the area will provide the solution to ensure enough overlap.
- In this project, a low cost solution for 3D georeferenced point cloud using consumer grade camera and GPS is attempted and evaluated.

The cost effectiveness, easy and rapid acquisition of data are the major advantages of terrestrial mobile mapping. The point cloud generated by the terrestrial mobile mapping image sequences provide finer results as compared to those generated with aerial imagery. It is also beneficial when compared to the data acquisition with ground based Laser scanners which is very expensive process, however, the point density is not as good as that of the Terrestrial Laser Scanner. The use of open source software is really the main advantage of this research work. A wide range of open source software are used in this research work for data processing. This project attempts to develop a common platform to integrate all the tasks in a single software to reduce the processing time. This forms the basis to easy, rapid and cost effective approach for the geospatial data collection.

1.3 Research Identification

1.3.1 Research Objective

To apply the Bundle block adjustment for the integration of SfM technique with GNSS for direct geo-referencing of point cloud.

1.3.2 Sub Objectives

- Developing methodology for field planning using CARTOSAT 1 DEM for data acquisition.
- To generate the geo-referenced point cloud from video (Frame of Images) using SfM technique.
- Developing methodology for generation of geo-referenced point cloud from GNSS unavailable intermittent camera stations.
- To develop an Android based App for Geotagging and GUI based Tool for point cloud Processing.

1.3.3 Research Questions

- Does video frame sampling improve the performance of SfM –GNSS integration over the static images?
- How to geo-tag the video frames?
- What is the effect of changing the camera resolution in SfM-GNSS integration?
- How to geo-reference the point cloud from camera station under GNSS unavailable environment?

1.4 Thesis Structure

The research work is organized as follows:

Chapter 1: Introduction

It explains about the general aspects and scope of the research work along with a little description of the past work done in this field. It covers the background, motivation and problem statement, research objectives, research questions of the project work.

Chapter 2: Literature Review and Theoretical Concepts

It explains the little background of the research field and the summarized description of the previous research in this field. It also explains the theoretical concepts implemented in this research work.

Chapter 3: Material and Methodology

This chapter gives a brief description of the study area in which the research is carried out. It gives the details of the software and hardware used in this research work. It also explains the whole workflow or methodology that is adopted to carry out this research work. It gives the detailed explanation of the processes involved in this research work.

Chapter 4: Results and Discussion

This section gives the description of the obtained results and provides the answer to each research question along with the discussion in support to that answer.

Chapter 5: Conclusion and Recommendation

It gives the answer of the research questions in concluded form and recommendations for further study.

CHAPTER 2: LITERATURE REVIEW AND THEORETICAL CONCEPTS

2.1 Literature Review

2.1.1 Mobile Mapping System

The work in the field of Mobile Mapping started in the early 1990s. The major advancements in the field of Digital Photogrammetric technology and Remote Sensing have proven it the most cost effective technique in comparison to the conventional surveying and mapping technique. Now days, mobile mapping is getting the popularity in both aerial and terrestrial mapping applications. High accuracy, fast acquisition of data and cost effectiveness are its major advantages, (Li, 1997).

Mobile Mapping always being a significant area for research and development have seen a lot of advancements since last three decades. In 1992, Department of Geomatics Engineering at The University of Calgary along with the Geofit Inc. started developing a Mobile Mapping System (MMS) called VISAT (Video, Inertial and Satellite GPS) and was implemented in 1995, (Tao et al., 2001).

(Tao et al., 2001) explains the automatic way to process an array of images implementing the visual motion theory. Image and object domain limits are especially considered. Various methods for automatic processing of image sequences are classified into two parts namely as Information extraction and Image based trajectory determination or Bridging, (Tao et al., 2001).

(Hassan et al., 2006) utilizes the VISAT MMS of The University of Calgary and explained the advantages of the land based MMS over aerial photogrammetry and traditional surveying. This paper presented the photogrammetric approach to georeference the array of images captured through MMS when there is unavailability of GPS signals or weak signals are available. Even the proposed method produces good results, however, the use of VISAT MMS makes it an expensive approach because of its complex setup and high cost instruments like INS.

2.1.2 Computer Vision Algorithms and Image Based 3D Reconstruction

Computer Vision Algorithm in mobile mapping is been largely attempted for 3D reconstructs and trajectory estimation. Apart from it, it is also used in various well known fields like games, image processing, artificial intelligence and augmented reality. According to (Szeliski, 2010), Computer Vision is an area that deals with the development of algorithms and other mathematical techniques to reconstruct the three dimensional structure of the objects through the two dimensional images. Computer Vision Algorithm especially Structure from Motion (SfM) is implemented in this research work to estimate the camera parameters. There are two types of camera parameters, one is intrinsic parameters and the other is extrinsic parameters. Intrinsic parameters are obtained through camera calibration and for obtaining the extrinsic parameters SfM is used. Conventionally, high cost Inertial Navigation System (INS) is used to obtain the

camera orientation information. Thus, SfM is implemented to make the technique cost effective. All these terms are explained in the upcoming Theoretical Background section.

3D reconstruction is recently the most emerging area for research. There's a plethora of books and research articles about image based 3D reconstruction. Earlier the 3D reconstruction is mainly done by the aerial photogrammetry and LiDAR technology. The implementation of the terrestrial photogrammetry for 3D reconstruction was quite difficult, (Yang et al., 2013). The evolution of Digital Photogrammetry and its integration with modern image processing techniques has given a new direction to the researchers. It has resulted in the simplification of the sophisticated photogrammetric techniques, such that complex photogrammetric task can be accomplished quickly and easily. Now it is possible to use the non-metric or conventional cameras for mobile mapping and 3D modelling such that a highly cost effective solution is now available to the users. This all is the result of the advancement in the field of digital photogrammetry and image processing techniques.

Camera Calibration is the most important step for the 3D reconstruction based on the image sequences. Several techniques have been developed to perform the automatic camera calibration. (Zhang, 2000) proposed a flexible and easy technique to perform the camera calibration with modeling of radial lens distortion. In this approach a planar design is observed by the camera from at least two distinct orientations.

The disadvantages such as ignoring lens distortion, knowledge of calibration environment in the camera calibration techniques based on single images are described by (Zhou, Cui, Gao, & Wang, 2013). To overcome these advantages, (Zhou et al., 2013) proposed a new technique for camera calibration especially considering the lens distortion parameters.

(Pollefeys, Koch, & Van Gool, 1999) proposed a self-calibration technique for the estimation intrinsic camera parameters. The orthogonality of image rows and columns is the common constraint for this approach.

(Karras & Mavrommati, 2002) describes a simple calibration technique for non-metric cameras. Several regular grids as well as manmade linear features are used in this technique. Simple adjustment of images captured from different locations is done and the radial distortion is modeled.

(Han, Paulson, & Wu, 2011) illustrates the procedure for 3D reconstruction from two dimensional (2D) video sequence. Surface fitting approach is suggested for 3D reconstruction in order to preserve the geometric structures. To segment the 3D point cloud into proper 3D geometric structures (Han et al., 2011) developed an expanded deterministic annealing algorithm. Present Feature extraction, feature matching, and projective reconstruction techniques are involved for 3D reconstruction, (Han et al., 2011).

3D Z-string-a knowledge structure for tracking the motion of an object and change in its size, obtaining the spatial and temporal relation between the objects in a video is

developed by (Lee, Yu, Chiu, & Hong, 2005). This technique produces good result and is very effective in terms of storage and processing time, (Lee et al., 2005).

An aerial mapping system with digital camera and GPS is described by (Novak, 1993). GPS is operated in kinematic mode to continuously record the position of the exposure station. Aerotriangulation is applied to find out the other camera parameters and then finally the data is processed, (Novak, 1993).

(Fathi & Brilakis, 2011) describes a method to generate sparse 3D point cloud using two videos captured through the calibrated cameras. Stereo vision approach is applied for 3D sparse point cloud generation. SURF algorithm for feature detection and matching, RANSAC is used for outlier removal and finally the bundle adjustment is applied for automatic registration of point clouds, (Fathi & Brilakis, 2011).

Different techniques for surface reconstruction like SfM, Clustering View for Multi-View Stereo (CMVS), Patch based Multi-View Stereo (PMVS) and Poisson surface reconstruction are clearly discussed by (Yang et al., 2013). 3D reconstruction from the images taken by conventional digital cameras, smartphones is well explained. The results are verified on various indoor and outdoor objects, (Yang et al., 2013).

(Tian et al., 2010) explains an automatic way for the 3D reconstruction from a video. They have implemented the structural knowledge of buildings into the reconstruction in order to have the topological relationships between the reconstructed points and edges. The mentioned procedure for reconstruction applies to complex building too and helps in obtaining the building structures in the areas of obstruction, (Tian et al., 2010).

(Lowe, 2004) developed an algorithms for feature detection called Scale Invariant Feature transform (SIFT). The detected features are independent of the scale of the image and rotation. These diverse detected features can be used for feature matching and identification, (Lowe, 2004). This patented algorithm is accepted worldwide and provides good accuracy along with the advantages of scale and rotation independency.

Multiple images of an object, taken from distinct points are used by (Jayson Jayeshkumar Jariwala, 2013) to obtain the 3D point cloud using the SfM approach. Camera positions are obtained through the DGPS survey. The mentioned approach uses the space intersection technique to georeference the 3D point cloud.

(Tissainayagam & Suter, 2005) describes the procedure implementing the Bayesian Multiple Hypothesis Tracking (MHT) method to track the objects in the sequence of images. Two MHT based algorithm, one for time-based tracking of a particular object and another for contour segmentation are used. Several key points are detected using edge map on an object to represent it and are tracked on the sequence of images. The major disadvantage of this approach is of contour grouping that occurs due to the blocking of the object on the image scene.

Bundle adjustment is used to generate the optimal 3D geometries with the help of camera parameters. To overcome the non-linearity and localization problem during optimal 3D

generation, (Lourakis & Argyros, 2009) developed a Sparse Bundle Adjustment Package (SBA) based on C/C++.

(İmre et al., 2007) developed the system for 3D reconstruction from a broadcast video, presuming the relative motion between the camera and the object. Ability to estimate the internal and external camera parameters, estimation of dense depth field and performing 3D model based segmentation are its distinctive advantages. However, its accuracy degrades in presence of small independently moving objects especially in indoor sequences, (İmre et al., 2007).

2.2 Theoretical Background

All the keypoints or major terms used in this research work are explained in the following subsections:

2.2.1 Camera Calibration

Camera calibration is the most important procedure for 3D reconstruction using computer vision algorithms. Generally there are two type of cameras, one is metric and the other is non-metric. Metric cameras are expensive and are mostly used for photogrammetric applications. Their camera parameters are accurately known and are fixed. However, the Non-metric cameras are not suitable to carry out the image based measurement until they are well calibrated or their camera parameters are known. Camera calibration is necessary so as to obtain the geometrical characteristics of the camera. In simpler terms it is a method of retrieval of camera parameters that are in the image coordinate system, (Horn, 2000). Intrinsic and Extrinsic parameters are two types of camera parameter. Intrinsic parameters are obtained through camera calibration while SfM is used to obtain the extrinsic parameters. According to (Zhang, 2000), a camera is always modeled as a pin hole camera model. A pin hole camera model is used to explain the link between a point in 3D world coordinate system and its location in 2D image plane mathematically, ("Pinhole camera model," 2014). In terms of pinhole camera model, camera calibration is done for finding out the parameters to establish a link between the features on the ground to their two dimensional representation on an image, (Zhou et al., 2013). These parameters are used to express the 3D position of a point in world coordinate system to camera coordinate system. A simple pinhole camera model is shown in Figure 2.1 below,

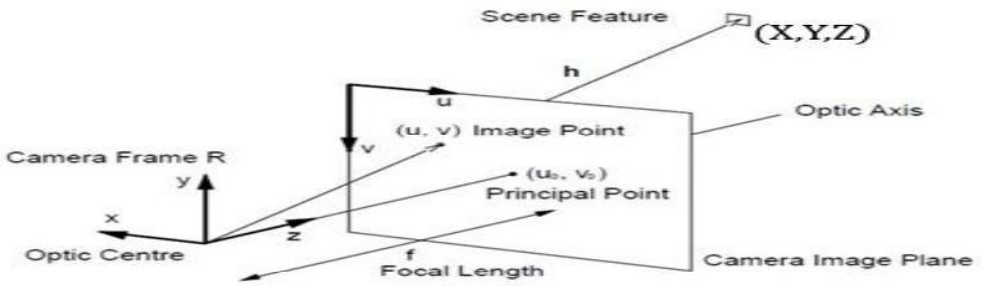


Figure 2.1 Pin hole Camera Model (Source: http://perso.ensta-paristech.fr/~filliat/Courses/2011_projets_C10-2/BRUNEAU_DUBRAY_MURGUET/monoSLAM_bruneau_dubray_murguet_en.html)

According to (Zhang, 2000), a pinhole camera model is expressed as in equation 2.1.

$$s\begin{bmatrix}u\\v\\1\end{bmatrix} = A_{int}[R \quad T]\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{2.1}$$

Here s is the scale factor, $[u \quad v \quad 1]^T$ is the augmented vector representing the position of a point in 2D image, A_{int} is the intrinsic parameter matrix, $[R \quad T]$ are the rotational and translational parameters, $[X \quad Y \quad Z \quad 1]^T$ is the position of the same point in 3D space. Rotational parameters, translational parameters and scale are the extrinsic parameters. Intrinsic and Extrinsic parameters are explained below:

2.2.1.1 Intrinsic Parameters

Intrinsic parameters are those that are used to determine the internal geometry of the camera and its visual features, (Zhou et al., 2013). Focal length, principal point location and lens distortion parameters are the intrinsic parameters that are retrieved in the process of camera calibration and are used for the interior orientation of the camera, (Horn, 2000). These are generally expressed as shown below in equation 2.1

$$A_{int} = \begin{bmatrix}f_x & 0 & u_0\\0 & f_y & v_0\\0 & 0 & 1\end{bmatrix} \tag{2.2}$$

Here f_x and f_y are the scale factors in x and y axis of image, (u_0, v_0) are the location of the principal point, and skew factor is considered zero.

2.2.1.2 Extrinsic Parameters:

Extrinsic parameters are used to determine the relationship between the world coordinate system to the camera coordinate system, (Zhang, 2000). It can also be explained as the parameters to determine to determine the position and orientation of the camera with respect to a world coordinate system, (Zhou et al., 2013). Rotational parameters R, translational parameters T and scale s are the extrinsic parameters. These parameters are used for the transformation of a point from camera coordinate system to the world coordinate system, (Zhou et al., 2013). Here SfM is used to obtain the extrinsic parameters which is explained in section 2.2.3.

Intrinsic and Extrinsic parameters together constitutes the Projection matrix. The transformation of the pixel coordinate system to the world coordinate system is done with the help of projection matrix as mentioned in equation 2.3, (Jayson Jayeshkumar Jariwala, 2013).

$$x = A_{int} * A_{ext} * X \tag{2.3}$$

Here A_{int} is the intrinsic parameter matrix, A_{ext} is the extrinsic parameter matrix.

$$x = A_{int} * R[I| - T] * X \tag{2.4}$$

Here I is the identity matrix.

$$x = PX \tag{2.5}$$

Here P is the projection matrix.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} * \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.6)$$

As per the literature available, there are various techniques available to perform the camera calibration. According to (Zhang, 2000), all the techniques for camera calibration can be categorized into Photogrammetric camera calibration and Self-calibration. An object whose 3D geometrical information is known accurately is used to perform the Photogrammetric camera calibration. However, due to the requirement of exclusive calibration arrangement, this approach is not so common, (Zhang, 2000). Self-calibration can be performed by capturing at least three images of a stationary scene with fixed internal camera parameters. Image data alone is sufficient to obtain camera parameters without any need of calibration object. This technique is also not popular as it does not provide the accurate results, (Zhang, 2000). A new method independent of the motion of the camera or the scene is proposed by (Zhang, 2000) to perform the camera calibration in which a 2D plane is captured by the camera from at least two distinct angles.

(Zhou et al., 2013) discussed the disadvantages of the single view based procedure to perform the camera calibration. Flouting the lens distortion and depending upon a particular situation for calibration were certain disadvantages with single view method. To overcome these problems (Zhou et al., 2013) proposed a new technique called "*Line based camera calibration with lens distortion correction from a single image using three squares of unknown length*". Firstly the radial distortion parameters were estimated and the inaccurate lines associated with the squares were rectified at the same time. Intrinsic parameters were estimated through the Homography matrix calculated using these lines, (Zhou et al., 2013).

2.2.3 Structure from Motion

Structure from Motion (SfM) is among the most widely used computer vision algorithms for 3D reconstruction and trajectory estimation. The SfM approach is based on the principle to form the relation between the images by applying relative and absolute orientation. The use of the photogrammetric techniques in SfM enable the estimation of the camera parameters, (Yang et al., 2013). The ability of SfM to estimate the camera parameters (interior and exterior orientation parameters) is of main importance, as it can directly replace the high cost Inertial Navigation Sensors (INS).

SfM acts as a medium to integrate the fields of Computer vision and Photogrammetry. By obtaining the pattern of the camera and 3D points, SfM can estimate the position and orientation of the exposure station along with the geometrical properties of the scene, (Yang et al., 2013). (Yang et al., 2013) mentioned the three main tasks of SfM as follows:

- Motion investigation of camera and target detection.
- Estimating the trajectory of moving camera.
- 3D reconstruction of the scene.

In SfM, firstly the feature detection is done using Scale Invariant Feature Transform (SIFT). SIFT algorithm is used for detection and description of the features present in the images, (Lowe, 1999). The extracted features from the images are further stored into the database for the estimation of the camera motion and tracking purpose. ANN or Approximate Nearest Neighbor is used to enhance the feature matching process between the detected features in the pair of images and thus leads to identify the motion of the camera, (Yang et al., 2013). RANSAC algorithm is used for the removal of outliers and for finding the Fundamental matrix for all the pair of images, (Yang et al., 2013). As determined by the Fundamental matrix, the primarily accurate image pairs are chosen for initializing the reconstruction process and the process of adding the new cameras starts to enhance the sparse bundle adjustment for the generation of sparse 3D point cloud, (Yang et al., 2013). All these processes and their theoretical background is given below:

2.2.3.1 Epipolar Geometry and Fundamental Matrix

Fundamental matrix is used to express the relationship between the same points imaged in two different images. It estimates the projection information of a point in one image to the corresponding point in the other image. In simpler terms, Fundamental matrix expresses the epipolar geometry. *"Epipolar geometry is the stereo view of two cameras when cameras view a 3D object from two different locations"*,(Jayson Jayeshkumar Jariwala, 2013).

2.2.3.2 Feature detection and matching using SIFT

Scale invariant feature transform (SIFT) was developed by David Lowe in 1999 and it is used to convert the image into the image into a group of local descriptors or features, ("Scale-invariant feature transform," 2014). "*This approach has been named the Scale Invariant Feature Transform (SIFT), as it transforms image data into scale-invariant coordinates relative to local features*", (Lowe, 2004). David Lowe's method for feature detection creates a database in which the local features called keypoints, found in the reference image are stored. Then the features in the new image are compared with this database and are matching is done on the basis of Euclidean distance of their feature vectors.

The basic steps in SIFT as mentioned by (Lowe, 2004) are shown in the block diagram below in Figure 2.2:

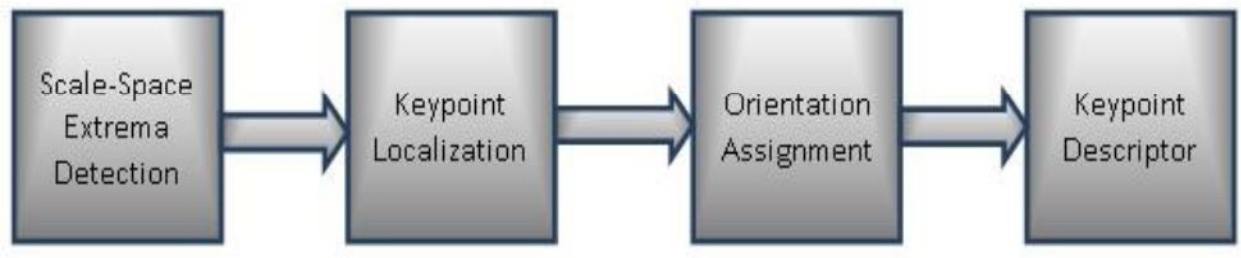


Figure 2.2 Block Diagram of SIFT

Detection of positions of interest that are independent of the variations in scale is the initial step in the process of scale space extrema detection and this is done by implementing the scale space

kernel for examining the features which are steady across all the probable changes in scales, (Lowe, 2004). (Lowe, 1999) describes the method that uses the Difference of Gaussian (DoG) scale space kernel for effective feature detection. To obtain the scale space images separated by a constant factor, the starting image is increasingly convolved with the Gaussian scale space function and then the neighboring images are subtracted to produce the DoG images, (Lowe, 2004). This procedure is shown below in Figure 2.3

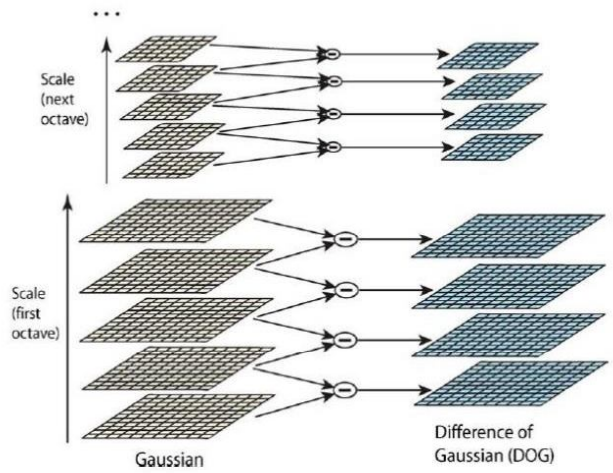


Figure 2.3 Generation of DoG images (Source: (Lowe, 2004))

Each pixel in an image is compared to its neighbors in the 3×3 region in the same image as well as in the adjacent scale space images above and below, conclusively each pixel is compared with 26 neighbors to obtain the maxima or minima of DoG images,(Lowe, 2004). If the result of comparison shows that the pixel is either maximum or minimum of its neighbors then it is marked as the candidate key point.

Keypoint localization is performed by the technique proposed by (Matthew Brown & David Lowe, 2002). For the localization of Keypoints, a 3D quadratic function is fitted to the local sample points in order to determine the estimated position of the maxima by applying the Taylor's expansion the Difference of Gaussian scale space function. The points having the low contrast and resides on the edges are discarded. The obtained value of the second order Taylor expansion of DoG scale space function is used to reject the points with low contrast if it is found less than the predefined threshold. Eigen values of the second order Hessian matrix (H) are calculated to obtain the principal curvatures which helps in determining the response of the DoG scale space function along and across the edges. Here only the ratio of the Eigen values is significant. The trace and determinant of the matrix are used to evaluate the Eigen values, (Lowe, 2004). (Lowe, 2004) explained it mathematically as

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$

Here D is the DoG scale space function

$$R = \frac{Tr(H)^2}{Det(H)} = \frac{(r+1)^2}{r}$$

Where $Tr(H)$ is the trace of (H) and $Det(H)$ is the determinant of (H) and are expressed as

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta$$

$$Det(H) = D_{xx}D_{yy} - \left(D_{xy}\right)^2 = \alpha\beta$$

$$r = \frac{\alpha}{\beta}$$

Here α and β are the Eigen values of (H) and r is their ratio.

So if for a candidate keypoint R is greater than $\frac{(r+1)^2}{r}$, then it is poorly localized and hence rejected.

Based on the local image characteristics an orientation is assigned to all the keypoints. With reference to this orientation a keypoint is made independent of the image rotation. This is done by obtaining the gradient magnitude and orientation for each pixel around the candidate keypoint. This gradient orientation data is used to plot an orientation histogram of 36 bins in which each added sample is weighted by gradient magnitude and Gaussian weighted circular window. Gradient image patch weighted by Gaussian window is shown in Figure 2.4.

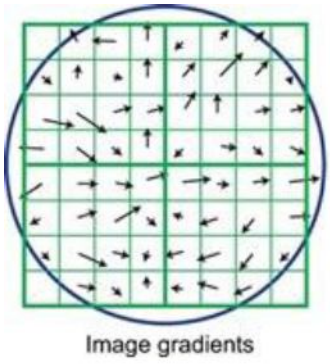


Figure 2.4 Gradient image patch (Source: (Lowe, 2004))

The direction of the local gradient are determined by the peaks in the histogram. The orientation is assigned to the highest peak and to the local peaks having the value equal to or within 80% of the highest peak. Many keypoints at same location are assigned multiple orientation because of the multiple peaks of identical magnitude. This confirms the image scale, rotation and position invariance. To obtain the more accurate results, a parabola is fitted to the 3 histograms values neighboring to each peak to obtain the interpolated peak position, (Lowe, 2004).

A keypoint descriptor is estimated for each keypoint which is independent of the illumination and viewing direction. It is already mentioned that the gradient image region around the sample keypoint is weighted by the Gaussian window. Descriptor coordinates are rotated by the orientation angle of the keypoint to gain rotation independency. Sub division of the region around the keypoint is done into 4×4 squares. For each sub divided region an 8 bin orientation histogram is plotted. This constitutes to the 16 sub orientation histogram which are used to form a 128-D vector called SIFT descriptor and is used for similarity matching among different SIFT features, (Lowe, 2004).

To enhance the process of feature matching Approximate Nearest Neighbor (ANN) is used. (Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, & Angela Y. Wu, 1998) describe the ANN technique for feature matching. In simpler terms, ANN is a searching algorithm which works on a data structure to locate the nearest neighbors. In image feature matching, an ANNkd tree is built consisting of all the keypoints of the image. It creates a database storing the information of each keypoint in the image. Then the keypoints of other image is compared with this database to perform the feature matching. Newly detected points are also added to the database and the process continues up to the last image. This approach is very fast and produces efficient results.

Outlier removal is done by using the Random sample consensus (RANSAC). RANSAC is highly efficient and robust method for outlier removal which works on the basis of model fitting approach. It takes a small set of data and analyses the parameters of the predefined models. If the parameters obtained by the taken data set do not satisfies the predefined model, then data points are treated as outliers, (Fischler & Bolles, 1981).

2.2.3.3 Bundle adjustment

Bundle adjustment is the procedure based on non-linear least squares to minimize the reprojection errors, (Yang et al., 2013). Here the term bundle refers to the 'bundle' of rays leaving 3D points and converging onto the camera station. Bundle adjustment is used to generate the optimal 3D geometries with the help of camera parameters. As mentioned by (Yang et al., 2013), nonlinear problems are generated in applying SfM due to the rotation, perspective division and radial distortion. It is implemented in Structure from Motion approach for the optimization of the point position estimation using multiple overlapped images. Levenburg- Marquardt algorithm is exclusively used to handle the nonlinearity problem in SfM, however its disadvantage is that it only confirms the local minimum and does not take care of the global minimum in large scale issues, (Yang et al., 2013). To overcome this localization problem the (Lourakis & Argyros, 2009) developed a sparse bundle adjustment package.

CHAPTER 3: MATERIALS AND METHODOLOGY

3.1 Study Area

This research is carried out at IIRS (Indian Institute of Remote Sensing, Dehradun) campus. IIRS is a premier institute under Indian Space Research Organization (ISRO), Department of Space, and Government of India. This beautiful campus is situated at Kalidas Road, Dehradun, Uttarakhand. It was established in 1966 and is a renowned name in the field of Remote Sensing and GIS with capacity building in the areas of Remote Sensing and GIS as the quality policy of IIRS.

The main building and Godavari Hostel building of the IIRS campus shown in Figure 3.1(a) and 3.1(b) respectively are chosen as the test scene. The average height of the terrain in front of the building is 652m in geodetic coordinate system with reference to WGS 84 datum.

(a) (b)

Figure 3.1 Study area, (a) IIRS main building (Source: Jayson Jayeshkumar Jariwala, 2013), (b) Godavari Building

3.2 Datasets

The datasets used in this research are as follows:

- Digital Elevation Model (DEM) of CARTOSAT 1 is used to extract the height of the target object for field planning.
- Videos and images captured using the conventional cameras of different resolutions are used for the point cloud generation.
- GPS data in static and kinematic mode is used for geo-referencing the 3D point cloud.

3.3 Hardware Used

The details of the hardware involved in this study is given in Table 3.1

Table 3.1 Details of the Hardware used

S. No.	Hardware	Model No.	Used for
1.	Cameras	NIKON D60,D90, Coolpix L810, Sony Cybershot DSC-H55, Nokia Asha 305, Canon PowerShot SX100 IS	acquiring images and videos of scene
2.	GPS system	Trimble R7 GNSS	Obtaining the position of the exposure station
3.	Total Station	Leica TPS 1200	Accuracy assessment of the georeferenced point cloud
4.	Laptop	HP 2000, (Intel core i3, 64 bit, 2.20GHz, 4GB RAM)	Data integration and processing

3.4 Software Used

A wide range of open source and commercial software have been used to perform this study. Table 3.2 gives the details of all the software used in this research work.

Table 3.2 Details of the Software Used

S. No.	Software/Packages	Used for
1.	Format Factory 3.00	rescaling of images
2.	Trimble Business Center	the post processing of GPS data in static and kinematic mode
3.	Arc GIS 10.1	height extraction using Cartosat DEM
4.	OSM Bundler	point cloud generation
5.	SfM_georef_v2.3	georeferencing of point cloud
6.	SQLite manager	updating the camera database
7.	MATLAB 2012a	coordinate transformation, georeferencing and statistical analysis
8.	Cloud Compare v2	visualizing the 3D point cloud
9.	LPS of ERDAS Imagine 2013	bundle adjustment
10.	Python v2.7	geotagging of image frames
11.	NetBeans v7.3	tool development for the generation of georeferenced point cloud
12.	MS office	reports and presentation slides preparation

3.4 Methodology

The adopted methodology for this study is divided into three segments and they are shown as flowchart in Figure 3.2. The workflow begins with the field planning followed by the data collection and data processing segments. All these segments and the complete methodology adopted to carry out this study is explained in the further sections.

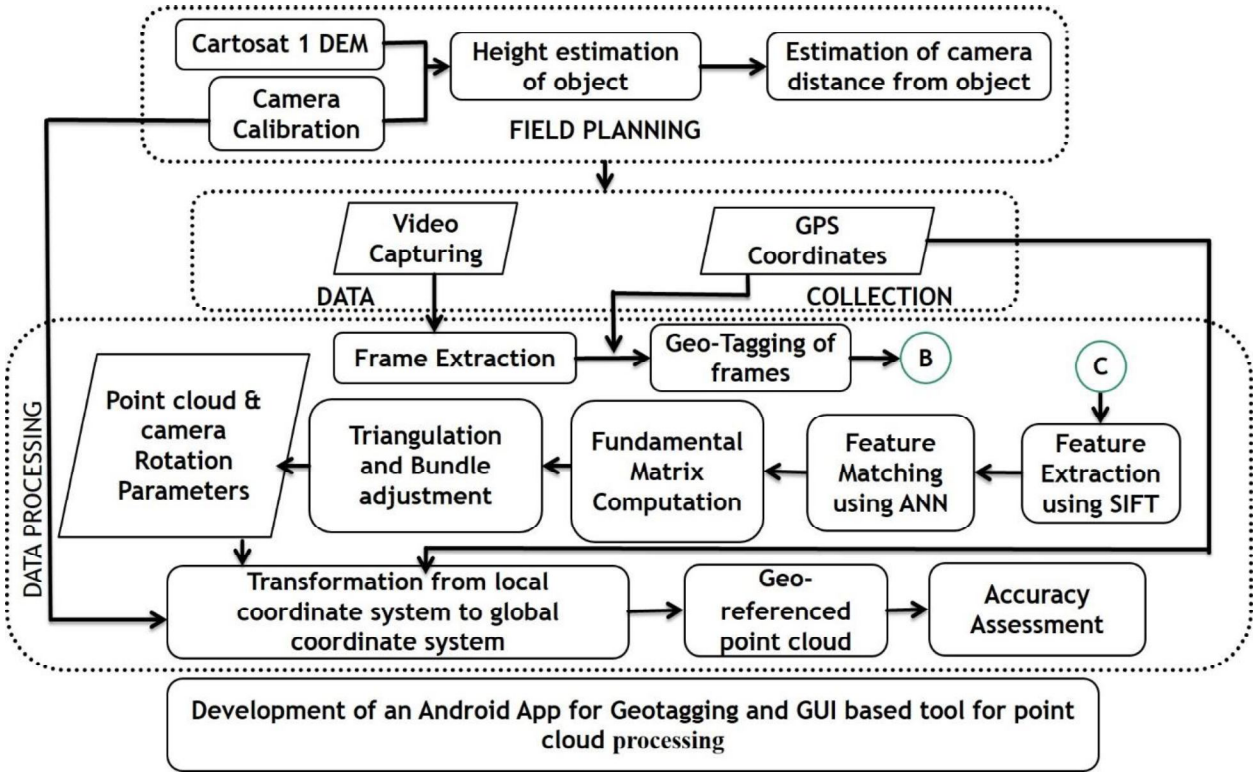


Figure 3.2 Flowchart of Methodology

3.4.1 Field Planning

As per the guidance of (ISPRS, 2010) for field Planning for close range photogrammetry, field planning includes site visiting, planning for the position of the exposure station (distance estimation from object, baseline estimation), note making, checking the weather suitability, taking care of safety measures etc. There are certain prerequisites for data collection like scale, minimum distance between the object and the exposure station and baseline. The minimum distance is calculated by the help of equation 3.1

$$\frac{D}{f} = \frac{H}{h} \tag{3.1}$$

Where D is the distance of exposure station from the object, H is object height, f is focal length of the camera and h is the sensor height.

Estimation of focal length *f* of a camera is done in camera calibration and sensor height *h* is obtained from the specifications of camera given in its user manual. As per the adopted methodology, CARTOSAT DEM is used to extract the object height *H*.

Since *f* is the intrinsic parameter of the camera, so camera calibration is performed to obtain *f*. Camera calibration is the first process which is very necessary to obtain the camera parameters. Here, a calibration sheet i.e., a chessboard is taken. Then, the images of this calibration sheet are taken from 12 different positions. The calibration is done by using openCV library in Python(x, y) environment. The calibration code displays the camera matrix, distortion parameters as output. It should be noted that the camera covers 80% of the calibration sheet and the focal length should be fixed or else there occurs error in the output of resultant parameters. The focal length is displayed in pixels which needs to be converted into millimeter using the formula,

Focal Length (in mm) = sensor width (in mm) * (focal pixels / width)

Where, focal pixels and width, both are in pixels. As mentioned in section 2.2.1.1 focal length, principal point and lens distortion parameters are obtained in camera calibration process.

Now, scale factor (M_b) can be calculated by using equation 3.2

$$M_b = \frac{D}{f} \tag{3.2}$$

To obtain the baseline i.e. the distance between the two exposure stations can be calculated by using equation 3.3

$$B = 20\ \%\ \text{of scale factor} * \text{sensor width} \tag{3.3}$$

Here, *B* is the baseline.

Sensor width can be obtained through camera specifications.

3.4.2 Data Collection

Once the minimum distance between the exposure station and the object and the baseline are known after proper field planning, then, data collection is the next step. Firstly, the GPS base is installed at the roof of the URSD building of IIRS campus and is kept on for 74 hours so as to obtain the accurate coordinates. For the data acquisition a mobile mapping system is developed as shown in Figure 3.3. In this MMS, Camera is mounted on a wooden frame and GPS antenna is mounted on the top of the camera. Video of the object is captured using the conventional cameras along with the GPS operating in the kinematic mode to obtain the coordinates of the camera trajectory. Images of the object from different locations are also captured with different grades of cameras to compare their accuracy. Coordinates of these different positions of exposure station are obtained by operating the GPS in fast static mode and with Leica total station.

Figure 3.3 Camera and GPS for data acquisition

3.4.3 Data Processing

This section deals with the generation of point cloud and transforming it from local to global coordinate system. Bundler software of Noah Snavely is used to generate the 3D sparse point cloud. Bundler uses the SfM computer vision algorithm for 3D sparse reconstruction. SIFT, ANN and RANSAC libraries are included in it along with the SBA package of (Lourakis & Argyros, 2009). The basic working of Bundler is shown as flowchart in Figure 3.4. As shown in the flowchart of methodology in Figure 3.1, the various steps involved in the data processing are explained in the further sub sections.

3.4.3.1 Frame Extraction

A video is a sequence of image frames that are moved at very high rate up to 8 to 25 frames per second or more. These image frames are extracted from the video using openCV library of python.

3.4.3.2 GPS Post Processing and Geo-Tagging of Image Frames

Once the image frames are extracted, the next task is to geotag them. Geo-tagging is the procedure to add the geographical information to various media such as images and videos etc. A python code is developed for the geo-tagging of these image frames. EXIF (Exchangeable Image Format) tags are used to tag the latitude and longitude on the images. The coordinates of the exposure stations obtained through the GPS in UTM format are firstly post processed in the Trimble business center software. The data of both the base and rover is transferred to the

computer through the data cable using the Trimble data transfer utility. The IGS (International GNSS Service) data of few stations is downloaded along with the final orbits ephemeris data to process the base data to obtain the accurate coordinates of base. This corrected value of base coordinates is used to process the rover data. The GPS coordinates of the exposure stations are converted to the degree decimal format and then added to the image metadata by executing the python code developed for geo-tagging.

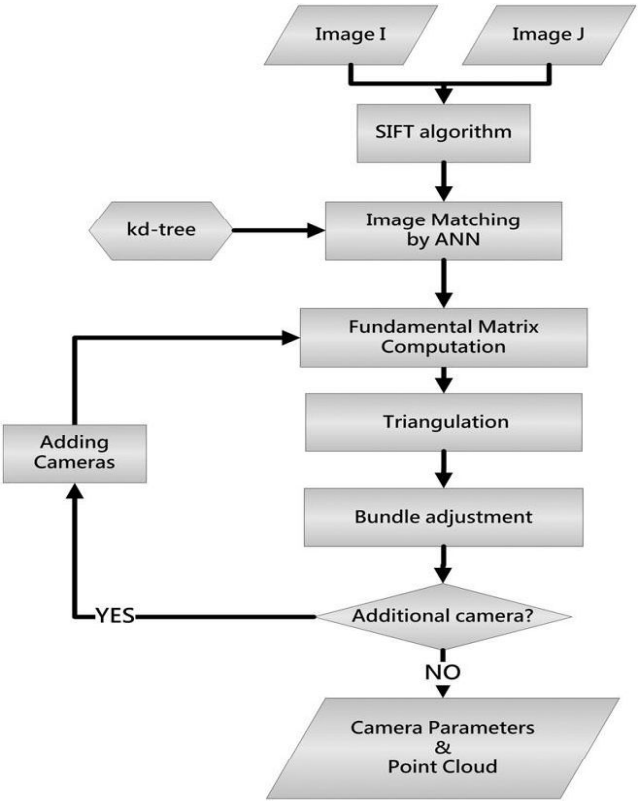


Figure 3.4 Flowchart showing the working of Bundler (Source: (Yang et al., 2013))

3.4.3.3 Feature Extraction

Feature extraction or keypoints identification on the images is the first process to generate the point cloud. Key features are extracted from the images by using the SIFT algorithm which is developed in C/C++. Working of SIFT algorithm is already explained in Chapter 2.

3.4.3.4 Feature Matching and Outlier Removal

Once the feature extraction is complete, ANN is used to enhance the feature matching task. ANN is a library developed in C++ and thus a C++ compiler is needed to execute it. Feature matching is performed between the extracted features of every pair of images.

RANSAC is used to increase the accuracy of feature matching by removing the outliers or bad matches. Python, Matlab as well as C++ version of RANSAC are available. All the wrong matches are removed after outlier removal and further the fundamental matrix computation is done.

3.4.3.4 Bundle Adjustment

Bundle adjustment is applied using the SBA package of (Lourakis & Argyros, 2009). It is very efficient package that process images simultaneously to reconstruct the scene. By using the matched keypoints and the projection matrices, SBA reconstructs the scene and calculates the orientation of camera and transformation parameters. Thus, it generates the 3D sparse point cloud of the scene along with the camera parameters file in local coordinate system.

Bundler provides the PMVS (Patch based multi view stereo) software of Dr. Yasutaka Furukawa's to convert the sparse point to dense 3D point cloud. It uses the camera parameter file generated during sparse reconstruction as the input file for the sparse to dense reconstruction.

3.4.3.5 Coordinate Transformation

The point cloud generated using Bundler is in the arbitrary coordinate system and needs to be transformed to global coordinate system with proper scale. Space Photo Intersection technique is used to obtain the global coordinates of the 3D points of the point cloud. Space intersection technique estimates the global coordinates of a point on the basis of the known camera parameters, (Jayson Jayeshkumar Jariwala, 2013). Space intersection technique uses the collinearity equations to obtain the ground coordinates of a point.

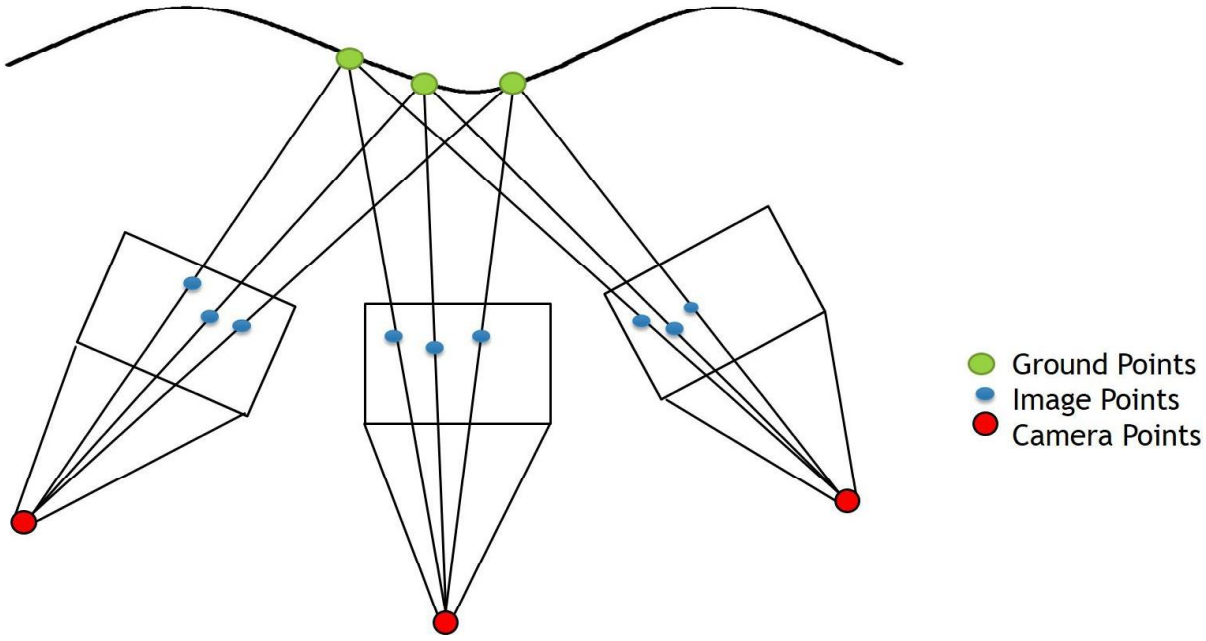


Figure 3.5 Collinearity for Space Intersection

Collinearity equations 3.4 and 3.5 are used to establish the relationship between the ground point, image point and camera point as shown in Figure 3.5

$$x_a = x_0 - f * \left[\frac{m_{11}(X_A - X_L) + m_{12}(Z_A - Z_L) + m_{13}(Y_L - Y_A)}{m_{31}(X_A - X_L) + m_{32}(Z_A - Z_L) + m_{33}(Y_L - Y_A)}\right] \quad (3.4)$$

$$y_a = y_0 - f * \left[\frac{m_{21}(X_A - X_L) + m_{22}(Z_A - Z_L) + m_{23}(Y_L - Y_A)}{m_{31}(X_A - X_L) + m_{32}(Z_A - Z_L) + m_{33}(Y_L - Y_A)}\right] \quad (3.5)$$

Here, (x_a, y_a) are image coordinates of the points, (x_0, y_0) are the principal point locations, f is the focal length of the camera, $X_L, Y_L, Z_L$are the exposure station coordinates, (X_A, Y_A, Z_A) are the ground control points, $(m_{11}, m_{12} \ldots m_{33})$ are the rotation parameters.

In collinearity equations, the principal point location can be obtained through the camera calibration and image coordinates, rotation parameters are obtained through the camera parameters file generated through the bundler and the global coordinates of the exposure station are obtained through the GPS. The global coordinates of the points can be calculated by using equation 3.4 and 3.5. In this case, there are three unknowns (X_A, Y_A, Z_A) so at least three equation are required to solve these equations. Since, a point is available in at least two images, so four equations are obtained for a point (two from each image). Then least square adjustment is applied to obtain the solution.

3.4.3.6 Accuracy Assessment

There are various factors like environmental factors, camera orientation and resolution of images etc. that contributes to the error in the point cloud. Accuracy assessment of the point cloud is done as internal and external assessment.

For performing the internal accuracy assessment few test planes are chosen. These test planes are fitted on a reference plane. Principal component analysis (PCA) that represents the axis of extreme deviation of the data is applied to obtain the parameters of the reference plane. Then the robust plane fitting of the reference frame is done using the RANSAC algorithm and the point to plane distances are obtained. This point to plane distance is used to perform the statistical analysis. Statistical parameters like mean, standard deviation, maximum distance, minimum distance, median,, 25^{th} percentile and 75^{th} percentile of the random errors are calculated. Here, the systematic errors in the data are represented by the mean while standard deviation represents the random errors of the observables and the exactitude of the obtained parameters.

External accuracy assessment can be performed in many ways. The global coordinates of few target points are measured with the help of Total Station. The total station is oriented by using the coordinates of two points that are already known by the help of GPS. The point cloud generated through the images taken by the conventional cameras is georeferenced and then viewed in the Cloud Compare software and the known target points are identified to measure the external accuracy. The error estimation is done by comparing the coordinates of the target points in the point clouds to the coordinates obtained by the total station.

3.4.3.7 Tool Development for point cloud processing

Finally a Graphical Users Interface (GUI) tool is developed on Java to perform all the data processing task. The codes written in Matlab and Python to perform various tasks are integrated on this GUI to form a common platform to perform all the data processing. An android based application is also developed to perform the task of Geo-tagging by enabling the device's camera and GPS.

3.4.4 Methodology to handle the GPS unavailability

As shown in the Flowchart of methodology in Figure 3.1, frames are geotagged based on the known camera station coordinates. During the data collection, the problems of poor GDOP (Geometric dilution of precision) or satellite unavailability may occur. In this case, few extracted frames remain untagged and thus it requires that the unknown locations of the camera stations should be determined. So in Figure 3.1, B represents both the geotagged and untagged frames. In this research, an attempt is made to address the issue of unavailability of GPS signals by developing the methodology shown in Figure 3.6.

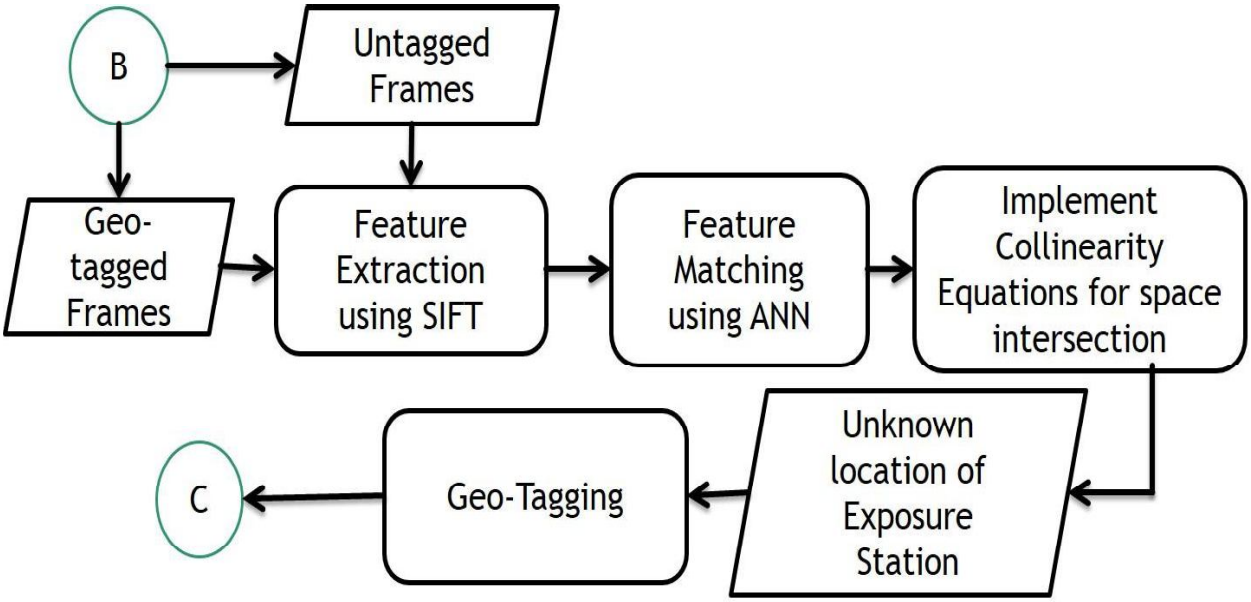


Figure 3.6 Methodology to handle the GPS unavailability

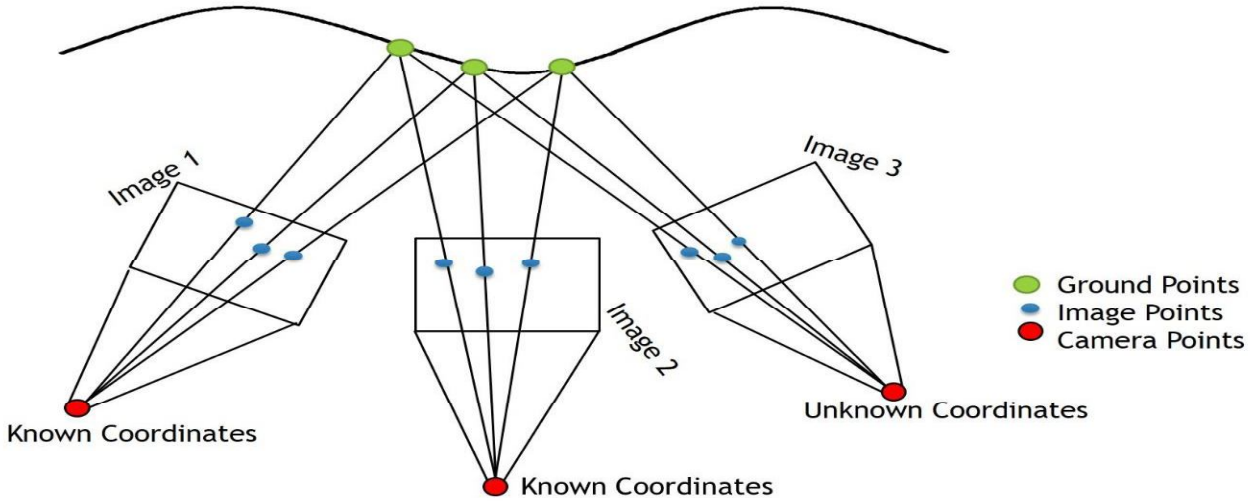


Figure 3.7 Figure depicting unknown camera station

Due to the unavailability of GPS signals or poor GDOP, the coordinates of few of the camera stations are unknown and this situation is shown in Figure 3.7. The collinearity equations for space intersection are used to obtain the unknown coordinates of the camera stations. The unknown coordinates of the camera station can be determined using the equations 3.4 and 3.5. In these equations, the camera coordinates X_L, Y_L, Z_L are unknown. Image coordinates are obtained through bundler output files, focal length is determined through camera calibration. In Figure 3.7, the camera coordinates of two camera stations are known so the global coordinates of the points belonging to image 1 and image 2 can be determined. Same common points on image 3 are taken so as to determine the unknown coordinates of the third camera station. Minimum two points are required to obtain the solution for the collinearity equations. Each point generates two equations such that four equations are obtained from two points. So, least square adjustment solution of these equations provide the location of exposure station where GPS signals are not available.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Field Planning

As per the Methodology explained in Chapter 3, the height extraction of the object using the CARTOSAT DEM was the primary task for field planning. The objective of the field planning is to analyze the study area and identify the approximate locations of the points (path for the camera motion) from where the photographs could be taken. Any satellite stereo pair could be used based upon the optimum resolution required to view the study area for proper field planning. In this study, Cartosat data is used due to the availability of the data and to develop a methodology for field planning showing the application of satellite data in Close Range Photogrammetry. Based upon the scene visibility and availability of the data any satellite stereo pair could be used. The methodology developed for the field planning uses the IIRS main building as test study area and the same approach was used in the field planning for Godavari building. The extracted heights of the main building and the administrative building of the IIRS campus are 9.97m and 3.37m respectively. The height of the main building of IIRS obtained through GPS is 11.36m. The difference of the extracted height and height obtained through GPS was calculated for finding the error between two heights and the calculated error is 1.39m. The focal length obtained through the camera calibration was used with the extracted height to find the other parameters for field planning.

4.1.1 Camera Parameters

Different grades of camera used in this research work are firstly calibrated to obtain the intrinsic parameters. PhotoModelerScanner software was used to perform the camera calibration task. NIKON D90 camera was used to obtain the parameters for field planning and its intrinsic parameters estimated were estimated using camera calibration. 10 Images of a calibration sheet (36 inches × 36 inches) consisting of 140 grid points and 4 control points taken by NIKON D90 at different angles were used for the camera calibration. Images of the calibration sheet are shown in Figure 4.1.

Figure 4.1 Calibration Sheet

Focal length, lens distortion parameters and location of the principal point are obtained through the camera calibration.

Parameters obtained through camera calibration are as follows:

Focal Length Value: 20.31 mm
Pixel Size, Sx=5.50µm, Sy=5.50µm
Xp - principal point x Value: 11.753564 mm Deviation: Xp: 0.007 mm
Yp - principal point y Value: 8.04432 mm Deviation: Yp: 0.010 mm
K1 - radial distortion 1 Value =-0.1136932061321540
K2 - radial distortion 2 Value=0.1195857117314250
K3 - radial distortion 3 Value=-0.0525860556555812
P1 - decentering distortion 1 Value: 0.000e+000
P2 - decentering distortion 2 Value: 0.000e+000
Point Marking Residuals Overall RMS: 0.688 pixels

As per the calibration manual of the PhotoModelerScanner the overall point marking residual should be less than 1pixels and the estimated overall point marking residual was 0.688pixels which was as per the standard. Using these intrinsic parameters the parameters for the field planning are estimated.

4.1.2 Parameters for Field Planning

Field planning is necessary for proper data acquisition and the steps included in it are already explained in section 3.4.1. Figure 4.2 Shows the shape file of IIRS campus which is digitized from Cartosat 1 imagery and is used along with cartosat DEM for calculating the zonal statistics

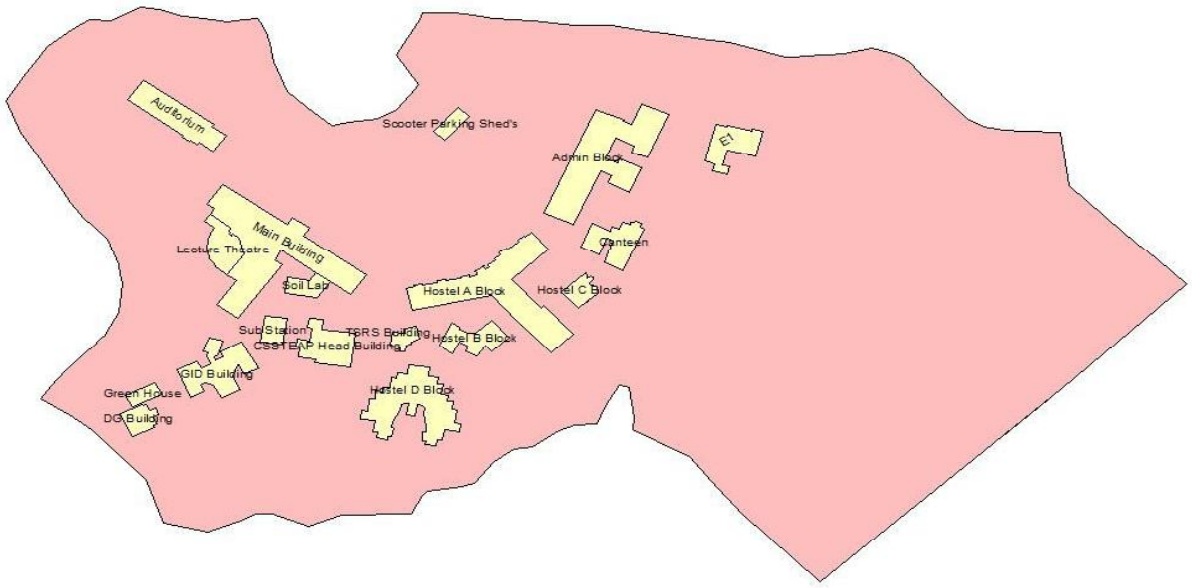


Figure 4.2 Shapefile of IIRS campus

The parameters obtained for the field planning for data acquisition are as follows:

Extracted height of building H=9.97m.
Focal length of camera f=20.31mm.
Sensor width w=23.6mm.
Sensor height h=15.8mm.
Minimum distance between camera and object was calculated using equation 3.1,
D=12.81m.
Scale $M_b \approx 630.73$.
Baseline= (20/100)*(630.73*23.6) =2.98m
Thus, the minimum distance between the camera station and the object is taken 15m for this study. Following the same methodology the points were marked in front of the Godavari Building of IIRS for taking the images.

4.2 Data Collection

The points marked in the field planning were used to capture the images of the object by different grades of conventional cameras. The coordinates of these points were also obtained through GPS. The cameras NIKON D60, D90, Coolpix L810, Sony Cybershot DSC-H55 and Canon PowerShot SX100 IS were used to capture the images. A total of 35 images were taken from each camera to cover the front face of Godavari building of IIRS campus. A video is also captured by NIKON D90 camera along with the GPS working in kinematic mode obtain the trajectory information of the camera.

To measure the GPS coordinates, base was installed at the roof of URSD building of IIRS campus and the rover was used to take the readings at the points from where the photographs were taken. The coordinates of two points were measured with the GPS and the rest coordinates of the other points were measured by the Total Station. Total Station was oriented by the help of the points measured by the GPS. The GPS data was post processed in Trimble Business center to obtain the corrected coordinates. The horizontal and vertical precision obtained in the processed GPS data of base station is 0.6cm and 0.7cm respectively. The horizontal and vertical precision obtained in the processed GPS data of the first rover point is 0.6cm and 0.8cm respectively whereas for the second point, the horizontal and vertical precision is 0.6cm and 0.9cm. The post processing of the GPS data provides the coordinates of the marked points on the ground and for that reason the tripod offset of 1.50m and camera electronic center height 6.50cm is added to the GPS height to obtain the position of exposure station.

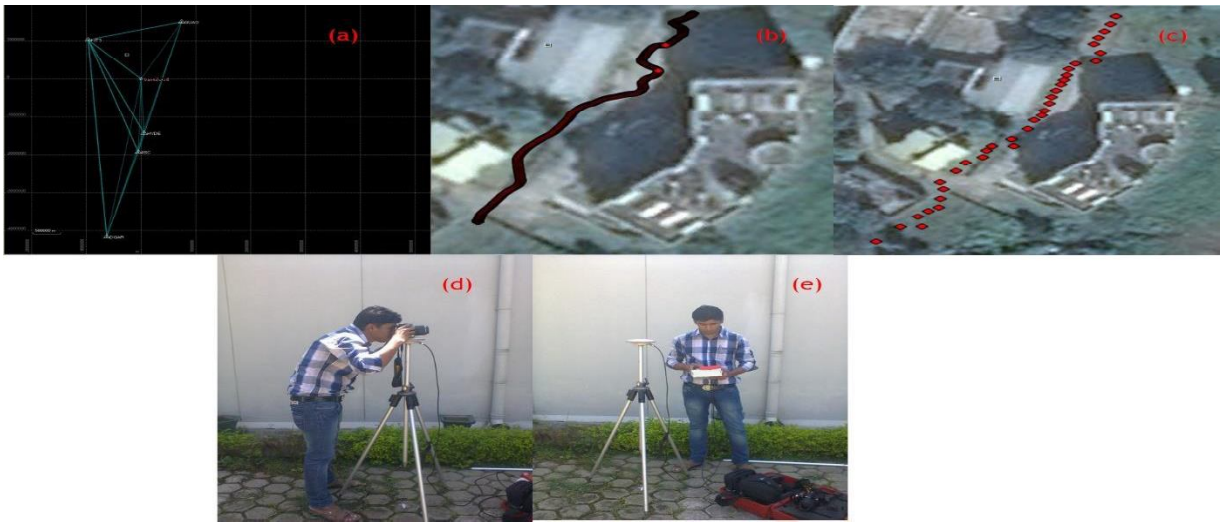


Figure 4.3 Field data collection & GPS post processing

The coordinates of base station was corrected by post processing the base station data with the few IGS stations data. The baselines are shown in Figure 4.3(a). Then the corrected coordinate values of the base station were used for the post processing of the rover data. Figure 4.3(b) represents the trajectory of the camera path followed for capturing the video of the Godavari building. The points used for capturing the images from different grades of conventional cameras are shown in Figure 4.3(c). Figure 4.3(d) and 4.3(e) shows the images of data collection.

The Terrestrial Laser Scanner RIEGL VZ-400 is used to for obtaining the highly dense point cloud of the object. The accuracy mentioned in the datasheet of RIEGL VZ-400 is 5mm with precision of 3mm. The laser point cloud was georeferenced by the known target point coordinates measured by total station and was used as a reference to compare the accuracy of the georeferenced point cloud generated through the video and images captured through the conventional cameras.

4.3 Data Processing

As mentioned in the section 3.4.3, firstly the image frames were extracted from the video and were geo-tagged using the python code. These geo-tagged frames were then processed in bundler software of Noah Snavely to generate the 3D point cloud. The processing of the bundler software involves feature detection and matching, outlier removal and bundle adjustment which were already explained in chapter 2. The Number of matched features is maximum in the adjacent images and decreases between any other pair of images. The graph showing the variation in number of matched features in different pair of images is shown in Figure 4.4. There were total 35 images. Feature matching was done by making the all possible pairs of images. This graph is showing the number of matched features in the image pairs formed using the first image. It can be easily seen that the number of matched features is maximum in image pair (0, 1) i.e. between the first and the second image. This value decreases between the first and third image i.e. image pair (0, 2) and keeps on decreasing with the other pair of images. This happens because of the lack of

overlap of images in the image pairs. The first pair of images has maximum overlap between them and thus the number of matched features for this pair is maximum.

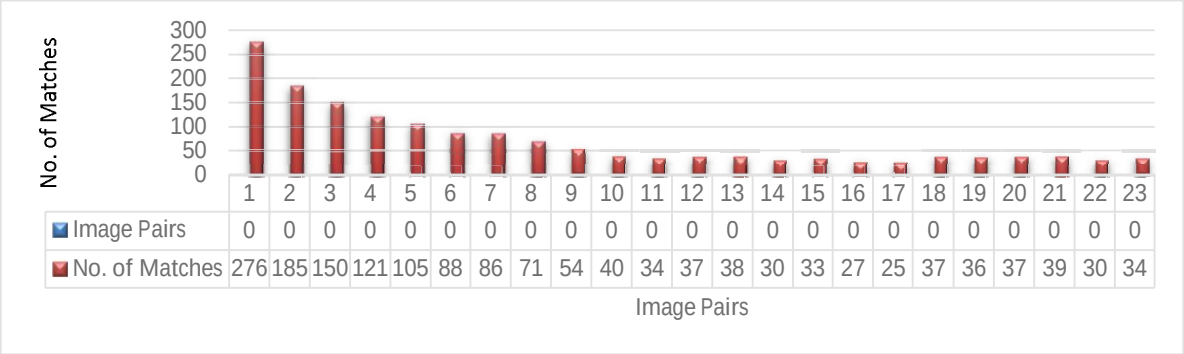


Figure 4.4 Number of Matches in different pair of Images

The output of the bundler contains the .txt, .out and .ply files. The information related to the matched features and camera parameters (rotational and translational) is present in these files. The .ply files contains the generated 3D point cloud and it can be viewed in the Cloud Compare software. The generated point cloud was sparse and therefore converted to dense point cloud by running the PMVS on bundler output. The same approach was applied to generate the 3D point cloud from the images taken from the different grades of conventional cameras.

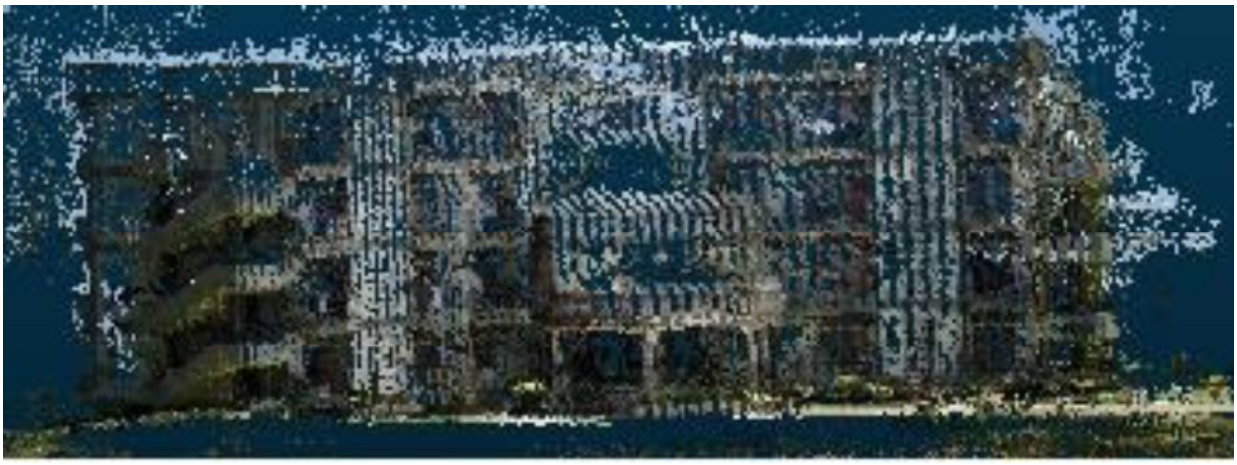


Figure 4.5 Point Cloud generated using the extracted frames from a video

The point cloud generated using the 198 extracted frames from a video is shown in Figure 4.5. Its top view is shown in Figure 4.6. The point cloud generated using the bundler was in the local coordinate system. It needs to be transformed into the global coordinate system. The transformation of the 3D point cloud from the local coordinate system to the global coordinate system was done using the collinearity equations for the space photo intersection as explained in the section 3.4.3.5. The global coordinates of the exposure stations measured through the GPS were used for the transformation of the 3D point cloud from the local coordinate system to the global coordinate system.

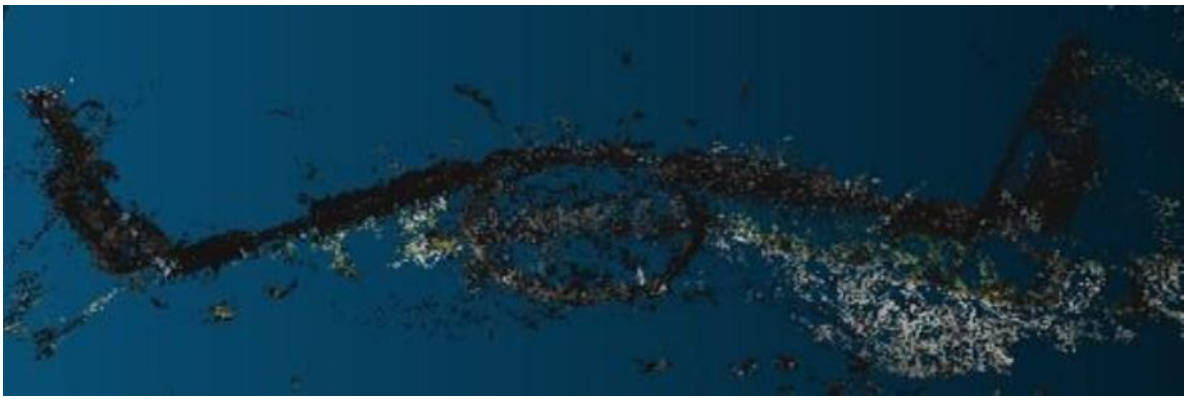


Figure 4.6 Top view of the point cloud generated from image frames extracted from the video

The images taken from different grades of conventional camera were also processed using the bundler to generate the 3D point cloud. PMVS was used to convert the generated point cloud to dense point cloud. The output of the bundler was used as an input to the PMVS for dense reconstruction. PMVS is available in the OSM Bundler package. The generated point cloud for all the cameras was georeferenced and the accuracy assessment was also done. The comparison of all the cameras was done on the basis of the external accuracy assessment.

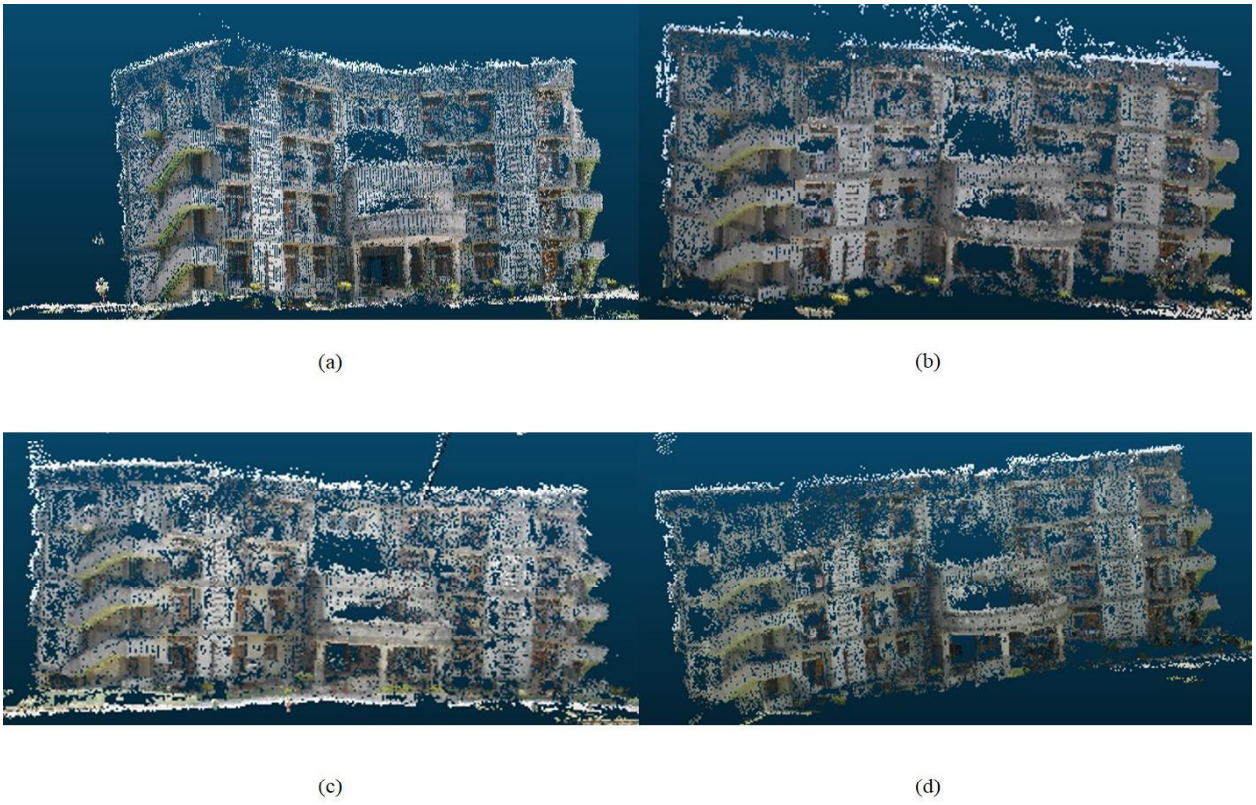


Figure 4.7 Point Cloud from the images taken from different grades of cameras, (a) point cloud by NIKON D60, (b) point cloud by NIKON D90, (c) point cloud by NIKON CoolPix L810, (d) point cloud by Sony Cybershot DSC-H55

The point clouds generated from the images taken from the different grades of conventional cameras are shown in Figure 4.7. Figure 4.7(a) shows the point cloud of the Godavari building of IIRS, generated from the images taken by NIKON D60 camera. Figure 4.7(b) represents the point cloud generated from the images taken by NIKON D90 camera. Point clouds generated by the images taken from NIKON CoolPix L810 and Sony Cybershot DSC-H55 are shown in Figure 4.7(c) and 4.7(d) respectively.

The georeferenced point cloud is shown in Figure 4.8. 4 Points along with their global coordinates are shown in it. In the label showing the coordinates of a point, first line represents the point id, second line displays the global coordinates of that point and third line shows the color vector of the point. Since the global coordinates of the camera exposure station were used for georeferencing the point cloud, this method of georeferencing is called direct georeferencing.

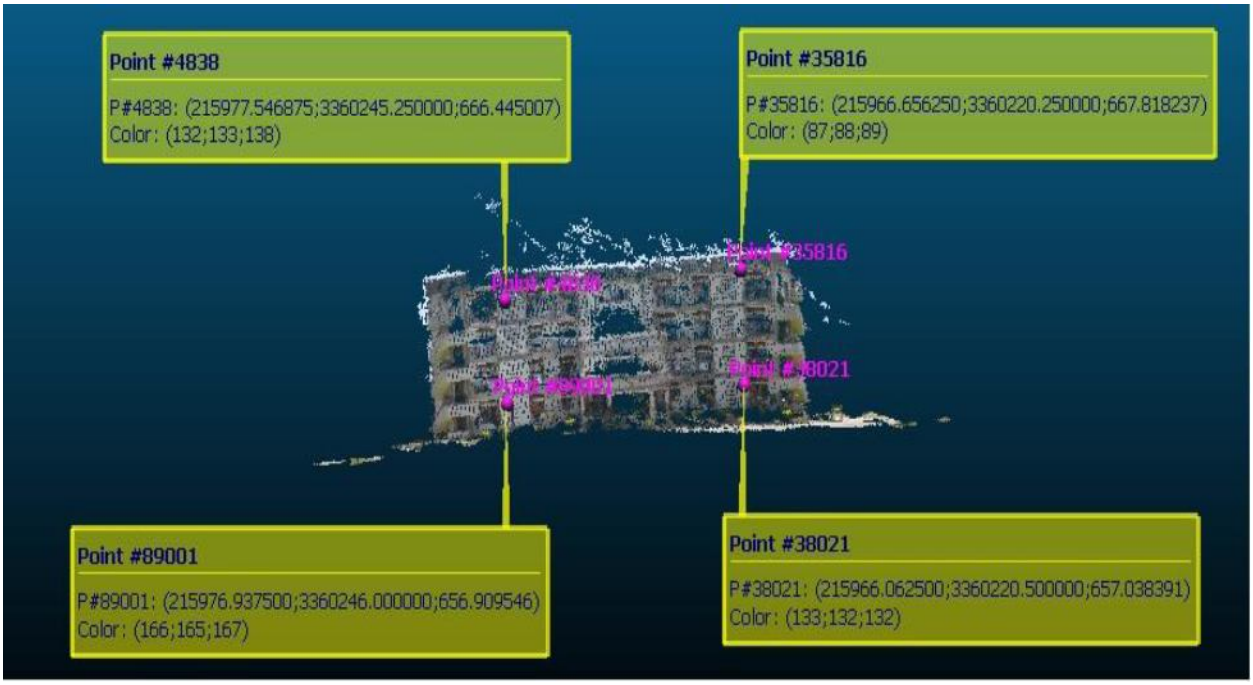


Figure 4.8 Georeferenced Point Cloud

4.4 Accuracy Assessment

As mentioned in the section 3.4.3.6, accuracy assessment of the point cloud is done in two ways, internal accuracy assessment and external accuracy assessment.

4.4.1 Internal Accuracy Assessment

In the internal accuracy assessment, the point cloud is analyzed to check the accuracy of the discrete points, density of the point cloud and gaps. It is generally done to check whether the generated points are forming a surface or not. The process to perform the internal accuracy assessment is already explained in section 3.4.3.6. As per methodology adopted for internal accuracy assessment, the points were chosen from three different areas. The estimation of the

points to plane distance was done by performing plane fitting on them. Several statistical parameters like mean, standard deviation, minimum and maximum distances, median distance were also estimated to perform the internal accuracy assessment. Figure 4.9 shows the areas from where the points are taken for internal accuracy assessment. The selected point samples are processed using the Jayson Jariwala (2013)'s MATLAB code for internal accuracy assessment.

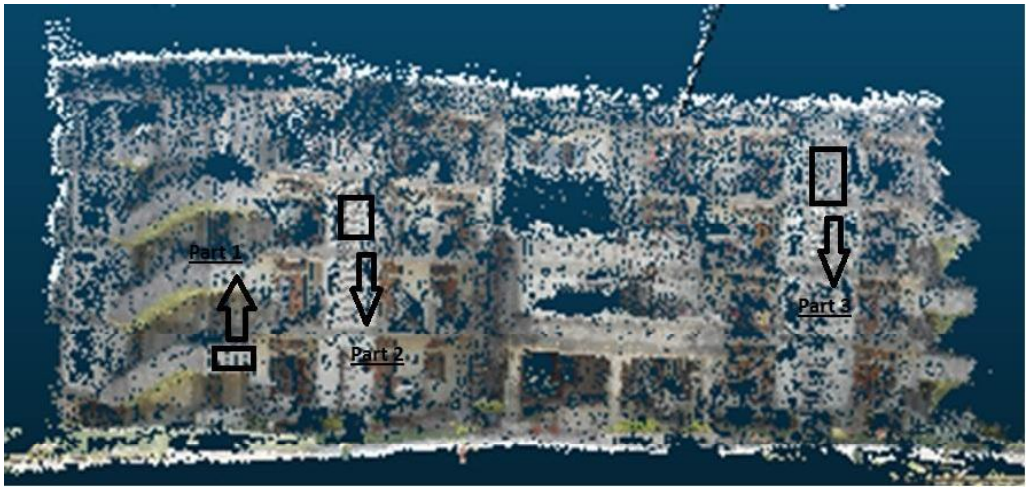
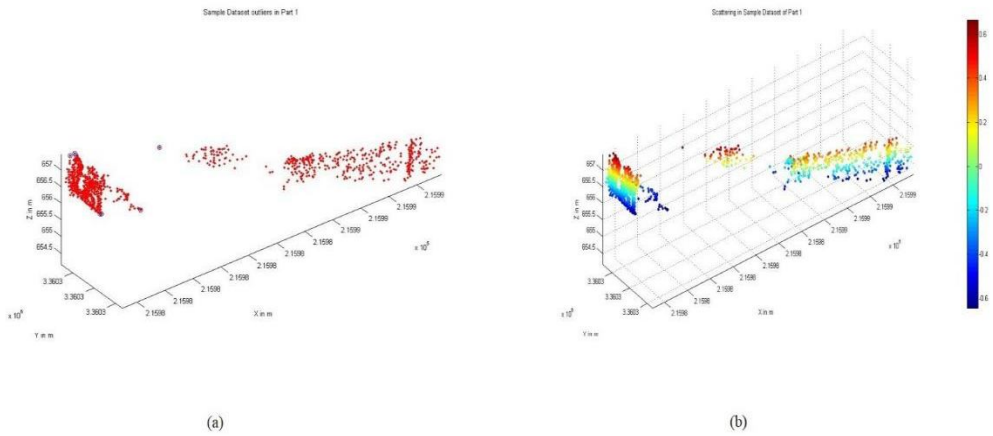


Figure 4.9 Test planes for Internal Accuracy Assessment

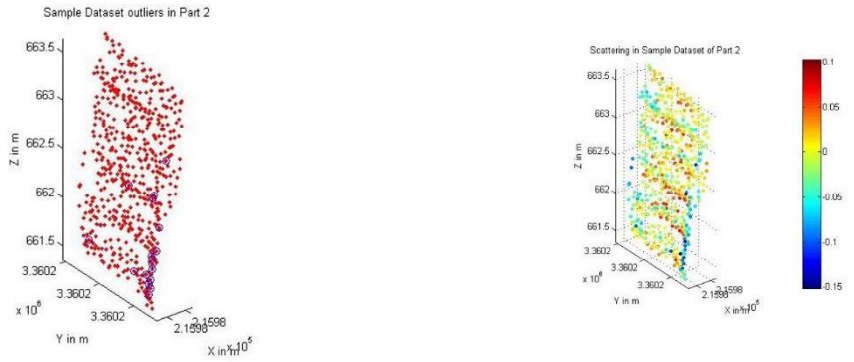
4.4.1.2 Internal accuracy assessment of NIKON D60

The internal accuracy assessment of the point cloud was done by taking the sample points from three different parts of the point cloud. The number of points in part 1 sample, part 2 sample and part 3 sample were 964, 568 and 1197 respectively. Figure 4.10(a) shows the outliers in blue color in point to plane matching and Figure 4.10(b) scattering of samples in Part 1 sample. The color bar in the Figure 4.10(b) represents the scattering of points from the reference plane in the northing direction. The outliers in point to plane matching and scattering of samples for part 2 are shown in Figure 4.11(a) and 4.11(b) respectively, Figure 4.12(a) and 4.12(b) shows the same information for part3.



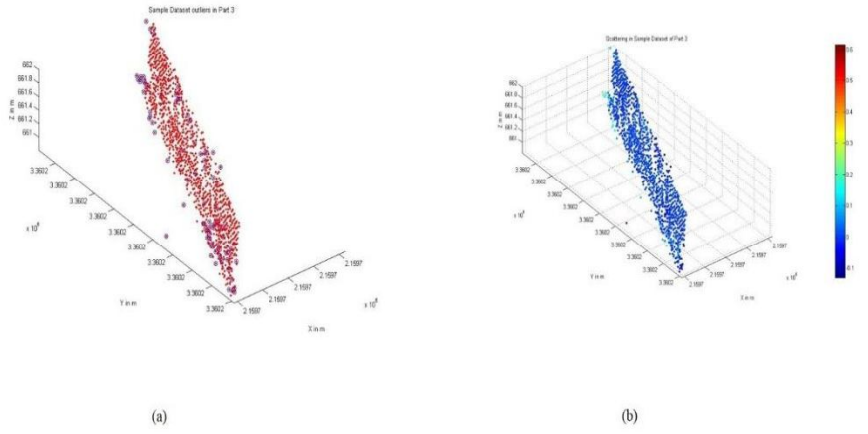
(a) (b)

Figure 4.10 (a) Outliers in point to plane matching, (b) scattering of samples in Part 1 sample



(a) (b)

Figure 4.11 (a) Outliers in point to plane matching, (b) scattering of samples in Part 3 samples



(a) (b)

Figure 4.12 (a) Outliers in point to plane matching, (b) scattering of samples in Part 3 samples

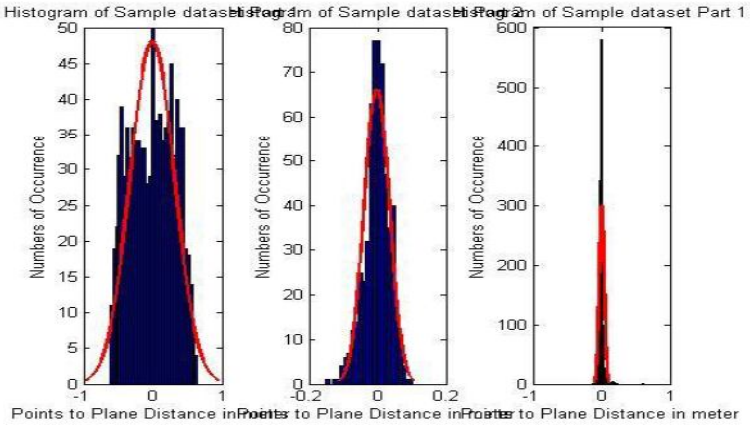


Figure 4.13 Histograms representing the Points to Plane Distance in Part 1, Part 2 and Part 3 samples

Figure 4.13 shows the plots of the histograms representing the points to plane distance in part 1, part 2 and part 3 samples. For performing the statistical analysis, box plots for all the sample parts are shown in Figure 4.14.

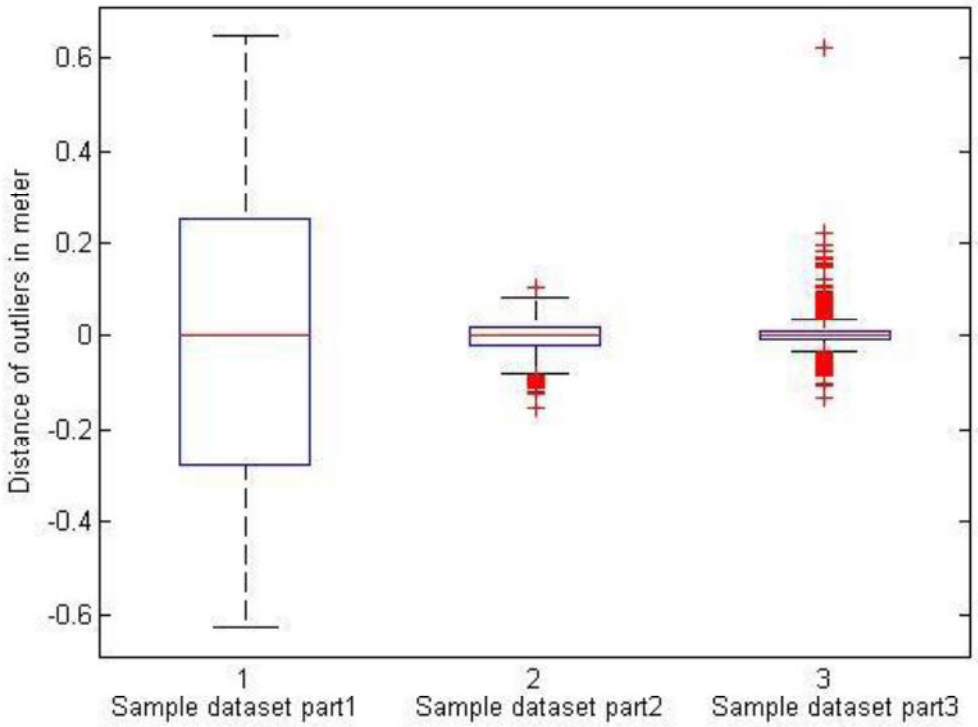


Figure 4.14 Box plot of sample parts for Points to plane distances

Table 4.1 Statistics of internal accuracy assessment of NIKON D60

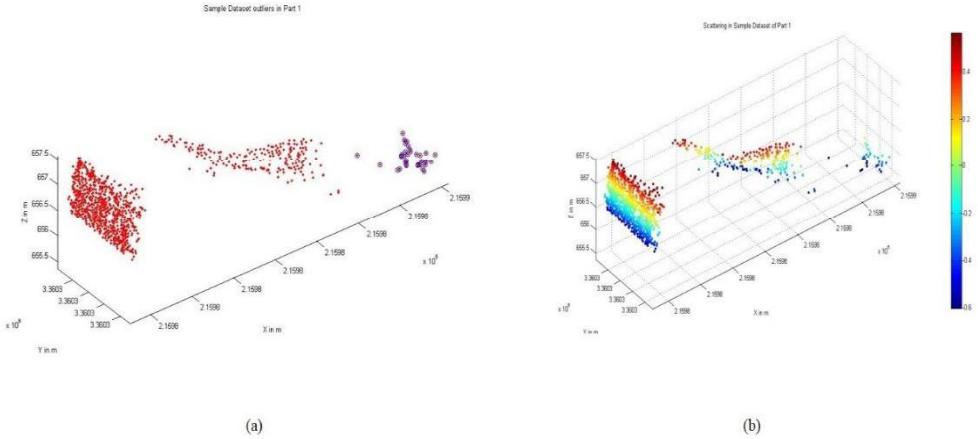
NIKON D60			
	Part 1(m)	**Part 2(m)**	**Part 3(m)**
Mean	-0.01017	-0.00254	0.001252
Standard Deviation	0.31802	0.036364	0.034448
Maximum Distance	0.64703	0.10341	0.62175
Median	0	0	0
Minimum Distance	-0.62585	-0.15177	-0.13281
Variance	0.10114	0.001322	0.001187

Table 4.1 shows the statistical parameters calculated for the internal accuracy assessment of the point cloud generated using the images taken from the NIKON D60 camera. As per (Jayson Jayeshkumar Jariwala, 2013), the red line in the boxplot shown in Figure 4.14 indicates that the 50% of the data is greater than the median value, the blue line at the bottom represents the lower quartile and indicates that the 25% of the data is less than this value whereas the blue line at the top represents the upper quartile and indicates that the 25% of the data is above this value. The red plus sign at the top and the bottom indicates that the outliers which is greater than 1.5 times the upper quartile and lesser than 1.5 times of the lower quartile respectively. The black lines at the top and the bottom represents the greatest values excluding outliers and least values excluding outliers respectively.

On the analysis of the histograms and box plots, it is clear that there exist both the symmetrical and asymmetrical distribution (positive skewed or negative skewed) in the point cloud. The box plot analysis is done to identify the scattering and skewness in the point cloud and for the identification of the outliers. In Table 4.1, the range of the points to plane distances in part 1, part 2 and part3 is -0.62585m to 0.64703m, -0.15177m to 0.10341m and -0.13303m to 0.62059m respectively. From Table 4.1, it is clear that the part 1 and part 2 are left skewed whereas the part 3 is almost symmetrical (mean≈median). As per the histograms in Figure 4.13 and box plots in Figure 4.14, the part 3 sample has normal distribution. The errors in the point cloud occurs due to the distortion in the positions of the epipolar lines in the overlapping images.

4.4.1.3 Internal accuracy assessment of NIKON D90

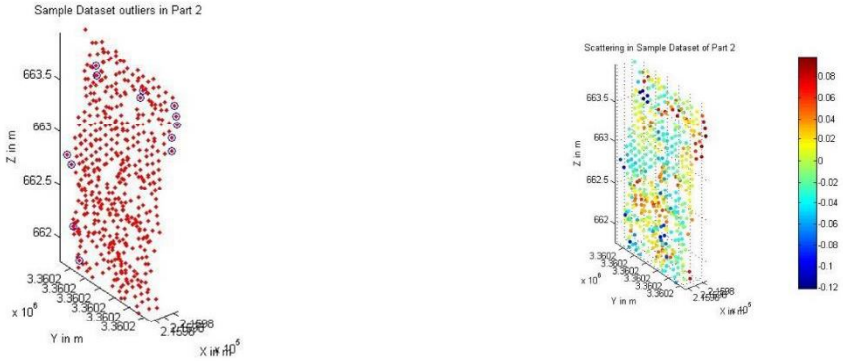
The procedure is same as applied for the internal accuracy assessment of the point cloud generated from the images taken from NIKON D60 camera. The number of points in part 1 sample, part 2 sample and part 3 sample were 987, 507 and 986 respectively.



(a) (b)

Figure 4.15 (a) Outliers in point to plane matching, (b) scattering of samples in Part 1 sample

Figure 4.15(a) shows the outliers in blue color in point to plane matching and Figure 4.15(b) scattering of samples in Part 1 sample. The color bar in the Figure 4.15(b) represents the scattering of points from the reference plane in the northing direction. The outliers in point to plane matching and scattering of samples for part 2 are shown in Figure 4.16(a) and 4.16(b) respectively, Figure 4.17(a) and 4.17(b) shows the same information for part3.



(a) (b)

Figure 4.16 (a) Outliers in point to plane matching, (b) scattering of samples in Part 2 samples

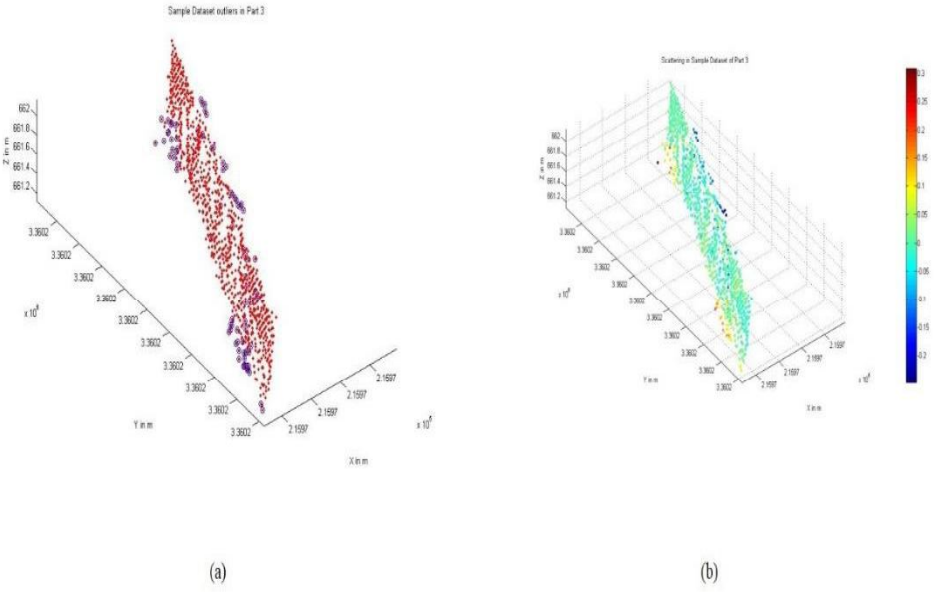


Figure 4.17 (a) Outliers in point to plane matching, (b) scattering of samples in Part 3 samples

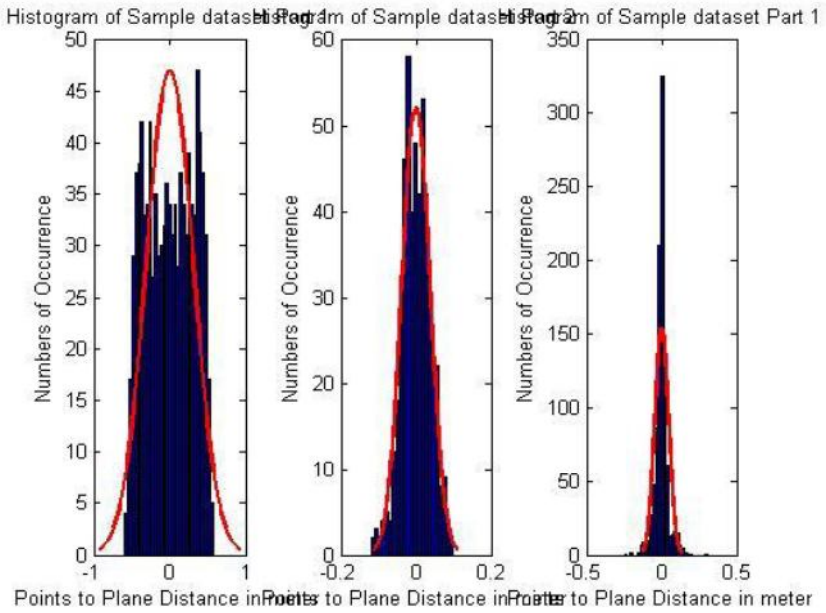


Figure 4.18 Histograms representing the Points to Plane Distance in Part 1, Part 2 and Part 3 samples

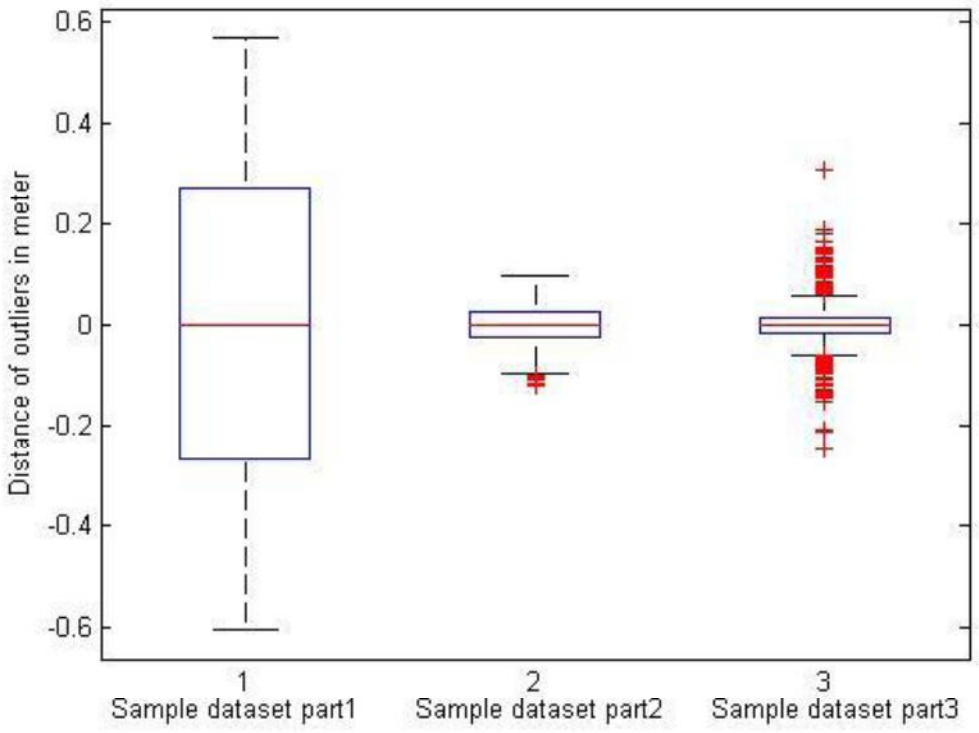


Figure 4.19 Box plot of sample parts for Points to plane distances

Figure 4.18 shows the plots of the histograms representing the points to plane distance in part 1, part 2 and part 3 samples. Box plots for all the sample parts are shown in Figure 4.19.

Table 4.2 Statistics for internal accuracy assessment of NIKON D90

NIKON D90			
	Part 1(m)	**Part 2(m)**	**Part 3(m)**
Mean	0.000992	-0.00051	2.73E-05
Standard Deviation	0.30697	0.037081	0.044191
Maximum Distance	0.56777	0.099002	0.30836
Median	0	0	0
Minimum Distance	-0.60311	-0.1208	-0.24738
Variance	0.094228	0.001375	0.001953

Table 4.2 shows the statistical parameters calculated for the internal accuracy assessment of the point cloud generated using the images taken from the NIKON D90 camera. The analysis is same as that performed for the NIKON D60.

4.4.1.4 Internal accuracy assessment of NIKON CoolPix L810

The procedure is same as applied for the internal accuracy assessment of the point cloud generated from the images taken from NIKON D60 camera. The number of points in part 1 sample, part 2 sample and part 3 sample are 474, 296 and 666 respectively.

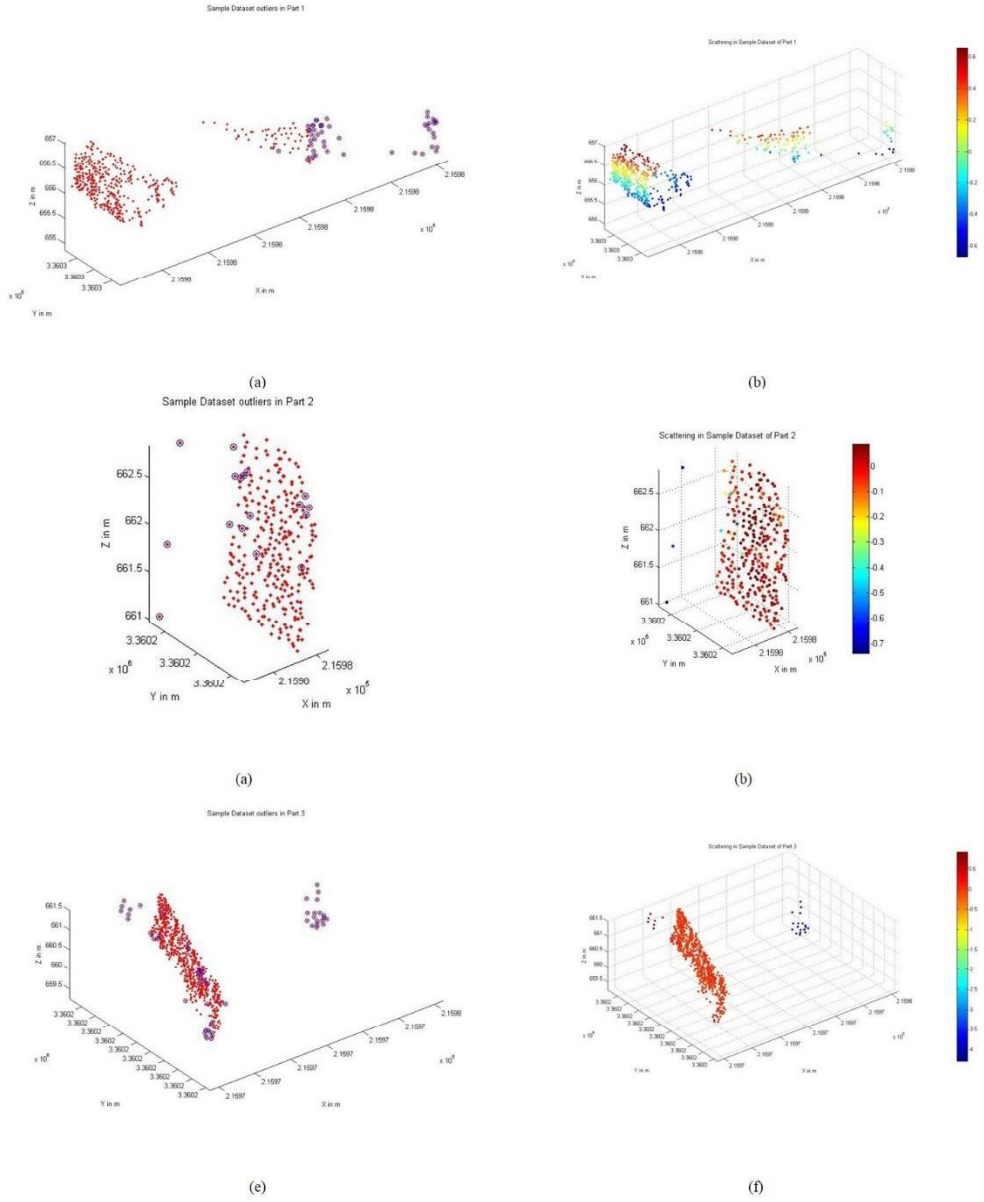


Figure 4.20 (a) Outliers in point to plane matching, (b) scattering of samples in Part 1 sample (c) Outliers in point to plane matching, (d) scattering of samples in Part 2 sample (e) Outliers in point to plane matching, (f) scattering of samples in Part 3 sample

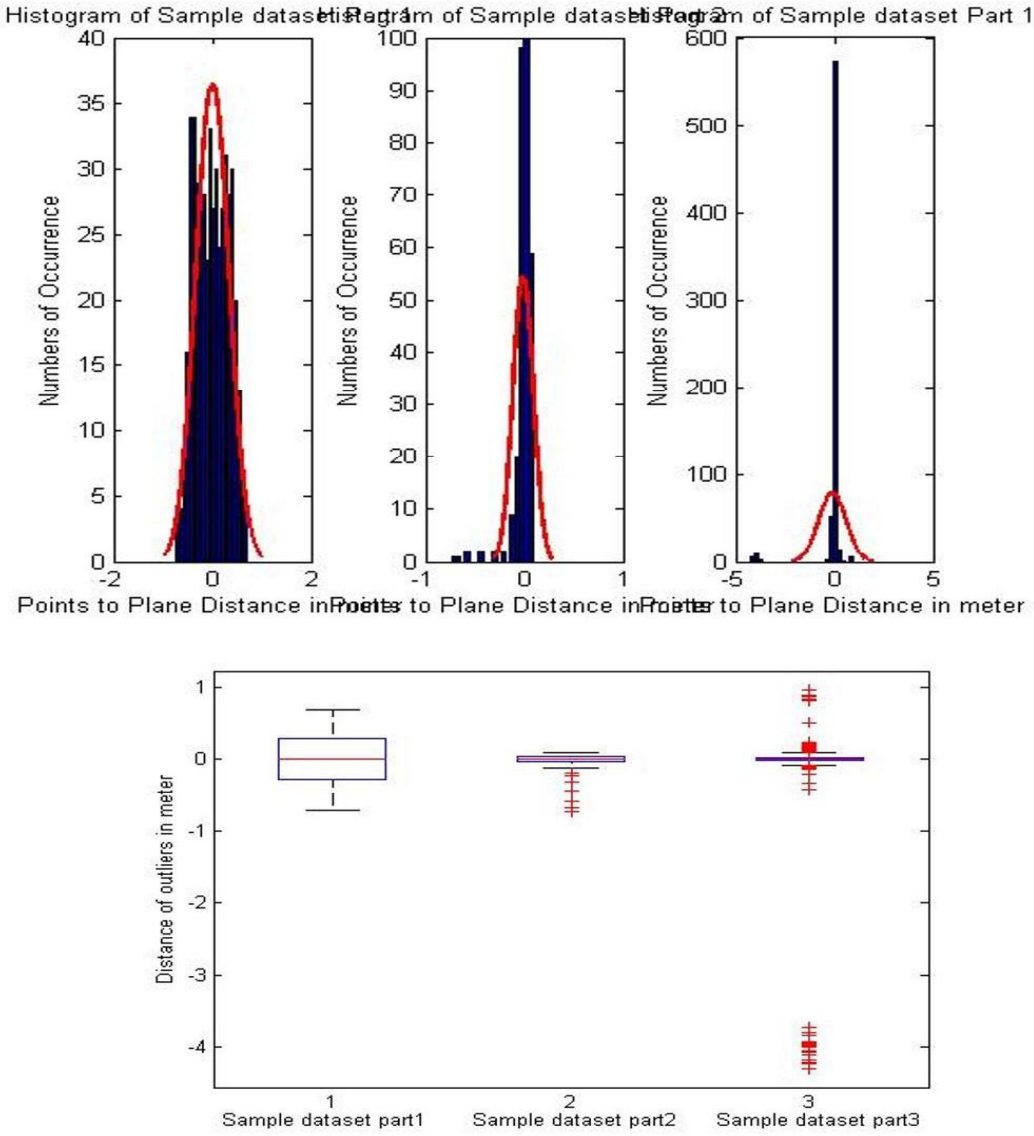


Figure 4.21 Histogram and box plot for three samples

Figure 4.20(a) shows the outliers in blue color in point to plane matching and Figure 4.20(b) scattering of samples in Part 1 sample. The color bar in the Figure 4.20(b) represents the scattering of points from the reference plane in the northing direction. The outliers in point to plane matching and scattering of samples for part 2 are shown in Figure 4.20(c) and 4.20(d) respectively, Figure 4.20(e) and 4.20(f) shows the same information for part3. The histogram and box plot for these sample parts are shown in Figure 4.21. Table 4.3 shows the statistical parameters calculated for the internal accuracy assessment of the point cloud generated using the images taken from the NIKON CoolPix L810 camera. The analysis is same as that performed for the NIKON D60.

Table 4.3 Statistics for internal accuracy assessment of NIKON CoolPix L810

NIKON CoolPix L810			
	Part 1(m)	**Part 2(m)**	**Part 3(m)**
Mean	0.001522	-0.01302	-0.10392
Standard Deviation	0.33359	0.099645	0.68255
Maximum Distance	0.69742	0.088208	0.953
Median	-5.82E-11	0	0
Minimum Distance	-0.72074	-0.73796	-4.296
Variance	0.11128	0.009929	0.46587

4.4.1.5 Internal accuracy assessment of Sony Cybershot DSC-H55

The procedure is same as applied for the internal accuracy assessment of the point cloud generated from the images taken from NIKON D60 camera. The number of points in part 1 sample, part 2 sample and part 3 sample are 723, 266 and 737 respectively.

Figure 4.22(a) shows the outliers in blue color in point to plane matching and Figure 4.22(b) scattering of samples in Part 1 sample. The color bar in the Figure 4.22(b) represents the scattering of points from the reference plane in the northing direction. The outliers in point to plane matching and scattering of samples for part 2 are shown in Figure 4.22(c) and 4.22(d) respectively, Figure 4.22(e) and 4.22(f) shows the same information for part3. The histogram and box plot for these sample parts are shown in Figure 4.23. Table 4.4 shows the statistical parameters calculated for the internal accuracy assessment of the point cloud generated using the images taken from the NIKON CoolPix L810 camera. The analysis is same as that performed for the NIKON D60.

Table 4.4 Statistics for internal accuracy assessment of Sony Cybershot DSC-H55

Sony Cybershot DSC-H55			
	Part 1	**Part 2**	**Part 3**
Mean	0.002959	0.001465	-0.50378
Standard Deviation	0.32856	0.039163	4.9879
Maximum Distance	0.61673	0.12329	13.272
Median	0	0	0
Minimum Distance	-0.70932	-0.10461	-38.891
Variance	0.10795	0.001534	24.879

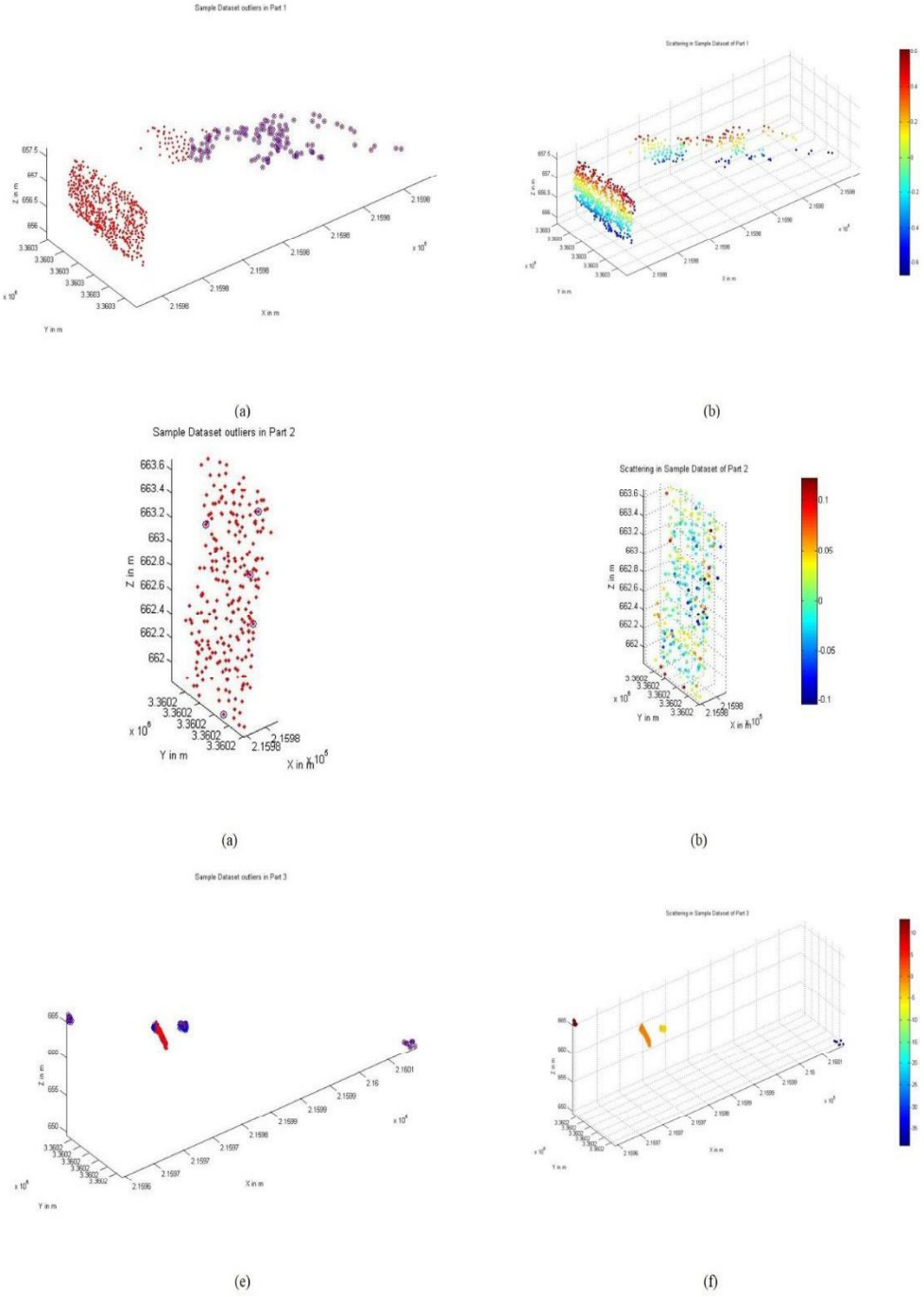


Figure 4.22 (a) Outliers in point to plane matching, (b) scattering of samples in Part 1 sample (c) Outliers in point to plane matching, (d) scattering of samples in Part 2 sample (e) Outliers in point to plane matching, (f) scattering of samples in Part 3 sample

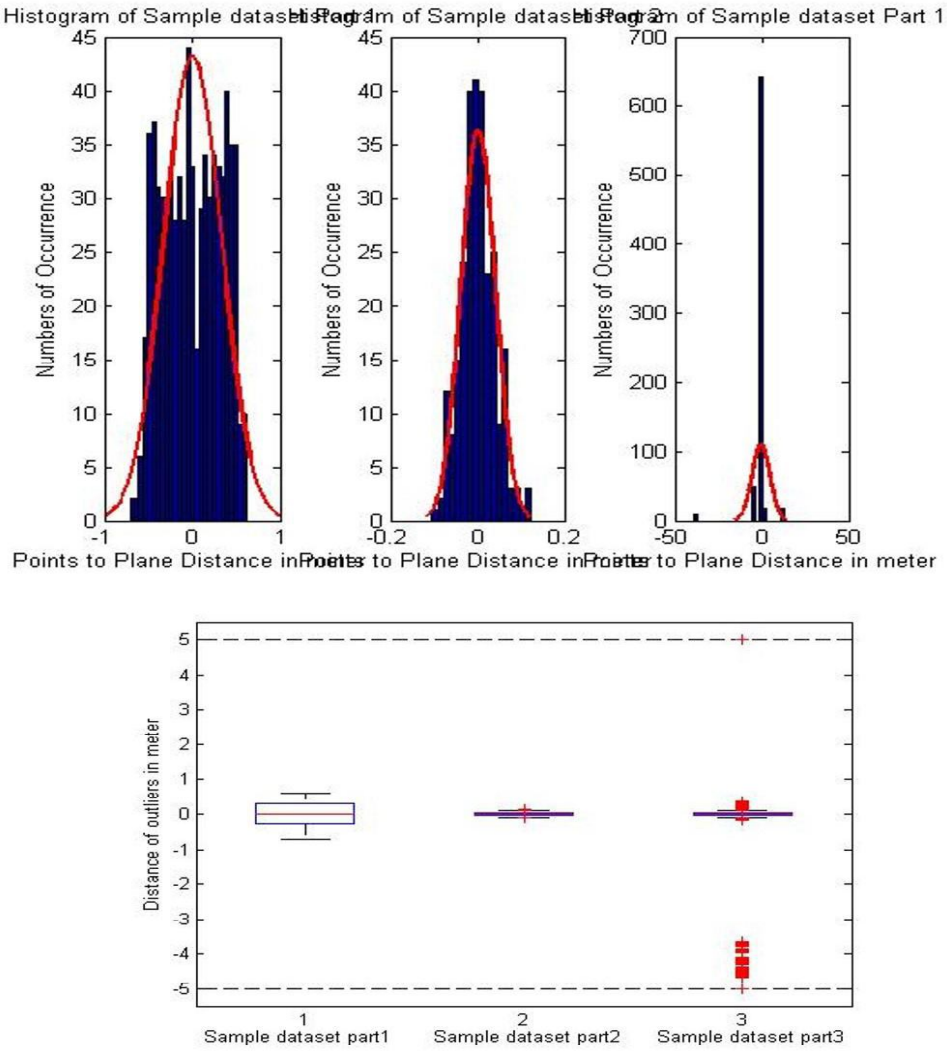


Figure 4.23 Histogram and box plot for three samples

4.4.2 External Accuracy Assessment

As explained in section 3.4.3.6, the global coordinates of few target points were measured with the help of Total Station. The total station was oriented by using the coordinates of two points that were already known by the help of GPS. The georeferenced point cloud generated through the images taken by the conventional cameras is viewed in the Cloud Compare software and the known target points are identified to measure the external accuracy. The locations (Easting, Northing and Height) of few points of the point cloud generated from the images is compared with the corresponding points measured by the Total Station. The locations of total 8 well distributed points are taken from the image point cloud for the comparison with the target points. The errors in the easting, northing and height are estimated for the point clouds generated from the images taken by different grades of conventional cameras. Figure 4.24 shows the chosen target points used for the external accuracy assessment.

Figure 4.24 Point Location used for external accuracy assessment

Table 4.5 shows the RMSE in Easting, Northing and height in meters for different grades of conventional camera used for image data acquisition.

Table 4.5 RMSE in Easting, Northing and Height

Camera Model	RMSE		
	Easting(m)	Northing(m)	Height(m)
NIKON D60	0.29	0.21	0.61
NIKON D90	0.28	0.26	0.50
NIKON CoolPix L810	0.83	0.62	0.97
Sony Cybershot DSC-H55	0.26	0.27	0.50
Video from NIKON D90	3.46	7.07	3.05

4.4 Results for the GPS unavailability problem

For handling the problem of GPS outage or poor GDOP the developed methodology is explained in chapter 3. A matlab code to determine the coordinates of the unknown camera station was developed by implementing the collinearity equations for space intersection. As explained in section 3.4.4, the image coordinates, global coordinates of minimum two points and rotational parameters, principal point location and focal length of the camera were required to obtain the unknown coordinates of camera stations. The coordinates of the required two points were obtained by processing the GPS data in Trimble Business Center software. The horizontal and vertical precision obtained for the first point is 0.6cm and 0.8cm respectively and 0.6cm and 0.9cm respectively for second point. More number of points provide better results. For this study, the main building of IIRS campus was chosen as the test scene. A total of 19 images of the main building were taken for this study. The location of camera station corresponding to the image number 8 was assumed to be unknown.

Image coordinates of two points used = $\begin{bmatrix} 534.2467 & -239.4600 \\ -318.0120 & -239.7760 \end{bmatrix}$

Principal Point coordinates = $\begin{bmatrix} 11.930593 & 8.037962 \end{bmatrix}$

Rotation Parameters= $\begin{bmatrix} 0.9957 & 0.0768 & 0.0520 \\ -0.0747 & 0.9964 & -0.0396 \\ -0.0548 & 0.0355 & 0.9979 \end{bmatrix}$

GCPs= $\begin{bmatrix} 215814.288539 & 3360261.242831 & 655.161734 \\ 215814.032010 & 3360262.015736 & 655.103009 \end{bmatrix}$

Calculated coordinates of exposure station= $\begin{bmatrix} 215804.785031 \\ 3360262.42872 \\ 654.86307229 \end{bmatrix}$

This method gives an error of 12.65m, 11.12m and 0.54m in Easting, Northing and Height. To improve the solution, triangulation and bundle adjustment was applied using the Leica Photogrammetric Suite of Erdas Imagine. For solving the bundle adjustment, the global coordinates of the exposure station (X_L, Y_L, Z_L) and the rotation parameters (omega, phi, kappa) are required. For finding the initial values for the unknown exposure station coordinates few (at least two) few tie points are required on the adjacent images whose exposure station coordinates are known. The coordinates of these tie points are estimated by solving the collinearity equations for space intersection. The same tie points are used to calculate the coordinates of the unknown exposure station by solving the collinearity equations for space intersection. Once the point cloud is generated by using the bundler software the rotation parameters and image coordinates of the tie points are obtained from the bundle.out file. These values are used as the initial values for bundle adjustment in Leica Photogrammetry Suite (LPS).

When 25 control points were taken the error in Easting, Northing and Height of camera station corresponding to image number 8 is shown in Table 4.6.

Table 4.6 Error in Coordinates when 25 control points are taken

Frame No.	Easting(m)	Northing(m)	Height(m)
8	3.25	4.66	2.18

From Table 4.6 it is clear that the bundle adjustment improves the accuracy and also with more number of control points accuracy improves. Figure 4.25 shows the total RMSE obtained after applying triangulation in LPS. This result was obtained in 10 iterations.

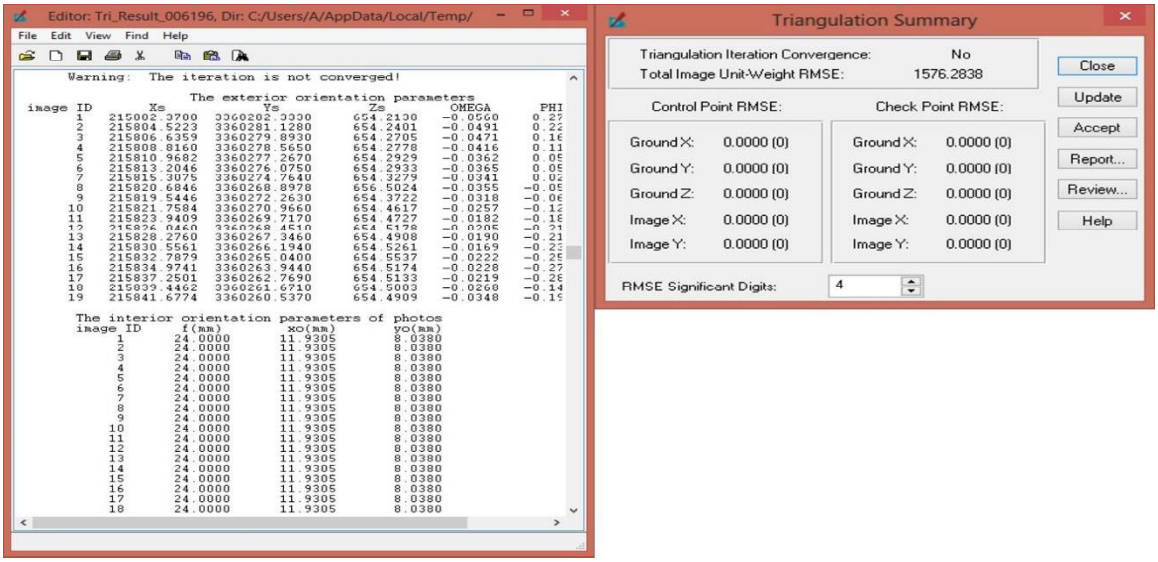


Figure 4.25 RMSE in Triangulation

By reducing the iterations to 5, accuracy is further improved and is shown in Table 4.7.

Table 4.7 RMSE after 5 iterations

Frame No.	Easting(m)	Northing(m)	Height(m)
8	2.20	2.43	1.40

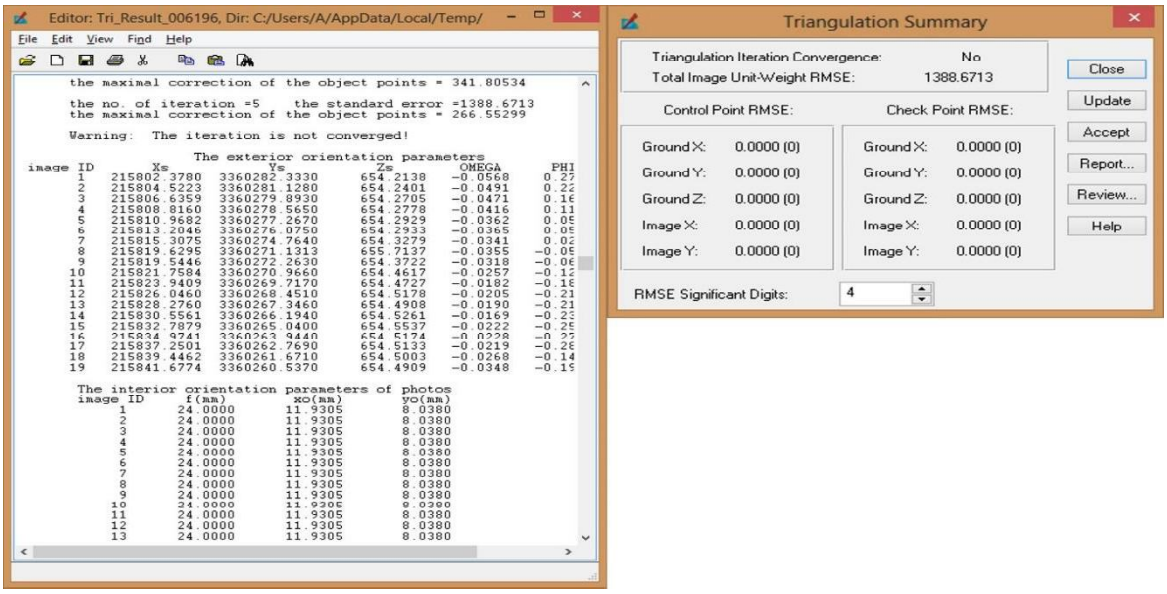


Figure 4.26 RMSE in Triangulation after 5 iterations

Figure 4.26 shows the RMSE in triangulation after 5 iterations.

Approach is also analyzed by reducing the number of Images. Now, total 4 images were taken and the coordinates of forth camera station were assumed to be unknown. Figure 4.27 shows the RMSE in triangulation for the 4 images. Error in Easting, Northing and Height is shown in Table 4.8.

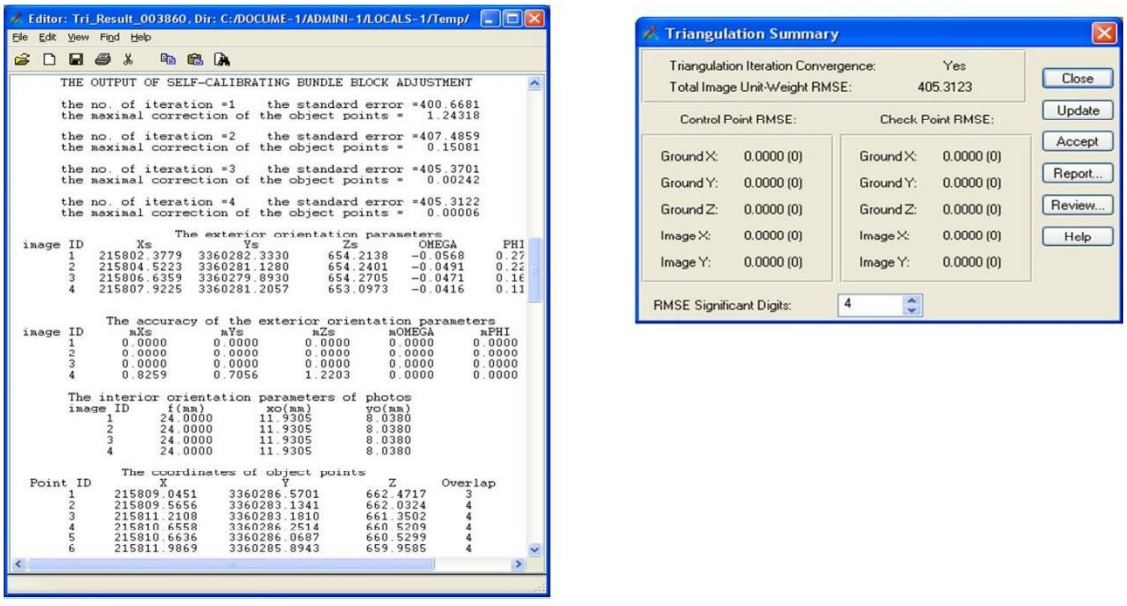


Figure 4.27 RMSE in Triangulation with 4 images

Table 4.8 Error in coordinates of 4th frame

Frame No.	Easting(m)	Northing(m)	Height(m)
4	0.90	2.65	1.19

Now, the approach is applied when a middle camera station is unknown. Here, it is assumed that the third camera station is unknown and the obtained error is shown in Table 4.9.

Table 4.9 Error in coordinates of 3rd frame

Frame No.	Easting(m)	Northing(m)	Height(m)
3	0.19	0.99	0.50

It is clear from the Table 4.9 that the accuracy improves if the coordinates of the adjacent camera stations around the station whose coordinates are unknown.

Consider the case when the coordinates of the two camera stations are unknown. It is assumed that the coordinates of third and fourth camera station are unknown. The error obtained in the coordinates of 3rd and 4th camera station are shown in Table 4.10 and RMSE in triangulation is shown in Figure 4.28.

Table 4.10 Error in coordinates of 3rd and 4th camera station

Frame No.	Easting(m)	Northing(m)	Height(m)
3	0.84	1.95	0.50
4	1.14	3.24	1.72

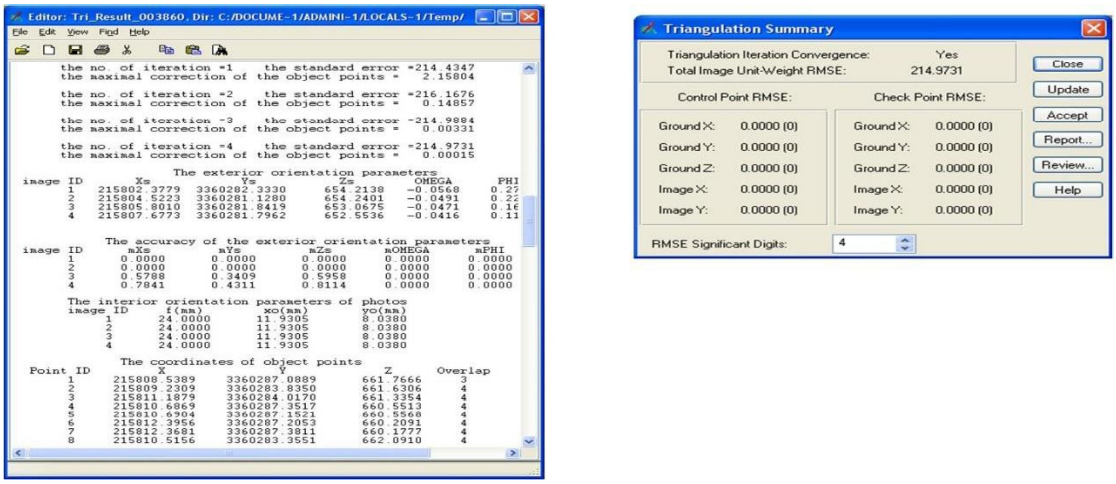


Figure 4.28 RMSE in triangulation when two unknown camera station are present

It is clear from the Table 4.10 and Figure 4.28 that as the number of unknown camera stations increases the accuracy reduces. The error in the coordinates of fourth camera station indicates that the error increases if the adjacent camera stations are unknown.

4.5 Android App for geo-tagging

An android application is developed using Eclipse SDK. The application allows the user to capture geotagged images. It uses the GPS feature of the android phone to obtain the current location. The home screen of the application is shown in Figure 4.29(a). The home screen displays a button which when presses allow the user to capture the images. Once the image is captured, it is stored in the memory card and can be seen on the image viewer available in the home screen of the android app.

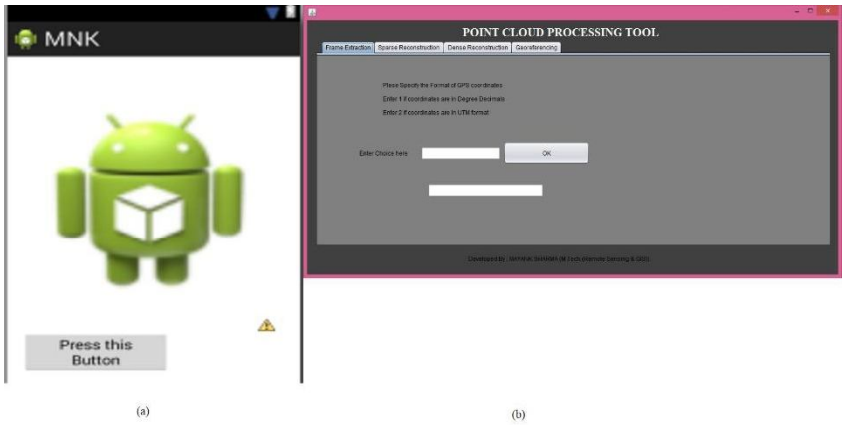


Figure 4.29 (a) Home screen of Android App, (b) Software tool for point cloud processing

4.6 Software tool for Point cloud processing

A GUI (Graphical User Interface) tool is developed in Java to perform the point cloud processing task. GUI consists of a tabbed pane providing different features like frame extraction, sparse point cloud generation, dense reconstruction, georeferencing of the point cloud and internal accuracy assessment. Figure 4.29(b) shows the GUI of the software tool. For the generation of point cloud is exe file of OSM bundler is used. For performing the georeferencing of the point cloud SfM_Georef software is used. The task of geotagging and frame extraction is done using the exe file of code written in Python for geotagging and frame extraction.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This research focuses on the development of the technology to generate the georeferenced point cloud from a video. Structure from Motion approach is used to generate the point cloud and estimate the extrinsic parameters of the camera. Collinearity equations for space intersection are used to georeference the point cloud. Point cloud is also generated from the static photos taken from different grades of conventional cameras. The point cloud generated by using the extracted video frames produces denser point cloud as compare to the static images due to more number of image frames and higher overlap between them. There are 253010 points in the point cloud generated from 442 frames, 298174 points from 552 frames and 24004 points from 35 frames. However, the georeferencing accuracy is very poor as compare to the georeferenced point generated through static images. This is because of the improper geotagging of extracted image frames due to the mismatch between the frame rate of video and logging rate of GPS. Accuracy may improve if proper geotagging is done. The RMSE obtained in Easting, Northing and Height is 3.46m, 7.06m, 3.04m respectively.

5.1.1 Answers to Research Questions

1) *Does video frame sampling improve the performance of SfM –GNSS integration over the static Photos?*

   The point cloud generated using the image frames extracted from a video produces denser point cloud as compare to the point cloud generated using the static images. The point density increases with the increase in the number of frames and vice-versa. There are 253010 points in the point cloud generated from 442 frames, 298174 points from 552 frames and 24004 points from 35 frames. However, the georeferencing accuracy is much degraded in comparison to the georeferenced point cloud generated from the static images. This occurs due to the following reasons:

   - Mismatch between the frame rate of camera and data logging rate of GPS results in the improper geotagging of image frames extracted from the video.
   - Poor accuracy of GPS in kinematic mode.

   The RMSE obtained in Easting, Northing and Height is 3.46m, 7.06m, 3.04m respectively.

2) *How to geo-tag the video frames?*

   A python code using the openCV library is written to automatically extract the image frames from a video and geotag the extracted image frames. This code provides the flexibility to control the frame extraction rate. Figure 5.1 shows the flowchart to geotag the video frames and also shows a geotagged frame. The latitude and the longitude are added to the image metadata by using the EXIF tags.

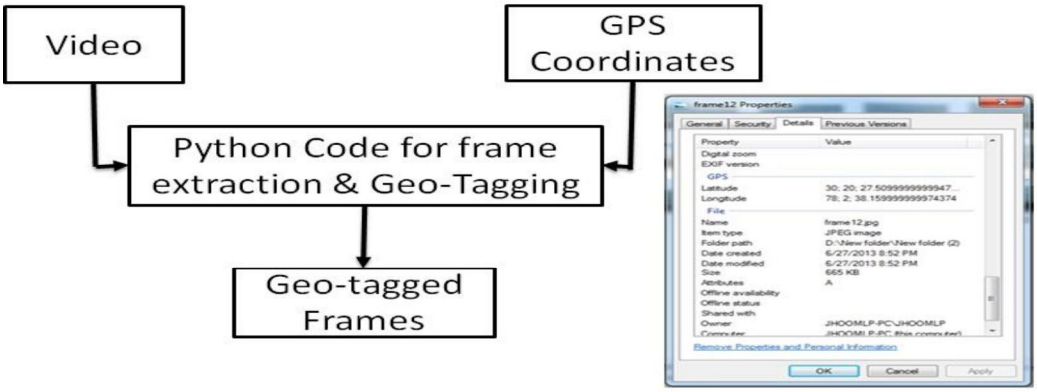


Figure 5.1 Geotagging technique

3) *What is the effect of changing the camera resolution in SfM-GNSS integration?*
Four cameras (NIKON D60, NIKON D90, NIKON Coolpix L810 and Sony cybershot DSC-H55) with different resolutions are used to acquire the images from the same locations. All the images are taken with constant focal lengths. A total of 35 images are taken from each camera and processed to generate the point cloud. Point clouds are georeferenced and the accuracy assessment is done individually for each camera. The results of accuracy assessment are shown in Table 4.5 in chapter 4. The accuracy comparison is shown in Table 5.1. The data shown in the table concludes that the type of camera or sensor type (CMOS or CCD) plays a major role in the accuracy of the point cloud. Also it is clear that improving the camera resolution with same camera type may not improve the accuracy and it is found that around 10-12MP cameras are sufficient to get an accuracy of around 25cm, 27cm and 50cm in easting, northing and height respectively.

Table 5.1 Accuracy Comparison of different grades of camera

Parameters	**NIKON D90**	**NIKON D60**	**Sony Cybershot DSC-H55**	**NIKON Coolpix L810**
Resolution	12.9MP	10.2MP	14.1MP	16.1MP
Type of camera	CMOS	CCD	CMOS	CCD
Focal Length	20mm	18mm	4mm	4mm
Sensor Width	23.6mm	23.6mm	6.17mm	6.16mm
Resolution (pixels)	3872*2592	4288*2848	4320*2432	4608*3456
No. of points generated	133904	128191	87206	86606
Feature identification	Possible	Possible	Possible	Possible
Accuracy	Easting: 0.2870031m Northing: 0.26563m Height: 0.496374m	Easting: 0.297368m Northing: 0.20900m Height: 0.611031m	Easting: 0.259153m Northing: 0.27398m Height: 0.50123m	Easting: 0.82854m Northing: 0.6194m Height: 0.97407m

4) *How to geo-reference the point cloud from camera station under GNSS unavailable environment?*
   The methodology is developed to find out the unknown locations of camera point. The unknown location is estimated using the collinearity equations for space intersection. To improve the accuracy bundle adjustment is applied. Once the unknown location of the camera points is obtained then the georeferencing of the point cloud is done.

5.2 Recommendations

1) The accuracy of the point cloud generated from the video frames is very low because of the improper geo-tagging due to the mismatch between frame rate of video and the data logging rate of the GPS. Either use the camera that provides automatic video tagging or use the GPS that have high logging rate.
2) In this study, the frames that are extracted from the video have poor resolution. An approach should be developed to extract the frames without degrading the resolution.
3) Instead of extracting the image frames, an approach could be develop to directly generate the point cloud.

REFERENCES

Android (operating system). (2014, April 18). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Android_(operating_system)&oldid=604712089

Computer vision. (2014, February 20). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Computer_vision&oldid=596272139

Fathi, H., & Brilakis, I. (2011). Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features. *Advanced Engineering Informatics*, *25*(4), 760–770. doi:10.1016/j.aei.2011.06.001

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.

Global Positioning System. (2014, March 10). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Global_Positioning_System&oldid=599015941

GPS Positioning Guide. (1995). Ottawa, Ontario: Natural Resources Canada. Retrieved from http://www.geod.nrcan.gc.ca/

Han, B., Paulson, C., & Wu, D. (2011). 3D dense reconstruction from 2D video sequence via 3D geometric segmentation. *Journal of Visual Communication and Image Representation*, *22*(5), 421–431. doi:10.1016/j.jvcir.2011.03.006

Hassan, T., Ellum, C., & El-Sheimy, N. (2006). Bridging land-based mobile mapping using photogrammetric adjustments. In *Proceedings of the ISPRS Commission I Symposium ''From Sensors to Imagery*. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.222.3424

Horn, B. K. (2000). Tsai's camera calibration method revisited. *Online: Http://people. Csail. Mit. edu/bkph/articles/Tsai_Revisited. Pdf*. Retrieved from http://200.2.147.154/courses/electrical-engineering-and-computer-science/6-801-machine-vision-fall-2004/readings/tsaiexplain.pdf

İmre, E., Knorr, S., Özkalaycı, B., Topay, U., Aydın Alatan, A., & Sikora, T. (2007). Towards 3-D scene reconstruction from broadcast video. *Signal Processing: Image Communication*, *22*(2), 108–126. doi:10.1016/j.image.2006.11.011

ISPRS. (2010). 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. Retrieved from http://isprsv6.lboro.ac.uk/tips.html

Jayson Jayeshkumar Jariwala. (2013, March). *Mobile Mapping by integrating Structure from Motion approach with Global Navigation Satellite System*. University of Twente.

Karras, G. E., & Mavrommati, D. (2002). Simple calibration techniques for non-metric cameras. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, *34*(5/C7), 39–46.

Lee, A. J. T., Yu, P., Chiu, H.-P., & Hong, R.-W. (2005). 3D Z-string: A new knowledge structure to represent spatio-temporal relations between objects in a video. *Pattern Recognition Letters*, *26*(16), 2500–2508. doi:10.1016/j.patrec.2005.04.018

Li, R. (1997). Mobile mapping: An emerging technology for spatial data acquisition. *Photogrammetric Engineering and Remote Sensing*, *63*(9), 1085–1092.

Lourakis, M. I. A., & Argyros, A. A. (2009). SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software*, *36*(1), 1–30. doi:10.1145/1486525.1486527

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (Vol. 2, pp. 1150–1157). Ieee. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=790410

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Marry E. Greece. (2000, April). Global Positioning System. Retrieved from http://infohost.nmt.edu/~mreece/gps/cover.html

Matthew Brown, & David Lowe. (2002). Invariant features from interest point groups (pp. 656–665). Presented at the British Machine Vision Conference, Cardiff Wales.

Novak, K. (1993). Application of digital cameras and GPS for aerial photogrammetric mapping. *International Archives of Photogrammetry and Remote Sensing*, *29*, 5–5.

Pinhole camera model. (2014, May 3). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Pinhole_camera_model&oldid=596336927

Point cloud. (2014, March 15). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Point_cloud&oldid=596051400

Pollefeys, M., Koch, R., & Van Gool, L. (1999). Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, *32*(1), 7–25.

Scale-invariant feature transform. (2014, February 22). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Scale-invariant_feature_transform&oldid=585477228

Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, & Angela Y. Wu. (1998). An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*.

Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer.

Tao, C. V., Chapman, M. A., & Chaplin, B. A. (2001). Automated processing of mobile mapping image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, *55*(5–6), 330–346. doi:10.1016/S0924-2716(01)00026-0

Tian, Y., Gerke, M., Vosselman, G., & Zhu, Q. (2010). Knowledge-based building reconstruction from terrestrial video sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(4), 395–408. doi:10.1016/j.isprsjprs.2010.05.001

Tissainayagam, P., & Suter, D. (2005). Object tracking in image sequences using point features. *Pattern Recognition*, *38*(1), 105–113.

Yang, M.-D., Chao, C.-F., Huang, K.-S., Lu, L.-Y., & Chen, Y.-P. (2013). Image-based 3D scene reconstruction and exploration in augmented reality. *Automation in Construction*, *33*, 48–60. doi:10.1016/j.autcon.2012.09.017

Zhang, Z. (2000). A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, *22*(11), 1330–1334.

Zhou, F., Cui, Y., Gao, H., & Wang, Y. (2013). Line-based camera calibration with lens distortion correction from a single image. *Optics and Lasers in Engineering*, *51*(12), 1332–1343. doi:10.1016/j.optlaseng.2013.05.010

APPENDICES

Appendix-1 Exposure Station coordinates

Table A.1 Coordinate of the Exposure Station locations marked in front of Godavari Building

Exposure Station ID	Easting(m)	Northing(m)	Height(m)
1	215930.509	3360196.52	656.785
2	215935.158	3360202.059	656.969
3	215939.101	3360202.105	656.972
4	215938.476	3360206.048	656.999
5	215940.439	3360208.179	657.055
6	215942.411	3360209.569	656.984
7	215942.803	3360212.885	657.014
8	215943.476	3360216.543	657.129
9	215942.906	3360219.216	657.16
10	215945.966	3360223.486	656.88
11	215947.794	3360226.643	656.977
12	215950.975	3360228.72	656.984
13	215951.713	3360231.268	657.086
14	215952.76	3360232.963	657.081
15	215956.744	3360232.809	656.078
16	215956.405	3360235.492	656.329
17	215959.449	3360239.308	656.481
18	215960.691	3360242.633	656.492
19	215962.122	3360244.377	656.5
20	215962.679	3360246.154	656.488
21	215963.623	3360248.982	656.478
22	215964.072	3360251.359	656.507
23	215964.943	3360253.81	656.499
24	215965.861	3360256.136	656.491
25	215966.448	3360257.87	656.495
26	215967.008	3360259.712	656.492
27	215967.424	3360262.106	656.524
28	215968.652	3360264.149	656.502
29	215972.411	3360265.069	656.355
30	215973.296	3360268.339	656.394
31	215972.208	3360271.872	656.68
32	215973.081	3360274.272	656.667
33	215974.244	3360276.696	656.638
34	215975.426	3360279.512	656.475
35	215976.5	3360283	656.557

Appendix-2 Accuracy assessment between TLS points and Generated point cloud using NIKON D60

Table A.2 Coordinates of Test points taken for Accuracy Assessment

Point Id	TLS			Point Cloud D60		
	Easting(m)	Northing(m)	Height(m)	Easting(m)	Northing(m)	Height(m)
1	215977.794	3360252.548	658.551	215977.7969	3360252.5	657.523621
2	215977.663	3360250.292	661.731	215977.7031	3360250	660.997803
3	215969.314	3360228.815	664.224	215969.8438	3360228.75	662.330688
4	215976.927	3360234.806	671.039	215977.375	3360234.25	668.827026
5	215967.533	3360213.327	671.096	215968.1094	3360213.5	668.929321
6	215969.459	3360238.096	659.973	215969.5313	3360238	659.21167
7	215976.532	3360250.11	671.101	215977.4375	3360249.75	669.026306
8	215973.041	3360237.896	663.25	215973.2813	3360237.75	662.190674

Table A.3 Error in Easting, Northing and Height

Error in Easting (m)	Error in Northing(m)	Error in Height (m)
0.002875	-0.048	-1.027379
0.040125	-0.292	-0.733197
0.52975	-0.065	-1.893312
0.448	-0.556	-2.211974
0.576375	0.173	-2.166679
0.07225	-0.096	-0.76133
0.9055	-0.36	-2.074694
0.24025	-0.146	-1.059326

Appendix-3 Accuracy assessment between TLS points and Generated point cloud using NIKON D90

Table A.4 Coordinates of Test points taken for Accuracy Assessment

Point Id	TLS			Point Cloud D90		
	Easting(m)	Northing(m)	Height(m)	Easting(m)	Northing(m)	Height(m)
1	215977.794	3360252.548	658.551	215977.7969	3360252.25	657.3943
2	215977.663	3360250.292	661.731	215977.9531	3360250.25	660.5467
3	215969.314	3360228.815	664.224	215969.4063	3360228.75	662.2935
4	215976.927	3360234.806	671.039	215977.1563	3360234.5	669.2051
5	215967.533	3360213.327	671.096	215967.6406	3360213.75	669.1592
6	215969.459	3360238.096	659.973	215969.4844	3360237.75	659.4269
7	215976.532	3360250.11	671.101	215977.4844	3360250.25	669.1552
8	215973.041	3360237.896	663.25	215973.375	3360237.5	662.1696

Table A.5 Error in Easting, Northing and Height

Error in Easting (m)	Error in Northing(m)	Error in Height (m)
0.002875	-0.298	-1.156713
0.290125	-0.042	-1.184346
0.09225	-0.065	-1.930482
0.22925	-0.306	-1.833861
0.107625	0.423	-1.936759
0.025375	-0.346	-0.54612
0.952375	0.14	-1.945758
0.334	-0.396	-1.080383

Appendix-4 Accuracy assessment between TLS points and Generated point cloud using NIKON CoolPix L810

Table A.6 Coordinates of Test points taken for Accuracy Assessment

Point Id	TLS			Point Cloud CoolPix L810		
	Easting(m)	Northing(m)	Height(m)	Easting(m)	Northing(m)	Height(m)
1	215977.794	3360252.548	658.551	215977.75	3360252	657.0601
2	215977.663	3360250.292	661.731	215977.5469	3360249.25	660.1569
3	215969.314	3360228.815	664.224	215969.7656	3360229.25	661.71
4	215976.927	3360234.806	671.039	215976.625	3360234	667.1805
5	215967.533	3360213.327	671.096	215969.7813	3360214	667.7358
6	215969.459	3360238.096	659.973	215969.2656	3360238	659.0425
7	215976.532	3360250.11	671.101	215977.7969	3360249	668.0498
8	215973.041	3360237.896	663.25	215973.2656	3360237.75	661.5591

Table A.7 Error in Easting, Northing and Height

Error in Easting (m)	Error in Northing(m)	Error in Height (m)
-0.044	-0.548	-1.49088
-0.116125	-1.042	-1.57414
0.451625	0.435	-2.513978
-0.302	-0.806	-3.858519
2.24825	0.673	-3.36016
-0.193375	-0.096	-0.930458
1.264875	-1.11	-3.051195
0.224625	-0.146	-1.690918

Appendix-5 Accuracy assessment between TLS points and Generated point cloud using Sony Cybershot DCS-H55

Table A.8 Coordinates of Test points taken for Accuracy Assessment

Point Id	TLS			Point Cloud Sony Cybershot		
	Easting(m)	Northing(m)	Height(m)	Easting(m)	Northing(m)	Height(m)
1	215977.794	3360252.548	658.551	215977.7969	3360252.25	657.5537
2	215977.663	3360250.292	661.731	215977.8438	3360250	660.8886
3	215969.314	3360228.815	664.224	215970.0156	3360228.75	662.3153
4	215976.927	3360234.806	671.039	215977.1406	3360234.75	669.531
5	215967.533	3360213.327	671.096	215968.1563	3360213.75	669.1948
6	215969.459	3360238.096	659.973	215969.4844	3360238	659.532
7	215976.532	3360250.11	671.101	215977.0156	3360249.5	669.5792
8	215973.041	3360237.896	663.25	215973.1094	3360237.75	662.3564

Table A.9 Error in Easting, Northing and Height

Error in Easting (m)	Error in Northing(m)	Error in Height (m)
0.002875	-0.298	-0.99735
0.18075	-0.292	-0.842389
0.701625	-0.065	-1.908692
0.213625	-0.056	-1.507994
0.62325	0.423	-1.901237
0.025375	-0.096	-0.441018
0.483625	-0.61	-1.521776
0.068375	-0.146	-0.893555

Appendix-6 Accuracy assessment between TLS points and Generated point cloud using Video

Table A.10 Coordinates of Test points taken for Accuracy Assessment

Point Id	TLS			Point Cloud by Video		
	Easting(m)	Northing(m)	Height(m)	Easting(m)	Northing(m)	Height(m)
1	215977.794	3360252.548	658.551	215965.0469	3360248	657.0753
2	215977.663	3360250.292	661.731	215965.2188	3360246.75	658.2596
3	215969.314	3360228.815	664.224	215961.4063	3360237.5	660.84
4	215976.927	3360234.806	671.039	215962.0781	3360242.5	661.9763
5	215967.533	3360213.327	671.096	215964	3360230.75	662.1576
6	215969.459	3360238.096	659.973	215962.4531	3360241.25	657.8248
7	215976.532	3360250.11	671.101	215965.1094	3360246.75	662.0427
8	215973.041	3360237.896	663.25	215964.2656	3360241.5	659.1089

Table A.11 Error in Easting, Northing and Height

Error in Easting (m)	Error in Northing(m)	Error in Height (m)
-12.747125	-4.548	-1.475683
-12.44425	-3.542	-3.471356
-7.90775	8.685	-3.384034
-14.848875	7.694	-9.062682
-3.533	17.423	-8.938407
-7.005875	3.154	-2.148232
-11.422625	-3.36	-9.058336
-8.775375	3.604	-4.141052