

**Improved Monocular Videogrammetry for Generating 3D Dense Point
Clouds of Built Infrastructure**

A Dissertation
Presented to
The Academic Faculty

By

Abbas Rashidi

In Partial Fulfillment
of the Requirements for the Degree
Ph.D. in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
August 2014

Copyright © 2014 by Abbas Rashidi

Improved Monocular Videogrammetry for Generating 3D Dense Point Clouds of Built Infrastructure

Approved by:

Dr. Reginald DesRoches, Advisor
School of Civil & Environmental
Engineering
Georgia Institute of Technology

Dr. Ioannis Brilakis
Department of Engineering
University of Cambridge

Dr. Nelson Baker
School of Civil & Environmental
Engineering
Georgia Institute of Technology

Dr. Patricio Vela
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Dr. Yong Cho
School of Civil & Environmental
Engineering
Georgia Institute of Technology

Date Approved: May 01, 2014

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to acknowledge and thank my advisors, committee members, industry partners, fellow students and my family members for all of their support and guidance throughout this process these past couple of years. I would like to express my deepest sense of gratitude to my advisor, Dr. Ioannis Brilakis, for seeing my potential and showing me the path to my academic aspirations. Many times when I felt I had reached a dead-end, Dr. Brilakis analyzed the problem from a broader perspective and presented new directions to follow. I am also grateful for my co-advisors, Dr. Reginald DesRoches and Dr. Patricio Vela, whose knowledge and determination have allowed me to achieve more than I thought was possible. Their advice and suggestions have been invaluable. Next I am fortunate to have Dr. Nelson Baker and Dr. Yong Cho as members of my thesis committee. They are truly world-class researchers and their feedback helped me gain new understandings about my research. I want to acknowledge the support of the faculty members at the Department of Civil Engineering and Construction Management at Georgia Southern University, where I have served as a visiting faculty member during the 2013-2014 academic year. I especially want to thank Chair Mike Jackson, whose support and knowledge has helped me to focus more on this research project. To Ms. Emily Christian, thank you for continuous support in different forms by partially proof reading my dissertation and providing very useful comments. Finally, I am deeply grateful to my parents for their hard work and sacrifice for the gift of education. They have encouraged me to believe I can succeed in whatever path I choose and they nurtured my curiosity from an early age. Thank you for the enthusiastic support every step of the way.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
SUMMARY	xi
<u>CHAPTERS</u>	
1. INTRODUCTION	1
1.1. Indoor Applications	4
1.2. Outdoor Applications	5
2. ANALYSIS	7
2.1. State of Practice: Close-Range Spatial Sensing of Civil Infrastructure	7
2.1.1. Time-of-Flight-Based Technologies	7
2.1.2. Computer Vision-Based Technologies	8
2.2. Comparison of Image- and Time-of-Flight- Based Technologies for 3D As Built Documentation of Indoor/Outdoor Scenes	10
2.2.1. Software and Hardware Selection	11
2.2.2. Ground Truth Data Collection	11
2.2.3. Evaluation of Accuracy	12
2.2.4. Evaluation of Quality	13
2.2.5. Time and Cost Estimation	14
2.2.6. Experiment and Results	14

2.3. State of Research: Applications of Monocular Videogrammetry in Construction	20
2.3.1. Strategies for Quality Control of Video Frames	22
2.3.2. Strategies for Key Frame Selection	24
2.3.3. Strategies for Computing the Absolute Scale of Dense Point Clouds	25
2.3.4. Strategies for Improving the Quality of Generated Point Clouds	32
2.4. Problem Statement and Objectives	37
2.5. Scope of Research	39
3. HYPOTHESIS	40
3.1. Overview	40
3.2. Optimized Selection of Key Video Frames	41
3.3. Automated Calculation of Absolute Scale of Point Clouds	46
3.3.1. Automated Absolute Scale Computation for Outdoor Settings (proposed solutions)	46
3.3.2. Automated Absolute Scale Computation for Indoor Settings (proposed solutions)	54
3.4. Point Cloud Data Cleaning	56
4. SYNTHESIS	61
4.1. Prototype	61
4.2. Basic Videogrammetric Pipeline for Generating Dense 3D Point Clouds	61
4.3. Key Frame Selection: Implementation	62
4.3.1. Identifying the Thresholds	62

4.3.2. Strategies for Low Quality Frames Filtering	63
4.4. Automated Computation of Absolute Scale: Implementation	68
4.4.1. Identifying Thresholds for the Minimum Acceptable Area of the Cube in Images	68
4.4.2. Identifying Thresholds for the Maximum and Minimum Roundness Factors	72
5. VALIDATION	74
5.1. Validating the Key Frame Selection Algorithm	74
5.2. Validating the Proposed Absolute Scale Calculation Algorithm	76
5.3. Validating the Improved Videogrammetric Pipeline for As-Built Documentation of Built Infrastructure	80
6. CONCLUSION	97
6.1. Conclusions and Remarks	97
6.2. Plans for Future Experiments	98
6.3. Achieved Contribution	98
REFERENCES	99

LIST OF TABLES

	Page
Table 2.1: Estimated camera calibration parameters in different experiments	16
Table 2.2: Time and cost estimation of different methods	17
Table 2.3: Accuracy, completeness, and density of videogrammetry under different settings	18
Table 2.4: Accuracy, completeness, and density of different methods (arch bridge case study)	19
Table 2.5: Accuracy, completeness, and density of different methods (indoor case studies)	20
Table 2.6: An overview of existing algorithms for filling gaps/holes on PCD	33
Table 4.1: Upper and lower thresholds for corresponding ratios	66
Table 5.1: Impact of blur on number of extracted feature points and re-projection errors	75
Table 5.2: Failure percentages and re-projection errors for different key frame extraction methods	75
Table 5.3: Summary of the results obtained from implementing the corner detection and matching algorithms	77
Table 5.4: Summary of the results obtained from evaluating the overall performance of the proposed method	80
Table 5.5: Summary of comparison results for 5 different datasets	82
Table 5.6: Summary of evaluation results for the 5 different outdoor datasets	91
Table 5.7: Summary of evaluation results for the 4 different indoor datasets	95
Table 5.8: Required levels of accuracy for multiple applications in AEC/FM domain (Dai et al., 2013)	96

LIST OF FIGURES

	Page
Figure 1.1:	Different types of 3D reconstruction methods based on the number of applied sensors: monocular (left) and binocular (right) 2
Figure 1.2:	Poor quality of a generated point cloud: Existence of holes 3
Figure 1.3:	3D point clouds of different indoor environments: a) office corridor including electrical and plumbing facilities, b) typical bedroom 5
Figure 1.4:	3D point cloud of a concrete column beam bridge 6
Figure 2.1:	Overview of the comparison workflow 10
Figure 2.2:	The point P is supposed to locate on the surface s of the ground truth model, but due to imperfection of the point cloud, there is a distance d between P and s . d is deemed as error of this point 13
Figure 2.3:	a) an offset Δ (exaggerated for illustration) is made on both sides of the edge l to form a region marked with dash lines, and b) a frustum is generated by extruding the region along the axis orthogonal to the surface 14
Figure 2.4:	Snapshots of 3D point clouds of the beam bridge 15
Figure 2.5:	Point cloud registration 16
Figure 2.6:	Snapshots of 3D point clouds of the arch bridge 19
Figure 2.7:	Projector and camera setup for extracting absolute measurements in manufacturing industry (Ganci and Brown, 2008) 28
Figure 2.8:	Pivot Ball Algorithm 36
Figure 3.1:	General framework for the proposed methodology 40
Figure 3.2:	Workflow of the key frame selection algorithm 42
Figure 3.3:	Selection of final candidate as the next key frame 44
Figure 3.4:	Acceptable ranges for upper and lower corresponding ratio thresholds 46
Figure 3.5:	Selected colors for surfaces of the cube (top) and snapshots of the cube (bottom) 47

Figure 3.6:	Overall workflow of the proposed algorithm for computing absolute scale of PCD	48
Figure 3.7:	Necessary steps for detection of the cube vertices	49
Figure 3.8:	Convex hull algorithm: a) non-convex shape, b) constructing an equal convex hull for the initial shape and c) reconstructed convex hull shape	50
Figure 3.9:	Corresponding corner points for the first view (left) are located on epipolar line in the next view (right)	52
Figure 3.10:	Possible locations for the letter-size sheet of paper in indoor settings	55
Figure 3.11:	Locations of corner points of the sheet of paper follow the same clockwise order in different views	55
Figure 3.12:	Overview of the proposed algorithm for point cloud data cleaning	56
Figure 3.13:	Different stages for 3D edge based reconstruction of holes	58
Figure 3.14:	Definition of hole (top) and definition of 3D edges on point cloud data (bottom)	59
Figure 3.15:	Different stages for balancing the density of point clouds	60
Figure 4.1:	Samples of corresponding ratios for different video streams	63
Figure 4.2:	Uniform scenes with texture-less areas (top) and complex scenes with texture-full areas (bottom)	65
Figure 4.3:	Completeness of generated dense point clouds versus number of processes key frames	66
Figure 4.4:	Proper resolution for outdoor (top) and indoor (bottom) settings	68
Figure 4.5:	Precision and recall ratios for detection of the cube surfaces (top) and sheet of paper (bottom)	71
Figure 4.6:	2D location errors (top) and re-projection errors (bottom) for both indoor and outdoor settings	72
Figure 5.1:	Sample of the implementation results for the cube corners detection algorithm: from top left: the original image of the cube – the results of filtering the image based on HSV thresholds – detected red, yellow and purple surfaces – detected lines based on the improved Hough transform –	78

and the intersections of the cube edges as the final results

Figure 5.2:	Actual distance measurements and preparation of ground truth: Leica TC805 total station (left) and Leica DISTO D5Laser measurer (middle and right)	79
Figure 5.3:	A sample of the generated PCD for indoor settings: bathroom- Sparse PCD generated by SfM (left) and PCD generated by PMVS (right)	79
Figure 5.4:	Samples of the generated PCD for outdoor settings: Campus building (top) and construction wall (bottom)	79
Figure 5.5:	Different stages for detecting and removing outliers	81
Figure 5.6:	Implementing the hole filling algorithm for one dataset as a case study	82
Figure 5.7.A:	Different approaches for filling holes on surfaces of point cloud data MLS (left), VD (middle) and 3D edge based (right)	84
Figure 5.7.B:	Results of filling the hole using MLS method	85
Figure 5.7.C:	Results of filling the hole using Volumetric Diffusion method	85
Figure 5.7.D:	Results of filling the hole using 3D edge based method	86
Figure 5.8:	Outdoor case study #01: steel bridge with concrete columns	87
Figure 5.9:	Outdoor case study #02: concrete arch bridge: point cloud data before improvement (middle) and after improvement (bottom)	88
Figure 5.10:	Outdoor case study #03: masonry wall in a construction site	89
Figure 5.11:	Outdoor case study #04: campus building	90
Figure 5.12:	Outdoor case study #05: country building	91
Figure 5.13:	Indoor case study #01: kitchen	93
Figure 5.14:	Indoor case study #02: bathroom	94
Figure 5.15:	Indoor case study #03: bedroom	94
Figure 5.16:	Indoor case study #04: office corridor	95

Summary

Videogrammetry is an affordable and easy-to-use technology for spatial 3D scene recovery. When applied to the civil engineering domain, a number of issues have to be taken into account.

First, videotaping large scale civil infrastructure scenes usually results in large video files filled with blurry, noisy, or simply redundant frames. This is often due to higher frame rate over camera speed ratio than necessary, camera and lens imperfections, and uncontrolled motions of the camera that results in motion blur. Only a small percentage of the collected video frames are required to achieve robust results. However, choosing the right frames is a tough challenge.

Second, the generated point cloud using a monocular videogrammetric pipeline is up to scale, i.e. the user has to know at least one dimension of an object in the scene to scale up the entire scene. This issue significantly narrows applications of generated point clouds in civil engineering domain since measurement is an essential part of every as-built documentation technology.

Finally, due to various reasons including the lack of sufficient coverage during videotaping of the scene or existence of texture-less areas which are common in most indoor/outdoor civil engineering scenes, quality of the generated point clouds are sometimes poor. This deficiency appears in the form of outliers or existence of holes or gaps on surfaces of point clouds. Several researchers have focused on this particular problem; however, the major issue with all of the currently existing algorithms is that they basically treat holes and gaps as part of a smooth surface. This approach is not

robust enough at the intersections of different surfaces or corners while there are sharp edges. A robust algorithm for filling holes/gaps should be able to maintain sharp edges/corners since they usually contain useful information specifically for applications in the civil and infrastructure engineering domain.

To tackle these issues, this research presents and validates an improved videogrammetric pipeline for as built documentation of indoor/outdoor applications in civil engineering areas. The research consists of three main components:

1. Optimized selection of key frames for processing. It is necessary to choose a number of informative key frames to get the best results from the videogrammetric pipeline. This step is particularly important for outdoor environments as it is impossible to process a large number of frames existing in a large video clip.
2. Automated calculation of absolute scale of the scene. In this research, a novel approach for the process of obtaining absolute scale of points cloud by using 2D and 3D patterns is proposed and validated.
3. Point cloud data cleaning and filling holes on the surfaces of generated point clouds. The proposed algorithm to achieve this goal is able to fill holes/gaps on surfaces of point cloud data while maintaining sharp edges.

In order to narrow the scope of the research, the main focus will be on two specific applications:

1. As built documentation of bridges and building as outdoor case studies.
2. As built documentation of offices and rooms as indoor case studies.

Other potential applications of monocular videogrammetry in the civil engineering domain are out of scope of this research. Two important metrics, i.e. accuracy, completeness and processing time, are utilized for evaluation of the proposed algorithms.

CHAPTER 1

INTRODUCTION

Over the past few years, computer vision-based technologies have been considered as cost-effective and easy-to-use alternatives of traditional spatial sensing methods, e.g., laser scanning and range cameras (Zhu & Brilakis, 2009). In the Architecture, Engineering and Construction (AEC) domain, the collected spatial data is useful for designers, engineers and inspectors to control/verify quality issues of construction projects (Dai & Lu, 2010) (Quiñones-Rozo, et al., 2008); identify deviations between as-built and as-designed structures (Klein et al., 2011); design site layouts more efficiently (Zhu & Brilakis, 2009); monitor the progress of projects in a more proactive manner (Golparvar-Fard, et al., 2009) (Kim & Kano, 2008); and assess damage caused by disasters (Memon, et al., 2005). Computer vision-based 3D reconstruction methods are generally based on processing a number of captured images (photogrammetry) or video streams (videogrammetry) from the scene. In comparison with arbitrarily taken images, sequential video streams provide more valuable information for processing (Kien, 2005). In the AEC domain, it is more convenient to videotape a relatively large civil infrastructure scene rather than taking hundreds or thousands of images.

Based on the number of applied sensors, 3D reconstruction algorithms are divided into three main categories (Figure 1.1):

1. Using a single camera or monocular video/photogrammetry (Kien, 2005)
2. Using a set of stereo cameras or binocular video/photogrammetry (Greenwood, 1999)
3. Using multiple cameras or camera rig (Pappa, et al., 2003)



Figure 1.1: Different types of 3D reconstruction methods based on the number of applied sensors: monocular (left) and binocular (right)

In practical cases, especially for civil engineering applications, using a single camera is the most straight forward, easy to use scenario since there is no need for extra settings. As the result, the main focus of this research is monocular videogrammetry, which is defined as generating dense point clouds of objects by processing video frames captured by a single off-the-shelf camera.

While videogrammetry works well for some specific applications, e.g. conducting measurements and visualization in manufacturing industry, applying it to civil infrastructure domain faces several practical constraints, therefore preventing its adoption in real construction practices (Brilakis, et al., 2011). Poor quality of captured frames is a major concern that significantly undermines the performance of the generated 3D point clouds. Unlike specific controlled areas in factories, it is very difficult to control the environment factors, e.g. speed of camera, light conditions and occlusion, in the civil engineering domain. Motion blur is inevitable due to random movements of cameras with different speeds. Moreover, considering size and level of complexity, videotaping civil infrastructure usually takes several minutes instead of a few seconds. This means there are millions of frames that need to be furthered processed. Estimating the camera poses and 3D scene structures is computationally expensive if it is performed with all the frames in a video sequence. If these frames can be decimated, then this process can become more efficient. Thus far, there is a lack of effective and automatic methods for selection of the informative, high quality frames for the 3D spatial sensing of infrastructure using videogrammetry.

There is another major obstacle in applying monocular videogrammetry in the civil engineering domain. For monocular photo/videogrammetry, it is known that generated point clouds are up to an unknown scale. This is where someone has to know at least one dimension of an object in the scene to scale up the entire scene into real 3D dimensions (Ahmed, et al., 2010). In most cases, it is not easy to measure dimensions of real scenes due to it being time consuming and possibly inaccurate. Moreover, after extracting the length of one dimension in the scene, someone has to manually find and register it to the point cloud which is time consuming. In addition, merely considering one 2D dimension for registering a big 3D point cloud might result in less accurate outputs. Thus, it is mandatory to use specific methods for automating the procedure of computing absolute scale of scenes (Hartley & Zisserman, 2004).

Poor quality of obtained point clouds is another major issue in applying videogrammetry in the civil engineering domain because there might be a number of gaps and holes on surfaces of generated dense point clouds. This phenomenon is mainly caused from several reasons including insufficient number of frames, poor coverage of entire scene and texture less areas which is fairly common in civil engineering applications. Fortunately, considering color and geometric properties of point clouds, it is possible to improve the quality and fill up some of those gaps and holes (Nistér, 2004) (Pollefeys, et al., 2004).



Figure 1.2: Poor quality of a generated point cloud: existence of holes

Considering the aforementioned gaps in videogrammetric research for surveying of civil infrastructure, this research focuses on three specific problems to improve the performance of a

basic videogrammetric pipeline and to prepare for practical applications in the civil engineering domain:

1. Optimized selection of key video frames for processing.
2. Automated calculation of absolute scale of scenes using a single camera.
3. Automated point cloud data cleaning and filling the gaps located on different surfaces of point clouds using visual/geometrical properties.

To achieve this goal for each problem, an innovative solution is presented and validated.

It is necessary to mention that the focus of this research is mainly in two specific areas:

1. 3D as built documentation of rooms and offices as indoor applications.
2. 3D as built documentation of bridges and buildings as outdoor applications.

Here is a brief description of the condition for each category.

1.1. Indoor Applications

For indoor applications, lengths of captured video clips are not very long; thus, the selections of key frames are not a major issue. Moreover, it is possible to use simple 2D patterns such as printed letter size sheets to obtain absolute dimensions. In comparison with outdoor applications, results are more accurate due to controlled environments and short distances between the camera and objects. However, considering poorly texture surfaces like surfaces of walls and interior facades, obtained point clouds are not dense enough and sections of the point clouds might be missing (Pollefeys, 2008) (Fathi and Brilakis, 2011). In this research, different indoor environments have been reconstructed as case studies:

1. 3D reconstruction of typical home spaces like bedrooms and bathrooms.
2. 3D reconstruction of office spaces and corridors.

Sample snapshots of obtained results are illustrated in the following figure:

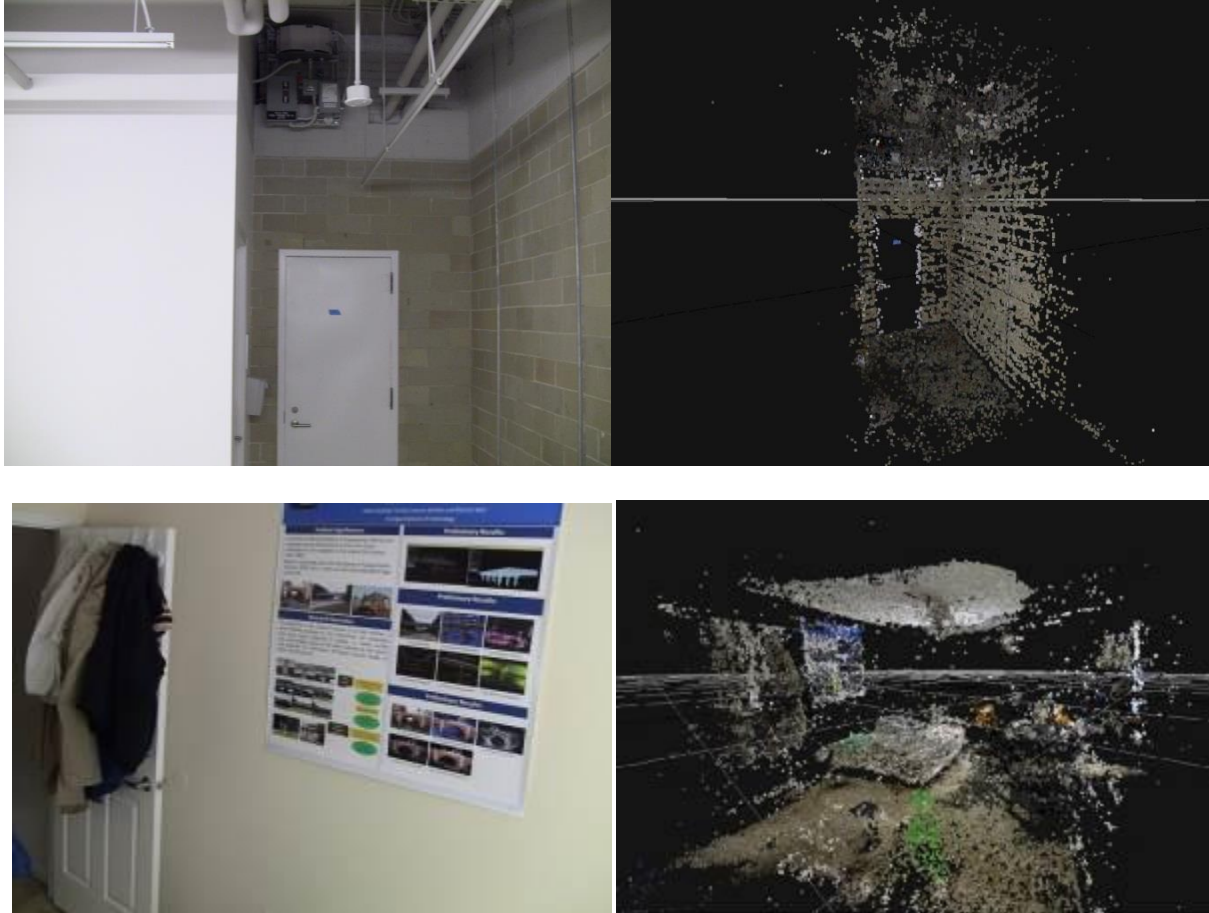


Figure 1.3: 3D point clouds of different indoor environments: a) office corridor including electrical and plumbing facilities, b) typical bedroom

1.2. Outdoor Applications

Based on the size of infrastructure scenes, captured videos are much longer than indoor cases. As a result, it is mandatory to extract a number of informative frames for processing. Also, considering practical purposes, it is not easy to use 2D patterns to compute absolute measurements. Distance between the camera and objects usually exceed 7-8 meters so obtaining a desired level of accuracy is a major challenge (Jian, et al., 2009). In this research, experiments of different building and bridge scenes are considered as the case studies:

1. 3D reconstruction of concrete column-beam bridges.
2. 3D reconstruction of arch bridges.

3. 3D reconstruction of various types of buildings including houses, campus buildings, sport facilities, etc.

Sample of obtained results is depicted in the following picture:



Figure 1.4: 3D point cloud of a concrete column beam bridge

As mentioned before, there are many other applications for generated point clouds using a videogrammetric pipeline which are out of scope of this research.

It is necessary to mention that the metrics utilized to measure the performances of different algorithms proposed in this research include accuracy and density of the generated point clouds. In this research, accuracy refers to the precision in 3D location of points in 3D space and density refers to the number of corresponding 3D points in a point cloud generated for each surface of the real scene. The rest of this dissertation is organized as following:

The current state of the art of practice/research for practical applications of monocular videogrammetry is reviewed in chapter 2. In chapter 3, an overview of the proposed methodology for solving each of the three problems is presented. As the next step, in chapters 4 and 5, experimental plans for testing and validating the proposed method are introduced. Finally, conclusions and future works are presented in chapter 6.

CHAPTER 2

ANALYSIS

2.1. State of Practice: Close-Range Spatial Sensing of Civil Infrastructure

The goal of close-range spatial sensing is to capture spatial/geometrical data from a distance with the help of spatial sensors. In infrastructure applications, close-range spatial sensors, such as laser scanners and vision-based software programs, are used to capture reality in the form of a point cloud. This cloud can be dense or sparse, depending on the sampling rate of the sensor. The sensed information is the x, y and z (depth) relative distance of each point from the sensor in the local coordinates system of the sensor. The depth information is either sensed directly by transmitting some sort of energy into a scene to receive reflected signal (Cho & Lee, 2009)(Shan, et al., 2008) (i.e. time-of-flight-based sensors) or indirectly by using natural light in the environment to capture surrounding information (i.e. vision-based technologies). These two established concepts in infrastructure spatial sensing are presented below.

2.1.1. Time-of-Flight-Based Technologies

Time-of-flight based sensors are typically laser scanners that operate with the time-of-flight principle, such as Light Detection and Ranging (LiDAR) (Fergus, et al., 2006). The operating principle is similar for all laser devices (Joshi, et al., 2010). A pulse of laser light is emitted to probe the object in the scene, and a time-of-flight laser range finder at the heart of this type of sensor finds the distance of a surface by timing the round-trip time of a pulse of light. The amount of time before the reflected light is detected is then calculated. Since the speed of light is known, the round-trip time determines the travel distance of the light, which is twice the distance between the scanner and the surface. The laser range finder only detects the distance of one point in its direction of view at a moment (Marziliano, et al., 2002).

2.1.2. Computer Vision-Based Technologies

The use of vision-based remote sensing technologies for 3D reconstruction of built environments has been the subject of many research initiatives both in computer vision and civil infrastructure management applications (Chung, et al., 2004) (Varadarajan & Karam, 2008) (Chen & Bovik, 2009). These technologies could provide an inexpensive solution for infrastructure's spatial data collection problem since they only require off-the-shelf digital cameras. In terms of the labor costs, they require minimal expertise and therefore an average construction employee could easily collect the required data. The recent advances in automating the computer vision algorithms allows the data processing step to be automatic which removes the need for manual post-processing of the collected data.

Other than the AEC industry, videogrammetry is a well-known technology for extracting measurements from different objects mainly in manufacturing industries and through a process called reverse engineering (Rashidi, et al., 2013).

For close-range (<2m) 3D reconstruction, it has been shown that vision-based approaches can produce models where the spatial accuracy rivals laser scanning (Crete, et al., 2007). Also, there have been several successful studies in the last couple of years that show how vision-based approaches can potentially be used for reconstruction of large scale environments (Seo, et al., 2003) (Torr, et al., 1998) (Seo, et al., 2008) (Thormahlen, et al., 2004). However, their performance for the range values and metric accuracy that are required in infrastructure applications has not been evaluated.

In general, vision-based technologies could be categorized into photogrammetry and videogrammetry. These two approaches are further detailed in the following subsections.

- **Photogrammetry**

Photogrammetry is the process of measuring geometric properties of real world objects from digital images. It requires a two-step procedure to provide spatial information. First, the engineer has to shoot the right source photos as input data. After collecting all of the photos, 3D

object point coordinates are calculated by employing multi-view stereo matching algorithms and Structure from Motion (SfM) techniques. SfM refers to the process of finding 3D structure of an object by analyzing local motions of the camera (Rashidi, et al., 2011). Camera parameters and 3D location of the matched features across multiple images are then recovered by minimizing the reprojection error. At least two images from different views of a point are needed for this purpose. These coordinates are calculated via triangulation, using the camera internal parameters and the corresponding image point pairs. Aside from spatial data acquisition, additional information such as object texture and color may be extracted to construct photo-realistic 3D models (Sheikh, et al., 2005).

Over the last decade, a number of commercial software packages have been developed based on photogrammetric principles. Photofly, developed by Autodesk Company, and Photosynth, presented by Microsoft, are two famous examples of software in this category.

- **Videogrammetry**

Videogrammetry is the process of determining 3D coordinates of object points using one or more video streams taken from different angles. The processing pipeline is typically similar to that of photogrammetry. However, in practical applications in the civil engineering domain, it is more efficient to videotape the scene rather than taking several pictures since (Zhang, 1999):

- a. An unorganized set of images that are used in image-based 3D reconstruction creates an exponential relationship between run-time and the number of images. The reason is that all the input images should be analyzed in order to find specific data such as a distinct object point. However, videogrammetry benefits from the sequential characteristics of video data which removes the necessity of this extensive computational load.
- b. In order to process images/video frames and generate a dense point cloud, there should be sufficient overlap between images/video frames. In videogrammetry, this goal would be automatically achieved and the user would only need to turn on the camcorder/camera, traverse around the scene and videotape it from all possible angles. In photogrammetry,

there should be a specific pattern for taking pictures and the user should receive specific instructions on how to take proper photographs from different views. Failure to meet such requirements in taking pictures might result in the failure of processing data.

On the other hand, videogrammetry suffers from a number of issues that needs to be carefully considered by the researchers in this domain (Lowe, 2004):

1. Quality and resolution of video frames are lower than images.
2. Even a short video clip contains millions of frames. For processing purposes, only a few portions of these frames are required.

2.2. Comparison of Image- and Time-of-Flight- Based Technologies for 3D as Built Documentation of Indoor/ Outdoor Scenes

In order to evaluate the overall performance of the mentioned technologies in the civil engineering domain, a number of experiments have been conducted. The comparison experiments divided into two categories, e.g. indoor and outdoor, and followed the procedures set forth in Figure 2.1.

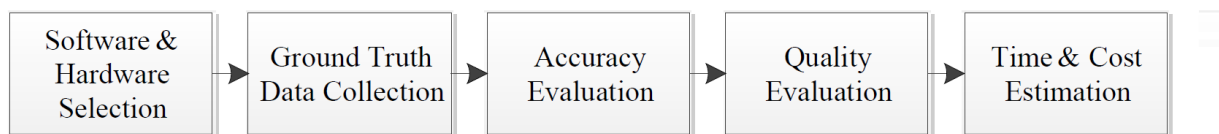


Figure 2.1: Overview of the comparison workflow

First, cutting-edge image-based 3D reconstruction software was selected. This representative software was used to evaluate different technologies in terms of accuracy, quality, time efficiency and cost. To this end, two bridges and two indoor scenes were selected, and a total station was used to collect the spatial coordinates of feature points (e.g., corners) on the surface of the infrastructure and indoor objects. Since total stations survey infrastructures only with an error of 2-3 mm (standard deviation) within 100 m, data collected by total stations can be deemed accurate enough as ground truth. Based on the feature points, surface models of these outdoor and two indoor scenes were created for the evaluation of the selected methods. Different

camera models, resolution configurations, and data collection distances will lead to different levels of accuracy and quality for the photo/videogrammetric results, while laser scanners might be affected by temperature and atmospheric pressure. Therefore, specific combinations of these variables were prepared to produce different data sets for testing using photo/videogrammetry and laser scanning respectively. By separately registering each data set onto the ground truth model, the accuracy in terms of the distances between the testing points and their actual ground truth model surfaces, and the quality in regard to the completeness of the testing point set can be consequently measured. For both technologies, the time for raw data collection in the field and processing time at office will be recorded and cost will be estimated in terms of dollar value. Finally, the measured results will be summarized as actionable information (i.e., performance/cost measure) to help engineers make cost-effective decisions such that not only the accuracy and quality required for their applications are met, but also the time and money spent are minimized.

2.2.1. Software and Hardware Selection

Considering public availability as the most significant selection factor, the following photogrammetric software has been chosen to use in the comparison studies:

1. Bundler (Snavely 2010) + PMVS (Patch-based Multi-view Stereo) (Snavely, et al., 2008; Rashidi, et al., 2011)
2. Photo Fly (Autodesk 2011)
3. PhotoSynth (Microsoft 2011)

In order to evaluate videogrammetry, a basic monocular videogrammetric pipeline capable of generating dense point clouds was developed. To scan the scenes with laser scanners, a Leica Scan Station C10, the latest product of the Leica Company as the pioneer in the field, was utilized (Nistér, 2004).

2.2.2. Ground Truth Data Collection

A total station (i.e., SOKKIA 30R) was used to collect data for the selected indoor/outdoor case studies. The accuracy level of a total station is around 1-2 mm at the range of 100 m, which will allow the collected data to be used as the ground truth. Corners and feature points located on different surfaces of the scenes' objects were collected in terms of 3D coordinates. As the next step, collected 3D points were rendered into a water-tight surface model using the Poisson Surface Reconstruction algorithm (Lourakis & Argyros, 2009), which was ready for the evaluation of photo/videogrammetry and laser scanner technologies.

2.2.3. Evaluation of Accuracy

The comparison variables considered will be cameras models, resolutions, distances, and temperatures. Selection of camera models will basically follow the criterion that cameras are off-the-shelf or from professional categories. Settings of resolutions will be configured based on the available range of a camera (e.g., for Canon Vixia HF S100, 1~5 megapixels). The distance ranges will be defined from 5~50 m, which is reasonable for infrastructure scenes. To evaluate the accuracy, each set of point clouds produced by the two technologies will be registered into the coordinate frame of the ground truth model using the Horn's absolute orientation method (Scharstein & Szeliski, 2002). Within the same frame, Euclidian distance (error) between a point from the point cloud and the surface of the ground truth model where this point is supposed to be located is measured. This distance is considered as the metric to measure the accuracy (illustrated in Figure 2.2), grounded on that for image-based methods, inaccurate radial lens distortion parameters have systematic effects (error) primarily along the depth direction (Furukawa, et al., 2010) (Hansen, et al., 1953). The i^{th} point's coordinate is denoted as (X_i^j, Y_i^j, Z_i^j) . It is supposed to lie on the j^{th} surface of the ground truth bridge model, as $a_j X + b_j Y + c_j Z + d_j = 0$. The average error of the point cloud can be calculated by:

$$err = \frac{1}{\sum_{j=1}^n m_j} \sum_{j=1}^n \sum_{i=1}^{m_j} \frac{|a_j X_i^j + b_j Y_i^j + c_j Z_i^j + d_j|}{\sqrt{a_j^2 + b_j^2 + c_j^2}} \quad (2-1)$$

In equation (2-1), m_j is the number of points supposed to belong to the j th surface, and n is the number of the surfaces. Note, if a point's distance to the surface is far beyond the average value, it will be deemed as an outlier and removed from the testing data set.

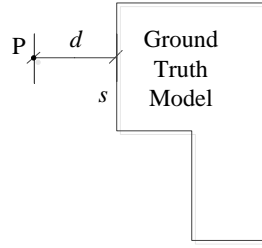
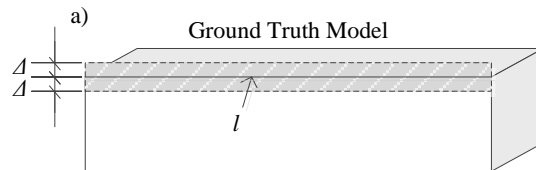


Figure 2.2: The point P is supposed to be located on the surface s of the ground truth model, but due to imperfection of the point cloud, there is a distance d between P and s . d is deemed as error of this point

2.2.4. Evaluation of Quality

The quality of a point cloud will be examined with respect to its completeness. To achieve this, the surfaces of the ground truth models were partitioned into small square regions (1×1 inch in this study). The existence (or not) and density of points were examined for each one of these regions. The outcome was depicted as the percentage of the coverage or completeness ratio. The percentage of the coverage is calculated by m/n , where m is the number of regions covered by any points and n is the total number of regions. The density is measured by $\frac{\sum_{i=1}^n x_i}{n}$, where x_i is the number of points covered by the i th region and n is the total number of regions.



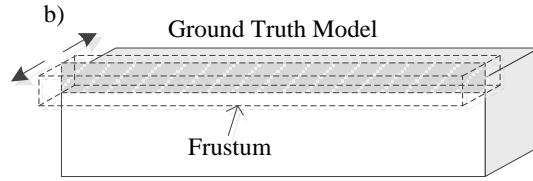


Figure 2.3: a) an offset Δ (exaggerated for illustration) is made on both sides of the edge l to form a region marked with dash lines, and b) a frustum is generated by extruding the region along the axis orthogonal to the surface

2.2.5. Time and Cost Estimation

The manpower time for collecting on-site data and post-processing the data at office will be recorded. These times will then be measured in terms of labor cost for the image-based and time-of-flight-based technology. Additionally, costs of actually purchasing or renting the devices will be considered (e.g., a typical inter-state bridge laser scanning by rental is \$4000 including 4~5 hour on-site data collection and 1~2 hour office data processing).

2.2.6. Experiment and Results

This research has conducted experiments to compare the performance of image and time-of-flight methods based on the metrics discussed above. As mentioned before, the test-bed cases include two indoor facilities (a bedroom and an office corridor) and two highway bridges. A calibrated consumer-grade Canon Vixia HF S100 camera was used to capture the image data of these objects. This Canon camera combined with an industrial Point Grey Flea-2 camera, was also used to collect the video streams. The software running is benchmarked on Intel Core i7 CPU with 4GBs RAM on a 32bit Windows platform, except for Photo Fly and PhotoSynth, which processes the images from the hosts' own remote servers.

Case 1: A concrete column beam bridge

For this set of tests, experiments have been conducted in two steps. First, the tests compared the image and time-of-flight methods at the data collection range of 25 m. Second, the tests evaluated the accuracy and quality of videogrammetric methods based on the metrics of: (1) type of cameras, (2) resolution configurations, and (3) data collection ranges. The interstate concrete beam bridge is four-spans and three rows with each row containing three rectangular

columns (Figure 2.4a). The resolutions of the cameras were set at 2 MP. The reasons for choosing this particular resolution will be explained in the validation section. A set of 130 images and two lengths of 5 minute videos were collected for comparison. As for the laser scan data collection, six scans were made on a sunny day in April 2011 with temperatures measuring around 30~35°C. Two of the scans were conducted under the bridge and the remaining four were from both sides at distance around 25 m. These raw images, videos, and laser scan data were processed using the representative software respectively and the results are 3D point clouds of the bridge as shown in Figure 2.4.

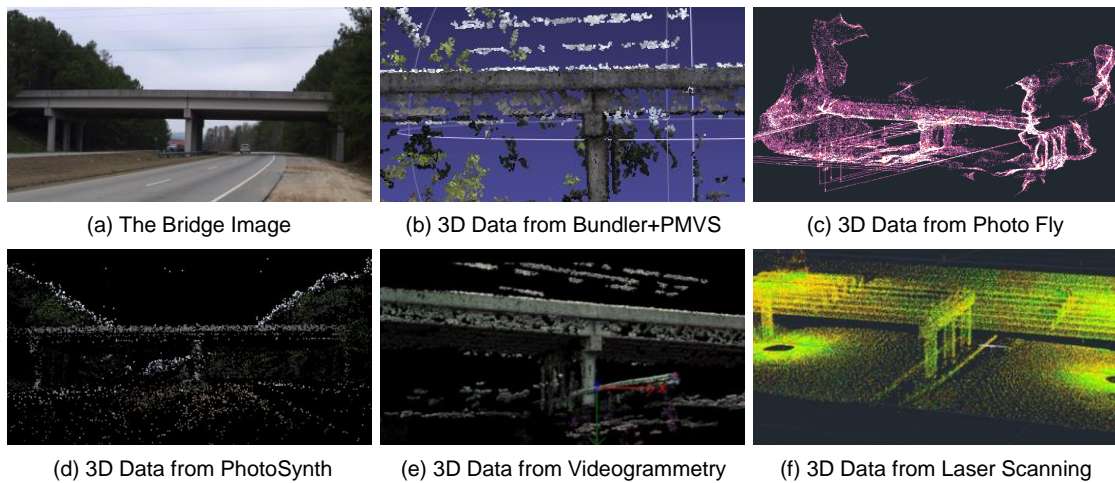


Figure 2.4: Snapshots of 3D point clouds of the beam bridge

The ground truth model (Figure 2.5a) was constructed from around 2000 points spreading the surface of the bridge captured by the total station. By manually picking the reference points on the bridge, the point cloud by each method was thereby registered into the coordinate frame where the ground truth model is located. Figure 2.5 shows a snapshot of a registered point cloud on the ground truth model of the bridge.

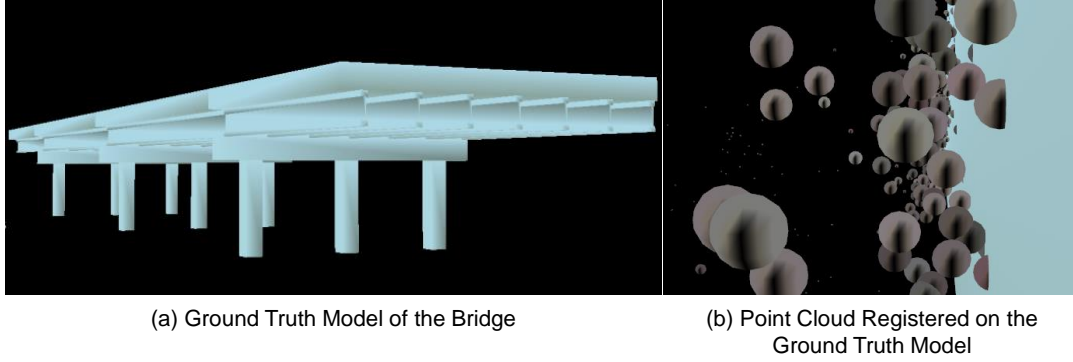


Figure 2.5: Point cloud registration

Table 2.1 shows the accuracy and completeness of different image and time-of-flight-based methods. Based on Table 1, the accuracy level of laser scanning is about 10 times as high as that of photo/videogrammetry with the completeness rate and density level being 20% and 75% higher. This research has shown that the density level of photo/videogrammetry is low due in large part to scarcity of features presented on the surfaces of this bridge. The difference between edge errors and overall errors is not significant (by 1.5%) for the image-based methods. The completeness rate of PhotoSynth is particularly low because the point cloud generated by this method is sparse. Also, observation showed that videogrammetry obtained highest accuracy and Bundler + PMVS was the second by less than 9% among the image-based methods. It shows the reason is intrinsic and lens distortion parameters of the camera are utilized for computation of their results. The difference is that the parameters used in videogrammetry were attained from calibration whereas Bundler + PMVS read the EXIF (exchangeable image file format) information of the captured photos to obtain the same parameters.

Table 2.1: Estimated camera calibration parameters in different experiments

Method	Avg. Error (cm)			Avg. Edge Error (cm)	Avg. Completeness (%)	Density (pt/m ²)
	Value	Limits at 95%				
Bundler+PMVS	7.21	6.44	7.98	7.56	73	8312
PhotoSynth	11.39	10.74	12.04	11.81	27	3511
Photo Fly	13.74	13.42	14.06	13.43	62	7617
Canon Vixia	6.60	6.19	7.01	6.61	79	8539
Point Grey	6.33	6.01	6.65	6.42	78	8350
Leica C10	0.52	0.48	0.56	0.64	98	14994

Table 2.2 shows the time and cost spent for collecting spatial data by different methods. It is obvious that image-based methods are time efficient for the on-site data collection. As for processing the data at the office, the time was counted based on these particular data sets provided. Admittedly, as the image number increases, the time for both collecting on-site data and processing videos and images at office will increase. Yet, it is still plausible for the image-based methods observing the trend that the total cost will not significantly increase since it is mainly accounted for by the equipment costs. Note that the cost calculated here is on one application base and it is the overall budget for the first time investment.

Table 2.2: Time and cost estimation of different methods

Method	Time (hr)			Cost (\$)		
	On Site	Office	Total	Equip.	Soft.	Total
Bundler+PMVS	1.0	5.0	6.0	1,000	Free	1,120
PhotoSynth	1.0	0.1	1.1	1,000	Free	1,022
Photo Fly	1.0	4.0	5.0	1,000	Free	1,100
Canon Vixia	0.8	3.0	3.8	1,000	Free	1,076
Point Grey	0.8	3.0	3.8	2,500	Free	2,576
Leica C10	5.0	2.0	7.0	100,000	w/ Equip.	100,140
Leica C10	5.0	2.0	7.0	by rent	w/ Equip.	4,000

Note: Manpower unit price is assumed as \$20/hr.

Later, the videogrammetric methods were evaluated in terms of data accuracy, completeness, and data density under varied settings. In Table 2.3, the average error shows anti-correlation with the level of resolution, i.e. at a certain distance, the higher the resolution, the lower the average error. As expected, there is strong correlation between the average completeness and the level of resolution. When the camera resolution is fixed, we observe the trend that the average error increases while the average completeness decreases with the increase of distance. Data density shows a similar trend as the average completeness. As for the performance of different cameras, under the same level of resolution 2.0 MP, the average error is reduced slightly by 6.3% using photos taken by Point Grey and the average completeness seems

to increase correspondingly except for the measurement taken at distance of 25m. Nevertheless, the difference is subtle.

Table 2.3: Accuracy, completeness, and density of videogrammetry under different settings

Camera	Resolution (MP)	Distance (m)	Avg. Error (cm)	Avg. Edge Error (cm)	Avg. Completeness (%)	Density (pt/m ²)
Canon Vixia	1.0	25	7.43	8.31	75	7531
Canon Vixia	1.5	25	6.92	8.07	80	7940
Canon Vixia	2.0	25	6.60	6.61	79	8539
Point Grey	2.0	25	6.33	6.42	78	8350
Canon Vixia	1.0	40	10.11	12.79	73	7324
Canon Vixia	1.5	40	9.75	12.46	76	7937
Canon Vixia	2.0	40	9.13	12.92	77	7953
Point Grey	2.0	40	8.47	10.65	80	8453
Canon Vixia	1.0	55	11.01	13.09	71	7233
Canon Vixia	1.5	55	10.27	12.37	74	7844
Canon Vixia	2.0	55	10.51	11.02	77	8038
Point Grey	2.0	55	9.72	10.1	79	7744

- **Case 2: A concrete arch bridge**

In this case, the tests compared the image-based methods using a concrete arch bridge as a test-bed (Figure 2.6) and evaluated the analytics derived in the previous section. The camera resolution was set to 2 MB and 130 images and a 5 minute video was collected within the range of 5~40 m. The 3D point clouds were generated and illustrated in figure 2.6 (b-e).

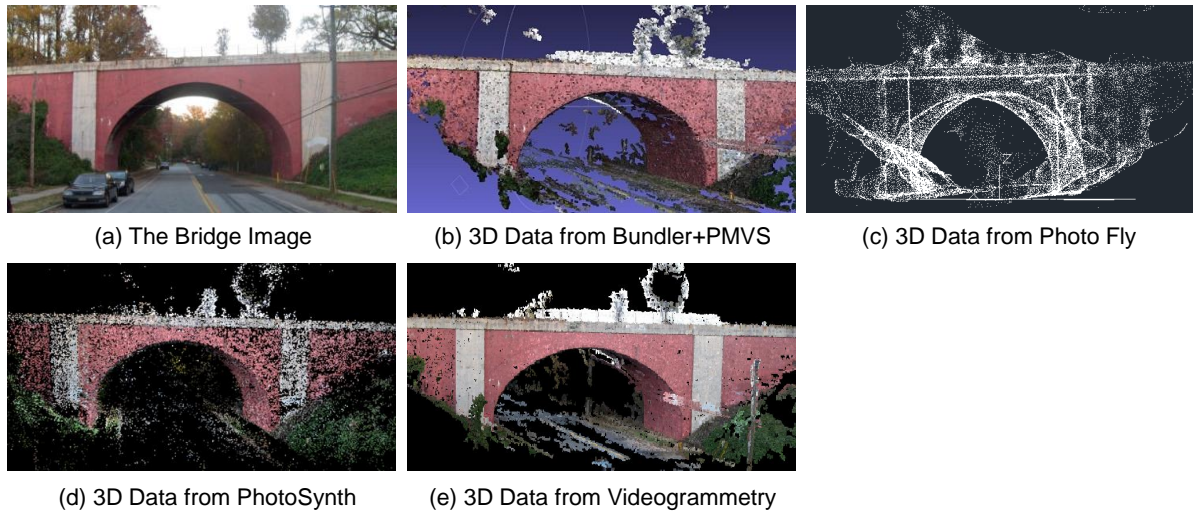


Figure 2.6: Snapshots of 3D point clouds of the arch bridge

Table 2.4 shows the comparison results of the four image-based methods with respect to accuracy, completeness, and density. Videogrammetry and photogrammetry - Bundler + PMVS based results are ranked the highest in accuracy, in consistency with the results in Case 1. It is noted that the completeness rate and density level are on average 15% and 13% higher than those in Case 2, because of a better coverage of the bridge in images and fewer repetitive patterns revealed on its surface. The better coverage of the bridge in images is likely due to a wide range of distances in data collection. While long range is able to cover a comprehensive view of the content, close range can grasp as many details of the bridge as possible.

Table 2.4: Accuracy, completeness, and density of different methods (arch bridge case study)

Method	Avg. Error (cm)			Avg. Edge Error (cm)	Avg. Completeness (%)	Density (pt/m ²)
	Value	Limits at 95%				
Bundler+PMVS	7.18	6.83	7.53	7.31	73	9543
PhotoSynth	10.02	9.58	10.46	10.25	27	3226
Photo Fly	7.67	7.00	8.34	7.86	72	7460
Canon Vixia	6.59	6.24	6.94	6.55	79	10062
Leica C10	0.63	0.59	0.67	0.7	97	11206

- **Case 3: indoor environments**

For each indoor case study, a number of significant 3D points such as the corners of a table and edges of a book, were collected by the total station. The entire scene was videotaped and different point clouds were generated using different software and following the same procedure as discussed before. Performance results of different methods are summarized in Table 2.5:

Table 2.5: Accuracy, completeness, and density of different methods (indoor case studies)

Method	Avg. Error (cm)			Avg. Completeness (%)
	Value	Limits at 95%		
Bundler+PMVS	1.45	1.41	1.68	65
PhotoSynth	2.26	2.18	2.45	22
Photo Fly	2.36	2.2	2.55	61
Canon Vixia	1.25	0.99	1.33	66

2.3. State of Research: Applications of Monocular Videogrammetry in Construction Area

Within the last two decades, researchers in the area of construction engineering have heavily focused on various applications of both active and passive sensors. A brief review of the current research efforts in this particular area is provided below.

- **Applications of active sensors:**

Laser scanners are the most popular type of active sensors and are commonly used for different applications in construction jobsites. In terms of construction research, the focus has been mainly on integrating laser scanning and BIM modeling of buildings and infrastructures (Tang, et al., 2010). From technical perspective, researchers have tried to solve specific problems related to generated results. These problems include lower quality of scanned data close to sharp edges (Tang, et al., 2010) and automated detection of objects from point cloud data (Cho & Gai, 2014). A number of researchers have also investigated potential applications of 3D range cameras in construction areas (Teizer, et al., 2008).

- **Applications of passive sensors:**

In terms of active sensors, the focus has been on different areas including progress monitoring of construction job sites (Golparvar-Fard, et al., 2009), as-built documentation of civil infrastructure (Dai, et. al, 2013) (Brilakis, et al., 2011), and motion tracking of construction resources (Teizer & Vela, 2009).

(Golparvar-Fard, et al., 2009), used the point-based approach on un-calibrated daily progress photographs of construction sites for sparse 3D reconstruction of the jobsite and construction progress monitoring. The main limitation of this study is that it only provides 3D coordinates of feature points. Edge points, such as points on the perimeter of a roof, are deleted in the feature point detection process (Lowe, 2004) and do not exist in the output. Dense point cloud generation algorithms such as Furukawa and Ponce (2010) could be used to overcome this limitation. The generation of a dense as-built point cloud for a construction site has been considered in (Golparvar-Fard, 2013). The study presents an automated approach for progress monitoring of activities at a jobsite based on two sources of information: a dense point cloud and a BIM file. The dense as-built point cloud is generated from unorganized daily construction photo collections.

In particular, several research undertakings have taken place in creating videogrammetric pipelines to reconstruct 3D buildings and infrastructure. The authors in (Brilakis, et al., 2011), created a framework that exploits a binocular stereo camera system to progressively reconstruct an ongoing building structure. (Fathi & Brilakis, 2011) proposed a method to generate the sparse point cloud of the scene from stereo images. These efforts are the basic steps toward applying videogrammetry in 3D spatial sensing of civil infrastructure. Considering the practical constraints encountered on site (i.e., low frame quality, video frame redundancy, generating up to scale and low quality point clouds), a robust videogrammetric method is needed. As mentioned before, the focus of this research is on three specific problems related to practical applications of videogrammetry in the civil engineering domain:

1. Automated selection of a number of high quality, informative key video frames for processing.

2. Automating the process of computing absolute scale of the scene.
3. Point cloud data cleaning and filling the gaps and holes existing on different surfaces of generated point clouds. State of research for each problem is presented separately within the following sections.

2.3.1. Strategies for Quality Control of Video Frames

One challenge in processing video frames is dealing with the low quality frames which are primarily affected by motion blurs. The motion blur is caused by camera shakes. Even a slight shaking may lead to serious blur effects. Normally, there are two strategies in dealing with this problem.

One of them is image de-blurring, which has recently received much attention among the computer vision community. De-blurring techniques can be divided into two main categories:

1. **Hardware methods:** The most common commercial method for reducing image blur is image stabilization (IS). These methods, used in high-end lenses and now appearing in lower-end point and shoot cameras, use mechanical means to dampen camera motion by offsetting lens elements or translating the sensor. Fundamentally, IS tries to dampen motion by assuming that the past motion predicts the future motion. However, it does not counteract the actual camera motion during an exposure nor does it actively remove blur – it only reduces blur.
2. **Software methods:** For software methods, different algorithms are utilized to model motion blur. If a motion blur is shift-invariant, it can be modeled as the convolution of a latent image with a motion blur kernel where the kernel describes the trace of a sensor. Then removing a motion blur from an image becomes a de-convolution operation. In non-blind de-convolution, the motion blur kernel is given and the problem is to recover the latent image from a blurry version using the kernel. In blind de-convolution, the kernel is unknown and the recovery of the latent image becomes more challenging. Several methods were proposed to handle the motion blur problem. While these methods might produce excellent de-blurring results, they necessitate intensive computation. It usually

takes more than a couple of minutes for the methods to de-blur even a single image of moderate size, making this method not applicable for high volume applications.

The aforementioned hardware and software methods reflect the state-of-the-art of current video processing techniques. However, these methods are not feasible for use in infrastructure reconstruction applications considering their computational complexity. Moreover, reconstructing infrastructure scenes only needs to process a few frames rather than process every frame contained in a video stream.

Another strategy is by removing blurred frames. Removing these frames relies on applying blur metrics that are capable of measuring numerical values that represent the extent of the effect. To measure the blur effect, the intensity change along edges was extensively studied by (Marziliano, et al., 2002) (Chung, et al., 2004) (Varadarajan & Karam, 2008). These algorithms are easy to use for predicting sharpness/blurriness of an image; however, evaluation of these algorithms are mainly focused on widths of intensity edges which does not completely reflect the quality of the measured images. To address this limitation, (Chen & Bovik, 2009) established a multi-resolution decomposition method capable of extracting reliable features regarding image blurs. The advantage of this method is that all the pixels of an image are taken into account and the estimate is more reliable. However, each time an image is assessed; a training process of sample images is needed, imposing extra work to the modeler.

Apart from the above methods, the BluM metric method (Cete, et al., 2003) is based on the human blur perception. It uses the blur discrimination properties of the human perception to create a reference image for the blur estimation, making this method simple yet practical. Here the blur discrimination properties means that human visual system is capable of differentiating the blurred and sharp image but cannot accurately discern a blurred image with the one that is re-blurred. By this method, the complete scene in the image is considered, and the resulting scores are simply denoted by the numerical range from 0 to 1 representing the worst blur and the best sharpness of the image respectively. Therefore, the BluM metric is the most suitable candidate to

measure the blur effect of video of an infrastructure. The end result of this step is selected high quality frames from a video stream.

2.3.2. Strategies for Key Frame Selection

Selecting a number of representative key frames from a large video sequence is a main step for efficient and robust 3D reconstruction. There are several undertakings in the research and development of video key frame selection criteria. For these undertakings, the criteria that have been used are primarily focused on four aspects: (1) long enough baseline, (2) sufficient overlap, (3) degeneracy avoidance, and (4) minimized reprojection errors. (Seo, et al., 2003) considered three criteria to extract key frames: (a) the ratio of the number of correspondences to the number of features, (b) the homography error, and (c) the distribution of correspondences over the frames. In their method, the criterion (a) is used to ensure the sufficient overlap of two frames, the criterion (b) serves as a good proxy for the baseline distances between two views, and the criterion (c) is used to increase the accuracy of calculating the fundamental matrices based on which to lead to higher accurate estimations of camera motions and an object structure. Later on, (Seo, et al., 2008) improved their method to extract informative key frames by incorporating a fourth criterion that the reprojection error of the reconstruction process is minimized. Nonetheless, these methods did not consider the degeneracy cases. In (Pollefeys, et al., 2004), the degeneracy problem is addressed by employing the Geometric Robust Information Criterion (GRIC). As a result, the next key frame is selected only once the fundamental matrix model explains the relationship between the pair of images better than the homography matrix model through the scores calculated by the GRIC.

In addition, (Ahmed, et al., 2009) proposed another key frame selection method which is based on the weighted score considering the GRIC difference and point to epipolar line cost. In their method, they also use the corresponding ratio to indicate whether long baseline and sufficient overlap between two frames exist. (Gibson, et al., 2002) proposed a method in which the sum of three weighted addends of: (1) the fraction of features that were matched in the

previous frame pair which cannot be matched in the current pair, (2) the inverse of the square of the homography error, and (3) the squared median epipolar error is minimized to select the frames. The long enough baseline, sufficient overlap, and degeneracy avoidance are ensured in their method. (Thormahlen, et al., 2004) proposed a criterion for the selection of key frames with the lowest expected estimation error of initial camera motion and object structure. In the same time, their method utilized the GRIC to guarantee a sufficient baseline, overlap, and avoid degeneracy cases.

In summary, the above discussed methods that have been developed are able to guarantee a long enough baseline, sufficient overlap, no degeneracy cases, and minimized reprojection errors of 3D reconstruction. However, these research efforts have not taken into account an optimized number of required frames for processing. The practical constraints such as speed of camera movements, and complexity of the scene also significantly impact the key frame selection algorithms, whereas they have not been addressed by the existing research.

2.3.3. Strategies for computing the absolute scale of dense point clouds

According to results of current studies conducted by (Golparvar-Fard, et al., 2013) and (Becerik-Gerber, et al., 2013), monitoring the health of infrastructure is one of the grand challenges faced by civil engineers in the 21st century. Lack of viable methods to map and label existing built infrastructure is an important component of this challenge. As-built 3D geometry comprises a significant portion of the total as-built information and any efforts towards automating its acquisition will translate to cost savings and improved quality assurance in the delivery and maintenance of the built environment.

The current state-of-the-art approach to collecting spatial data and converting it to as-built geometry of built environment scenes is through active sensors (total stations and laser scanners) and surveying methods. This approach encapsulates the 3D geometry in a set/cloud of 3D points. Although as-built geometry generation is assisted by recent technological advancements both in hardware and software, most of its steps are costly in terms of equipment

and labor, and time consuming. As a result, there is increasing demand for automated, cost effective methods for collecting spatial data of built infrastructure scenes and converting the data to as-built models (Brilakis, et al., 2011).

Within the last two decades, advances in high resolution digital photography and increased computing capacity have made it possible for image/video-based 3D reconstruction methods to produce promising results. Over the past few years, researchers in the fields of computer vision and civil engineering heavily focused on developing algorithms to improve the performance of this technology.

Based on the number of cameras, photo/videogrammetric-based algorithms are divided into two major categories: (1) monocular, defined as using a single camera; and (2) binocular, defined as using a stereo set of cameras. Additional cameras can also be used if needed in multi-camera systems. For binocular, the relative position and orientation of one camera in relation to the other camera is measured in advance and considered as a known parameter; thus, making it directly possible to obtain 3D measurements in Euclidian space. However, stereo cameras are specialized equipment and far less feasible hardware solutions than monocular setups such as the cameras in most smart phones that on-site personnel carry. In general, a single camera (monocular setting) is a much more practical way to capture images/video data since most individuals on a jobsite has access to a single digital camera or smart phone. However, implementing a monocular camera setup only generates unknown global scale PCD (Scaramuzza, et al., 2009). In order to compute the absolute scale, the operator needs to know the base line of the camera motion or at least one dimension of the scene. The traditional way of solving the problem is measuring the distances between a set of predominant points in the scene before or after the data collection. The corresponding 3D locations of these predominant points should be manually identified by the operator from the generated PCD. The ratio of the real Euclidian distance between the predominant points compared to the computed distance in the PCD is the absolute scale of the scene.

Measuring such dimensions in a job site is a manual task that increases the time and effort needed to collect the geometry and increases human error in one of the most sensitive parts of the 3D reconstruction process; consequently, the results can be inaccurate. In addition, there is no guarantee that the corresponding measured points is successfully reconstructed and already exist in the PCD. As explained in section 6, this research conducted experiments and measured a number of dimensions in outdoor built environments using a total station. These experiments indicate that it takes an average of 15 minutes to manually measure one dimension of the scene, find the corresponding points in PCD, and calculate the scale factor within a reasonable error tolerance.

Several new methods have been proposed for automatically retrieving the absolute scale of a scene using a monocular setup. These methods, however, either lose the practicality of the monocular setup by adding extra sensors or are limited to explicit scenes and are not general enough to be useful by Architecture/Engineering/Construction (A/E/C) practitioners in their daily tasks (Scaramuzza, et al., 2009). A general method for automatically computing the absolute scale of PCD from monocular video without the use of additional sensors is proposed in this research. The proposed method is based on using pre-measured simple standardized objects that are commonly available or easily obtained, in particular a letter-size sheet of paper for indoor settings (up to approximately 7 meters distance from target) and a simple colored cube made of plywood material for outdoor environments (up to 25 meters distance from target). The vertices of these predefined objects are detected in video frames using a novel algorithm. The detected vertices in 2D frames are then reconstructed along with the other feature points extracted from the scene. Knowing the distance between the vertices, the entire PCD is then scaled up using an existing method.

In manufacturing practices, the entire measurement procedure takes place in indoor, controlled settings so it is feasible to arrange specific settings for directly extracting real dimensions of objects. One popular approach is using specific target projectors called PRO-SPOT. This structured-light system works like an ordinary slide projector. A light source

illuminates a target slide. As the next step, the illuminated pattern (usually a dot pattern) passes through a number of lenses which magnify the slide and project it onto the object's surface. By knowing the dimensions of the pattern, it is possible to extract the actual dimensions of the objects (Figure 2.7).



Figure 2.7: Projector and camera setup for extracting absolute measurements in manufacturing industry (Ganci and Brown, 2008)

The proposed solution is feasible for indoor, controlled manufacturing environments; however, it does not practically fit the random, uncontrolled built infrastructure scenes. Theoretically, for built infrastructure scenes, it is possible to compute the global scale of the PCD by measuring only one dimension in the scene. However, in practice, a number of issues would occur:

1. The common practice to precisely measuring dimensions in a built infrastructure jobsite is using a total station (Coaker, 2009). Using total stations for measurement purposes leads to very accurate results (average error $\approx \pm 1$ mm). However, the entire procedure is not straight forward and requires certain levels of training. A surveyor should carefully setup the equipment in a proper location of the job site and conduct the measurements (Coaker, 2009). The surveyor then goes back to the office and implements relevant software for post processing steps including visualization of PCD, extracting corresponding measured dimensions

from it, and scaling up the entire PCD. Obviously this procedure is time consuming and labor-intensive.

2. Unlike scanning senses using laser scanners, in some cases processing images and video frames does not result in generating PCD that are uniformly dense enough (Rashidi, et al., 2013). There might be poorly reconstructed areas due to several reasons such as insufficient coverage during sensing, reconstruction errors and texture-less areas, and there is no guarantee that the corresponding points used for actual measurements already exist in the PCD.
3. The devices used for measuring dimensions of the scene are either expensive, e.g. laser measurer and total stations, or inaccurate, e.g. tape measurer (Dai, et al., 2013).

For monocular camera settings, two major approaches are suggested to automatically recovering the absolute scale.

The first approach relies on the application of supplemental electronic sensors for acquiring extra information about the scene or motion of the camera. Global Positioning System (GPS), inertial measurement units (accelerometers, gyroscopes, magnetometers), and odometry measurements are examples of the applied sensors for providing supplemental measurements for absolute scale computation purposes (Tribou, 2009). (Nutzi, et al., 2011) fused inertial measurement unit (IMU) and visual data for absolute scale estimation in monocular SLAM (Simultaneous Localization and Mapping). (Eudes et al., 2010) solved the scale drift problem observed in long monocular video sequence using a standard odometer installed on a car. (Kneip, et al., 2011) combined accelerometer and attitude measurements with feature observations in order to compute the metric velocity estimation of a single camera. Supplemental sensors can also be applied in the form of range measurement devices or additional monocular cameras (Gutierrez-Gomez & Guerrero, 2012). (Jung et al., 2008) implemented a range finding device for use in a SLAM context by projecting a structured light on the environment and measuring the resulting distortions with a monocular camera. 2D laser range finder (LRF) is another popular

sensor used by the robotics and computer vision community to address the global scale issue (Castellanos, et al., 2000).

Applying additional sensors is not always a cost effective solution, so other researchers have tried to use prior knowledge about the scene obtained through predefined existing objects and visual fiducials (Tribou, 2009). In the SLAM area, different classes of objects and artificial landmarks are utilized to acquire necessary information about the environment and therefore, solve the robot positioning or localization problem. (Olson, 2011) proposed a visual fiducially system based on 2D planar targets with specific bar code patterns for accurate localization of robots. Obtained results for localizing groups of robots in indoor and outdoor settings have been promising. (Botterill, et al., 2012) proposed an innovative solution to the problem of scale drift in single camera SLAM based on recognizing and measuring different classes of objects. (Anati, et al., 2012) developed a robot which can localize itself by recognizing specific groups of objects (bins, clocks, ticket machines) on a simple map of a train station. (Li, et al., 2011) incorporated the structure of instances of known objects into the 3D reconstruction of a scene. Specific poles have been used for 3D reconstruction of large scale, cultural heritage in absolute scales (Pavlidis, et al., 2007)

Acquiring extra information from existing objects in the scene or visual fiducials is a feasible solution. However, the selected objects are not simple enough (from points of material, shape and pattern) to be commonly found/built in regular jobsites. Furthermore, the success rate of the suggested algorithms for reconstructing the predefined object(s) should be high enough to be reliably used in various conditions and environments.

Other than the two major approaches, there have been attempts to mathematically solve the problem for explicit settings by imposing extra constraints/assumptions. (Kuhl, et al., 2006) proposed a method based on a Depth-from-Defocus approach to calculate the absolute scale of monocular settings by combination of geometric and real-aperture methods. The proposed method does not require any prior knowledge about the scene; however, it is based on tracking objects and is not a feasible solution for large scale civil infrastructure scenes. (Scaramuzza et

al., 2009) mounted a single camera on a specific wheeled vehicle to automatically recover the absolute scale of the scene. The method is applicable for large scale scenes. Though, mounting the camera on a wheeled vehicle is not feasible in common construction job sites.

In the area of A/E/C, specific settings might be applied to solve particular problems. (Golparvar-Fard, et al., 2012), used 3D coordinates of predominant benchmarks such as corners of walls and columns, and the building information modeling (BIM) of the built infrastructure to solve the absolute scale calculation and registration problems. Later on, (Golparvar-Fard, et al., 2012), proposed a solution based on placing specific registration targets on rebar meshes to compute the absolute scale and 3D locations of rebars and embedments. In a NIST report, (Saidi, et.al, 2011), introduced the application of fiduciary markers combined with specific elaborated patterns to extract the absolute scale of built infrastructure PCD. The proposed solutions are all practical, yet limited to specific settings and are not general enough to be considered for a vast range of indoor and outdoor built infrastructure scenes. Fiduciary markers with specific elaborated patterns cannot easily be found at job sites, yet not easy to build, there is no guarantee that corners of walls and columns are reconstructed properly.

In the area of structural health monitoring, (Jahanshahi, et al., 2011), proposed an innovative approach for measuring dimensions of cracks on concrete surfaces. They assumed that the working distance (the distance between camera and the object) is known. This extra known dimension was implied to calculate the Euclidian dimensions of cracks. (Zhang, et al., 2012), utilized an unmanned aerial vehicle-based imaging system equipped with GPS and INS for 3D measurement of unpaved road surface distresses. (Carozza, et al., 2012), proposed a mark-less monocular vision based approach for localization within an urban scene based on an offline map of the environment. Their method requires a manual learning stage and manually matching several 3D model points with their corresponding image points.

As observed, most of the proposed solutions either required specific extra electronic sensors/equipment or are limited to particular settings/scenarios and are not generic enough to

immensely be applied by practitioners in the areas of construction engineering and facility management.

2.3.4. Strategies for improving the quality of generated point clouds

As mentioned before, two major approaches are utilized for reconstructing built environments and generating PCD: (1) using active sensors like laser scanners, and (2) using passive sensors such as digital cameras. It is well known that using laser scanners results in generating denser PCD with higher qualities; however, neither of the approaches is able to generate nearly perfect PCD. Some areas might be poorly reconstructed and there are always gaps or holes on the surfaces of PCD. This phenomenon usually happens due to a number of reasons including accessibility issues and occlusions (for laser scanners) and insufficient coverage of the scene, occlusions, lack of sufficient number of textures on the surfaces of elements and possible errors in calibration and optimization algorithms (for image/video based techniques).

One possible solution to this issue would be generating the PCD, identifying the missing/poorly reconstructed areas and trying to re-scan the scene to acquire extra 3D points. This solution is not always feasible since there might be changes in the original scene over time. In addition, re-scanning the scene is not helpful if gaps/holes occurred due to reasons such as lack of texture or computational errors. As the result, adding a post processing step for cleaning generated PCD and filling gaps/holes is crucial.

For most applications, gaps on surfaces of PCD are not appropriate. Depending on the size/location, gaps may cause two problems. First, they negatively impact the quality of PCD when it comes to visual representations. PCD entails several gaps and poorly reconstructed areas that are not eye catching. The second problem is related to proper extraction of geometric information from PCD. Gaps/holes and incomplete areas might be located on edges and intersections of elements and those locations are extremely significant for conducting length measurements and extracting geometric information of the scene.

The process of hole-filling on PCD can be divided into two steps: (1) determining locations of gaps/holes, and (2) smoothly covering the hole and reconstructing missing parts using available data (Wang & Oliveira, 2002). For several applications, gaps/holes on surfaces of built infrastructure PCD present simple topology, i.e. the holes are located on flat surfaces. The situation is more challenging when holes are located on intersections of different surfaces, or geometrically complex surfaces. Existing algorithms for filling gaps/holes are classified into three major categories:

1. Algebraic methods which reconstruct the missing surfaces by fitting appropriate functions based on sets of neighbor points as the input. This approach works very well for holes with simple structure located on flat/curved surfaces but is not able to successfully cover the holes located on more geometry complex or twisted surfaces.
2. Computational geometry: under this class of algorithms, different computational approaches such as region growing and Delaunay triangulation and usually leave under sampled areas.
3. Implicit functions: this group of algorithms uses various implicit functions for reconstructing missing parts and covering holes. This category is the most popular approach for filling gaps.

The following table summarizes existing algorithms suggested by researchers in each category:

Table 2.6: An overview of existing algorithms for filling gaps/holes on PCD

Category	algorithm	Reference	Input data	Geometric complex surfaces?
Computational geometry	Voronoi-Based surface reconstruction	(Amenta, et al. 1998)	meshes	no
	Delaunay triangulation	(Edelsbrunner & Muech, 1994)	points	no
	Projection-based approach for region growing	(Gopi & Krishnan 2000)	meshes	no
	Graph-based approach for region growing	(Mencel, 1995)	meshes	no
	Ball Pivoting Approach	(Bernardini, et al.	points	yes

		1999)		
Algebraic methods	Generalized implicit functions approach	(Scarloff & Pentland, 1991)	meshes	no
	Dynamic implicit functions approach	(Terzopoulos & Metaxas, 1991)	points	no
Implicit methods	Volumetric Diffusion	(Davis, et al. 2002)	Partial meshes	yes
	Moving Least Square Method	(Wang & Oliveira, 2007)	Points and meshes	no
	Compactly supported radial basis functions	(Morse, et al. 2001)	points	yes

A more detailed explanation for one famous algorithm for each category is summarized here:

Moving Least Square Method (Suggested by Wang and Oliveira, 2007)

A brief description of Moving Least Square method is presented in this section (Wang and Oliveira, 2007):

In order to fill holes, extra 3D points should be added to the un-sampled regions. For this end, the algorithm first identifies the hole's boundaries and its vicinities. For each hole, the algorithm fits a plane through the vicinity points and for each corresponding point, calculates the distance of the new point to this plane as well as its projection onto the plane. This set of distances forms a height field around the hole which is then applied for surface fitting. Following this procedure, the problem of reconstructing holes in 3D is converted to a simpler interpolation problem. Once a surface has been fitted to the height field using MLS, new points for filling the hole can be obtained by re-sampling the fitted surface. The basic version of the algorithm is presented as the following step-by-step algorithm:

1. Generate the triangle mesh from the input point cloud
2. Repeat
3. Automatically identify a hole's boundary and its vicinity
4. Compute a reference plane for the hole's vicinity

5. Calculate the distances between the vicinity points and the above mentioned
6. plane
7. Fit a surface through this height field using MLS
8. Cover the hole by re-sampling the fitted surface until no holes exist.

The Ball-Pivoting Algorithm (Suggested by Bernardini et al., 1999)

The Ball-Paving Algorithm is based on a pretty simple principle. Assume that the manifold M is the three dimensional surface of an element of the scene and S is a point sampling of M . If S is considered dense enough, it can be assumed that a ball with ρ radius is not able to pass through the surface without touching sample points. The ρ -ball can also be placed in contact with three sample points. That being said, the ball can be kept in contact with two sample points and pivot the ball until it touches the third point. It can pivot around each edge of the hole boundary so the ball contacts new triangles. The set of triangles which are formed while the ball is traversing on the hole's area continuously covers the surface of the hole. The procedure is also depicted in Figure 2 where the pivoting ball touches three vertices of the triangle $\tau = (\sigma_i, \sigma_j, \sigma_o)$ whose normal is n . The z axis is perpendicular to the page surface and points towards the viewer and m is the origin of the coordinate system. The circle s_{ij0} is located at the intersection of the ρ -ball with $z = 0$. During the pivoting procedure, the ρ -ball touches two edges with endpoints σ_i, σ_j and the ball center represents a circular trajectory γ whose center is m and the radius $\|c_{ij0} - m\|$. During the pivoting stage, the ball hits a new data point, σ_k . Consider s_k as the intersection of a ρ -sphere centered at σ_k with $z = 0$. The center c_k of the pivoting ball when it hits σ_k is the intersection of γ with s_k located on the negative half-plane of oriented line l_k . More details can be found at (Bernardini et al., 1999).

undefined areas. In particular, the diffused function propagates inward across the holes, eventually spanning them. Once diffusion is complete, the zero set of this function is the desired hole-free surface. More information about the diffusion method can be found at (Davis et al., 2002).

All of the above mentioned algorithms suffer from one issue: they perform fairly well on smooth surfaces; however, they are unable to maintain sharp edges and corners on different locations such as intersection of planes.

2.4. Problem Statement and Objectives

While videogrammetry has been conceptually proven to be viable for collecting spatial data of civil infrastructure, a number of major issues have not been fully addressed yet:

1. Low quality (i.e., blurry) and large quantity of frames existing in captured video clips from large scale civil structures. Poor quality frames mainly results from motion blur, which is a common issue in videotaping civil infrastructure scenes. It would further affect frame processing that significantly increases the reprojection errors in a 3D reconstruction pipeline. In worse cases, it might lead to failures of 3D reconstruction if a sufficient number of feature points cannot be extracted from those blurry frames. As the other issue, by increasing the number of frames, computation costs of the 3D reconstruction pipeline increases on a logarithmic scale. Capturing 2 minutes video clip by using a 25 fps camcorder, which is a common case for several off-the-shelf cameras, will result in 3000 frames. Obviously, processing all of those frames would be costly in terms of computation and performance. In civil infrastructure applications, only a few portion of these frames, e.g. 3-10 percent, suffices to generate high quality dense point clouds of scenes. Processing more frames is redundant make the procedure computationally ineffective. As the result, instead of processing all frames in a video sequence or uniformly selecting a number of frames based on capturing rates using an automated algorithm for selecting a number of informative, high quality frames are vital. As mentioned before, a number of researchers proposed algorithms for selecting key frames. In their works, common criteria which theoretically affect a 3D reconstruction pipeline, e.g. length of base line, sufficiency of overlap and degeneracy are taken into

account. None of them considered blur as a serious issue. Moreover, there is a lack of optimization steps in current key frame selection approaches. Key frame selection algorithms reduce the number of frames which are supposed to pass to processing pipeline but there is no guarantee that obtained number of frames is optimized. If number of extracted frames is less than enough, it is not possible to generate a high quality point cloud. On the other hand, processing extra frames, rather than minimum requirements is redundant. For practical purposes, any robust automatic key frame selection algorithm has to take this point into account. The objective of this research is proposing a robust algorithm for automating selection of key frames from a video sequence. The proposed method not only considers common issues that might happen while implementing a 3D reconstruction pipeline, but also covers two major practical problems: quality of frames and optimizing the number of extracted key frames.

2. Unlike most of computer vision applications, results of reconstruction methods are mainly used for measuring dimensions in civil engineering. Various applications in civil engineering such as 3D as-built documentation, defect detection, quality control of elements and evaluating deviations between as planned and as built structures demand accurate measurements. It is well known that generated point clouds are up to scale, i.e. at least one dimension of the scene should be known to scale the entire scene. Current state of practice of scaling point clouds is manual which is labor-intensive. As mentioned previously, a number of researchers have proposed algorithms for automated scaling of point clouds obtained from a single camera. Unfortunately, none of the proposed methods are general enough to effectively use for 3D reconstruction of civil infrastructure. To fill in the gaps mentioned above, one of the objectives of this research is to propose an innovative method for automatically extracting the absolute scale of generated point clouds by single cameras using pre-defined 3D cubes and with sufficient level of accuracy.
3. In most cases, the quality of point clouds obtained from videos is not excellent. Some areas of point clouds might fail to be reconstructed due to poor texture or uniform areas. The current state of research for filling those gaps and holes is not able to maintain sharp edges and corners. In this research, a new point cloud data cleaning algorithm is proposed

and validated. The hole filling component of this algorithm is able to maintain sharp edges and corners.

2.5. Scope of research

The term “built infrastructure” is too generic. In order to quantitatively define the scope of this research, the following assumptions/ limitations have been considered:

- The maximum size of an object is limited to 25 m (outdoor cases) and 5 m (indoor cases).
- It is assumed that data collections and videotaping are conducted in normal construction job site conditions (sunny or cloudy day). Evaluating the effects of light or occlusion is out of the scope of this work.
- Maximum distance between the camera and sensor is limited to 15 m (outdoor settings) and 5 m (indoor settings).
- There is no specific pattern for videotaping the scene; however, the operator should try to videotape the scene from all possible views to maximize the coverage.

CHAPTER 3

HYPOTHESIS

3.1. Overview

The proposed research plans to improve the basic prototype for monocular videogrammetric surveying of indoor/outdoor facilities. To aim for this goal, a number of pre-processing and post-processing steps are added to the main pipeline as shown in Figure 3.1:

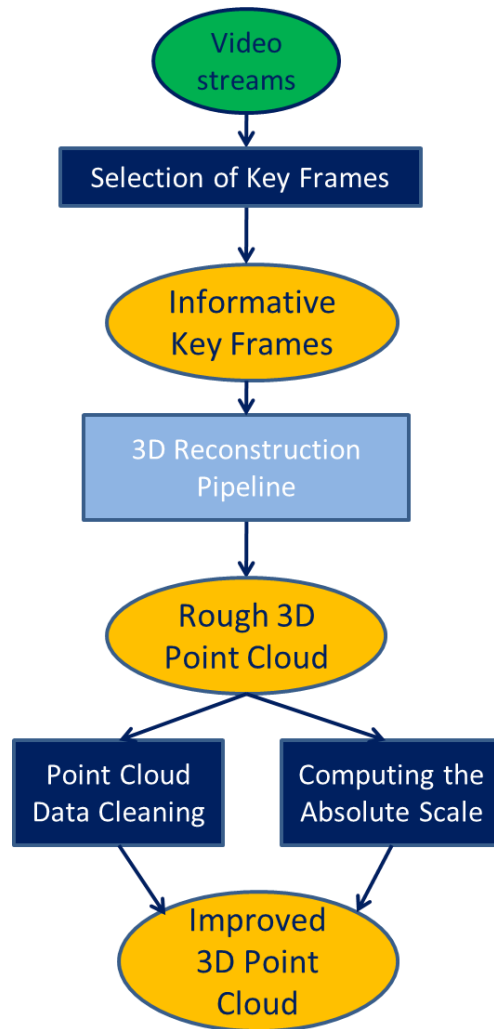


Figure 3.1: General framework for the proposed methodology

As can be observed in Figure 3.1, the main focus of this research is on the following steps:

1. Optimized selection of key video frames
2. Automated computing of absolute scale of scenes
3. Point cloud data cleaning and automated filling the holes of planes and surfaces of the generated point clouds

The details of each step are described in the following subsections:

3.2. Optimized selection of key video frames

In this section, this research presents a novel method to select key frames for the purpose of robust 3D reconstruction of infrastructure objects. The proposed method takes into account six significant factors (as mentioned in the literature: (Torr & Pollefeys, 2009) (Seo, 2008) in creating the key frame selection algorithm including: high quality frame extraction, sufficient overlap between two frames, long enough baselines, degeneracy avoidance, uniform distributions of features in the frame, and optimization of the number of the extracted key frames. Considering that the existing key frame selection criteria has been well established in dealing with sufficient overlaps, long enough baselines, degeneracy avoidance, and high quality frame extraction, this paper does not intend to re-invent these criteria. Instead, this research adopts the criteria and expands the key frame selection pipeline by incorporating the criteria that focuses on ensuring uniform distributions of features and optimizing the extraction of the number of required key frames. The main contribution of this research is the creation of the two criteria for the purpose of augmentation of the existing key frame selection algorithms.

Figure 3.2 illustrates the main workflow of the proposed key frame selection algorithm. According to Figure 3.2, the procedure starts from taking the first high quality frame as a key frame. Then a number of subsequent frames are nominated as possible key frame candidates of sufficient overlaps and enough baselines. Among the remaining candidate frames, those that lead to degeneracy cases and large re-projection errors will be further removed. Finally, the most

suitable candidate is selected based on the uniform distribution of features over the frame. To ensure that the number of extracted key frames is within an optimized range, a linear programming method is applied. Detailed explanation on each step is presented below:

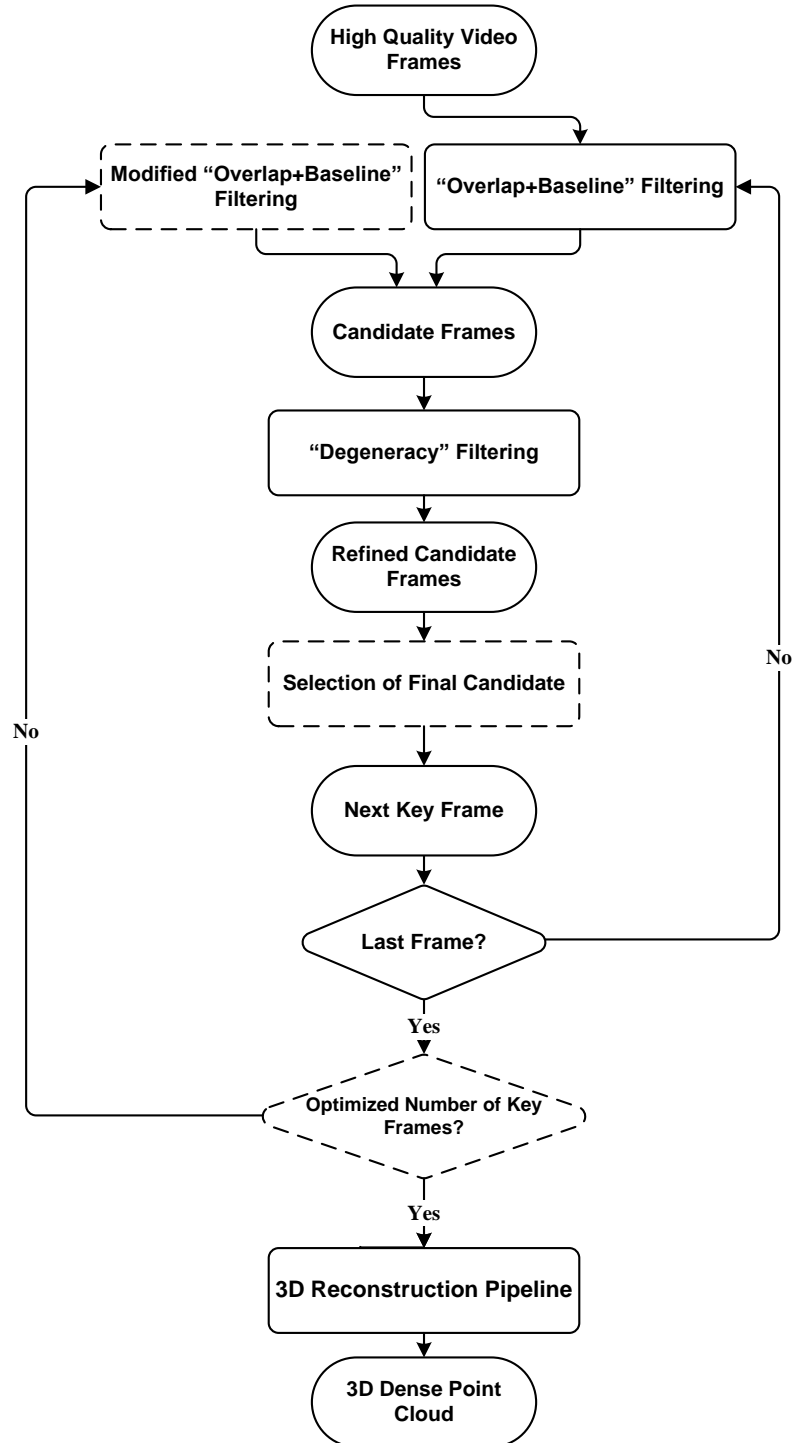


Figure 3.2: Workflow of the key frame selection algorithm

Input: High quality video frames: In order to address the blur issue, this research follows a low quality frames filtering approach and utilizes the BluM metric to measure the quality of frames and thereby remove the ones with blur effects.

Step1: Overlap and Baseline filtering: After selecting the first frame as a key frame, it is necessary to select a number of consecutive frames as key frame candidates that guarantee both enough baseline and sufficient overlap between the candidate and the first frame. To achieve this goal, this research work used the corresponding ratio defined by (Seo, 2008) and (Tahir, 2011).

Step2: Degeneracy filtering: In order to avoid degeneracy cases, this research work followed a similar strategy suggested by (Torr & Pollefeys, 2009) by calculating GRIC scores. Using GRIC score also results in maintaining frames with minimum re-projection errors.

Step 3: Selecting the final candidate as the next key frame: After filtering a number of frames from the video stream, the next step is selecting the final candidate among those remaining frames. While calculating either fundamental or homography matrices, this research realized the percentage of inliers to total number of features is a good indicator of how good two frames are modeled by these matrices. Evenly scattered corresponding points in the frame are desired since a fundamental matrix can be calculated more accurately. A more accurate fundamental matrix is useful to achieve higher accuracy of camera poses and object structure. As the result, it is preferred to choose frames that contain evenly scattered corresponding points. Considering these two criteria, i.e. percentage of inliers in calculations of homography and fundamental matrix and uniform distribution of features, this research proposes the following procedure to select the final candidate as the next key frame.

After calculating fundamental and homography matrices between the candidate and the key frames using RANSAC, it calculates the percentage of inliers to the total number of correspondences. Then it calculates the F-score to select the final candidate:

$$S = (1 - \sigma) \frac{S_F - S_H}{S_F} \quad (3-1)$$

Where:

S_H is the percentage of inliers for homography,

S_F is the percentage of inliers for fundamental matrix, and

σ is the standard deviation for measuring how uniform is the distribution of features over the frame.

To calculate the σ , the frame is divided into sub-regions. The point density for sub-regions and the entire frame will be calculated separately. Then, the standard deviation will be calculated using the following equation:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(N_i - \frac{N}{2}\right)^2} \quad (3-2)$$

Where N_i and N are number of features in each sub-region and the entire frame respectively; n is the number of sub-regions. The frame with highest S score will be chosen as the next key frame (Figure 3.3).

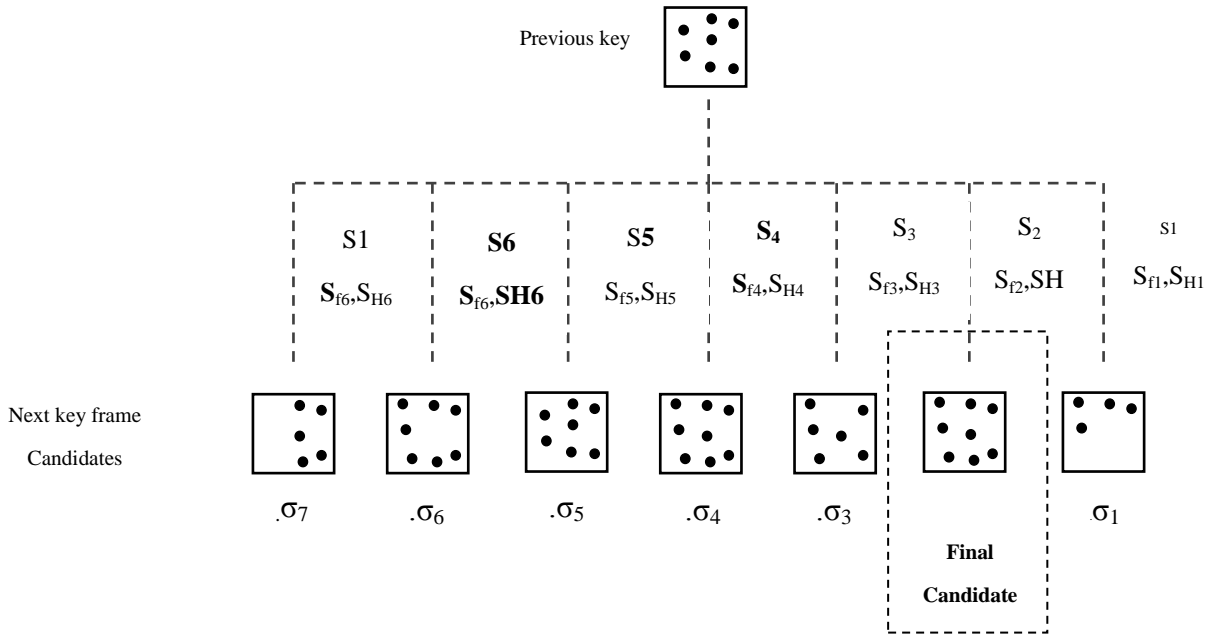


Figure 3.3: Selection of final candidate as the next key frame

Step 4: Optimizing the number of extracted key frames using a linear programming approach:

What makes this work different from the previous research is the optimization of the number of key frames needed for use in the 3D reconstruction pipeline. The corresponding ratio is defined as follows:

$$R = \frac{R_c}{R_T} \quad (3-3)$$

$$\tau_1 < R < \tau_2 \quad (3-4)$$

In equation (3-3), R_c is the number of corresponding points between the key frame and the next candidate while R_T is the total number of feature points for the first key frame. τ_1 and τ_2 are lower and upper thresholds respectively. R is inversely proportional to the length of camera motion since; as the camera moves, features tend to leave the scene. Researchers in computer vision have usually set fixed thresholds for τ_1 and τ_2 in equation (3-4) based on experiments conducted on a few datasets. However, it is not ensured that an optimum quantity of key frames can be selected. In this case, instead of assigning fixed values as the upper and lower thresholds, this research defines a specific range for each one based on three important factors: desired number of extracted key frames, approximate speed of camera while traversing, and complexity of the civil infrastructure scene.

Given a defined set of ranges for the upper and lower thresholds, this research uses a linear programming method (equations 3-5 and 3-6) to optimize the number of the required frames (Figure 3.4):

$$Goal: p_1 \leq p \leq p_2 \quad (3-5)$$

$$Constraint s: \left\{ \begin{array}{l} \tau_{l \min} \leq \tau_1 \leq \tau_{l \max} \\ \tau_{u \min} \leq \tau_2 \leq \tau_{u \max} \end{array} \right\} \quad (3-6)$$

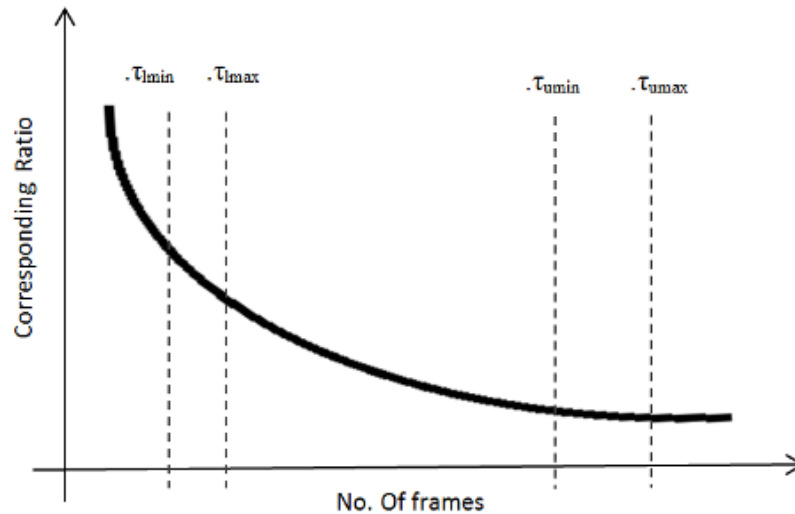


Figure 3.4: Acceptable ranges for upper and lower corresponding ratio thresholds

In equation (3-5), P is the percentage of key frames over the entire number of frames existing in the sequence; $P1$ and $P2$ define the optimum range for the number of key frames. In equation (3-6), τ_{lmin} , τ_{lmax} , τ_{umin} and τ_{umax} are acceptable ranges for the upper and lower thresholds which can be obtained through experiment for particular scenarios. The result of this step is an optimum number of key frames identified that can be used for post-processing to produce 3D points with maximum quality while maintaining the efficiency of the algorithms.

3.3. Automated calculation of absolute scale of point clouds

3.3.1. Automated absolute scale computation for outdoor settings (proposed solution)

Many A/E/C practices take place in outdoor settings so it is necessary to choose a simple, consistent object which is easily detectable and easy to use at most job sites. Among geometrical objects, a cube is the simplest. The dimensions of a cube are equal and it is typically possible to view three of its surfaces from various perspectives simultaneously. This research chose a cube made of plywood which is solid and light weight and it can be built at nearly any job sites. The size of the cube should be big enough to use in large scale infrastructure scenes, yet small

enough to be carried out and handled by only one person. Considering those factors, and based off of a number of experiments, 0.8 meter was chosen as the standard dimension for the cube.

In order to better detect the object in the scene, there are three different colors for the cube's surfaces. Two criteria should be considered while choosing the right colors for the cube surfaces: (1) the colors should be distinct from the colors of existing features in the scene, and (2) there should be a maximum difference between RGB (HSV) values of the selected colors so they can easily be identified using color detection algorithms. Considering the above constraints, different colors of blue and green were removed since those colors frequently appear in outdoor settings. Examining what remains and distributing the color values as evenly as possible across the remaining spectrum leads to the three distinct colors whose HSV values are depicted in Figure 3.5.

Given the selected colors, the overall method for calculating absolute scale mainly relies on detecting the cube in video key frames; identifying, matching and reconstructing the cube vertices along with other feature points of the scene; and scaling the obtained PCD given the known dimensions of the cube (distances between the vertices). Figure 3.6 depicts the proposed framework for absolute scale estimation.

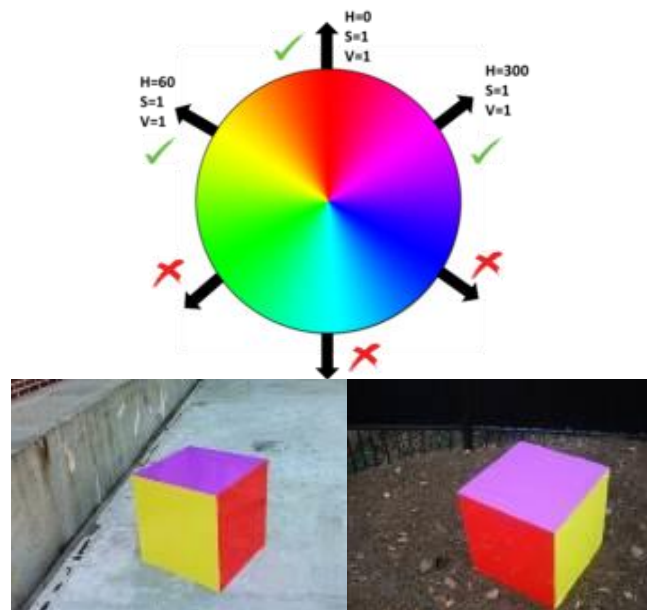


Figure 3.5: Selected colors for surfaces of the cube (top) and snapshots of the cube (bottom).

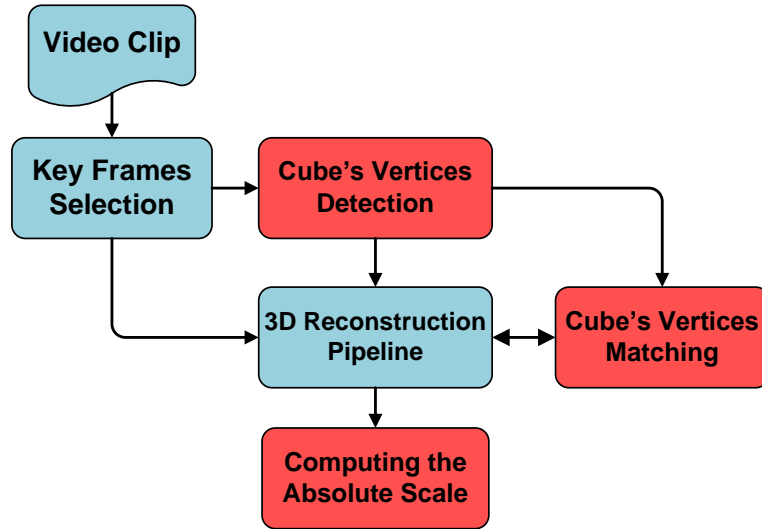


Figure 3.6: Overall workflow of the proposed algorithm for computing absolute scale of PCD
 The proposed algorithm consists of the following three steps:

Step 1: Detection of the cube's vertices

Figure 3.7 describes the necessary steps for detecting the vertices of the cube in 2D video frames captured from the scene.

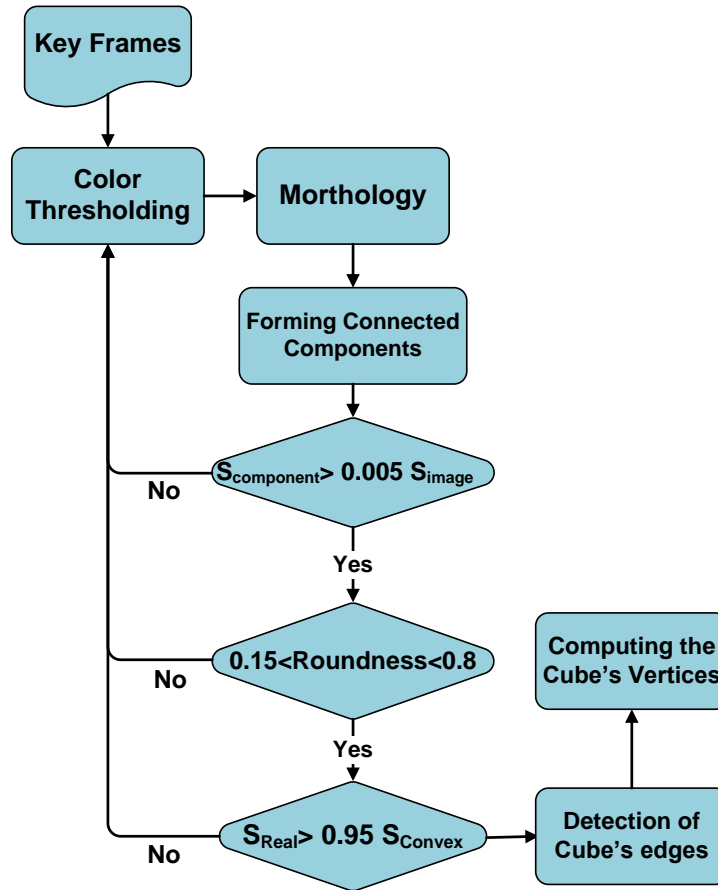


Figure 3.7: Necessary steps for detection of the cube vertices

The procedure starts with detecting the surfaces of the cube by filtering the HSV values. For each detected surface, the connected components are analyzed and an opening morphology operator (size of structuring element=3×3 pixels; two iterations) is applied to remove small areas with the same color values which do not belong to the cube's surface (Chi & Caldas, 2011). To ensure that detected areas belong to the cube surfaces, the following constraints should be met:

1. The area of the surface should be bigger than 0.005 times the area of the entire image. This criterion removes false detections of small areas that might match and also ignores detected boxes that are too far from the camera which often introduce estimation error. As explained later, the threshold value, 0.005, was experimentally obtained.

- It is assumed that each surface of the cube should look neither too long nor too circular in the image. Accordingly, the roundness of the surface, calculated by the following equation, should be located between an upper and a lower threshold:

$$Roundness = \frac{4\pi \times Area}{(perimeter)^2} \quad (3-7)$$

- Due to the perspective projection equations describing image formation, the imaged surfaces of a cube are trapezoid in shape, which is convex. To isolate potential cubes by removing non-convex objects, the real area of the surface should be approximately equal to the convex hull of the surface (Figure 3.8).

After identifying the surfaces of the cube, the edges of the cube are detected using a modified version of the Hough transform. Due to nonlinear lens distortions, the cube edges may not appear straight in the 2D images, but will be slightly curved. In order to address the issue, a modified Hough transform algorithm was implemented. The details of the modified algorithm are below:

A dilation procedure, which is a common function in image processing applications, is applied to remove some of the noises. In the modified Hough transform algorithm, all edges in different directions with a radial resolution equal to 2 degrees are recognized in the polar coordination system.

Finally, the cube vertices are identified by determining neighboring edges through their intersection points.

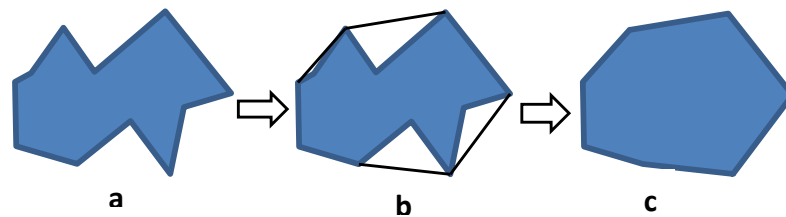


Figure 3.8: Convex hull algorithm: a) non-convex shape, b) constructing an equal convex hull for the initial shape and c) reconstructed convex hull shape

Step 2: Matching the cube's vertices across key frame views

In parallel with extracting the cube's vertices, other feature points of the scene are also recognized using SURF feature detection algorithm (Rashidi, et al., 2013). As the next step, camera intrinsic and extrinsic parameters are computed using two standard approaches: camera calibration and structure from motion (SfM). In the case of processing images, instead of manually calibrating the camera, it is possible to automatically extract the intrinsic parameters using the Exchangeable image file format (Exif) (Golparvar-Fard, et al., 2012). In this case, the camera calibration step, which might be a slightly challenging task for job site personal, is eliminated.

After detection of the cube's vertices and calculating the camera parameters, the next step is to match these vertices within two key frame views. The matching strategy of this research work consists of two components:

1. The corresponding point for each vertex in one key frame view should be located on the epipolar line for the other view (Dias, 2006). If P and P' are the camera matrices for the first and second view, the ray which is projected onto the point x in the first view is defined as:

$$X(\lambda) = P^+ x + \lambda C \quad (3-8)$$

Where C is the common camera center for both P and P' , λ is a scalar, P^+ is the pseudo inverse to P , i.e, $PP^+ = I$ and $PC = 0$. The line intersects the points P^+x and C . These points are mapped into the other camera P' at $P'P^+x$ and $P'C$. The epipolar line l' intersects these projected points and can be written as:

$$l' = (P'c) \times (P'P^+ x) \quad (3-9)$$

The point $P'C$ is the epipole e' or the projection of the first camera center into the second camera. Thus, the epipolar line can be formulated as:

$$l' = e^x \times (P'P^+ x) = [e^x]_x (P'P^+)_x = Fx \quad (3-10)$$

Where, e^x is the corresponding skew-symmetric of e' and F is a 3×3 non-zero matrix known as the fundamental matrix. Applying this criterion always limits the search area into a few candidates, usually 1 or 2, located on the corresponding epipolar line on the second view (Figure 3.9).

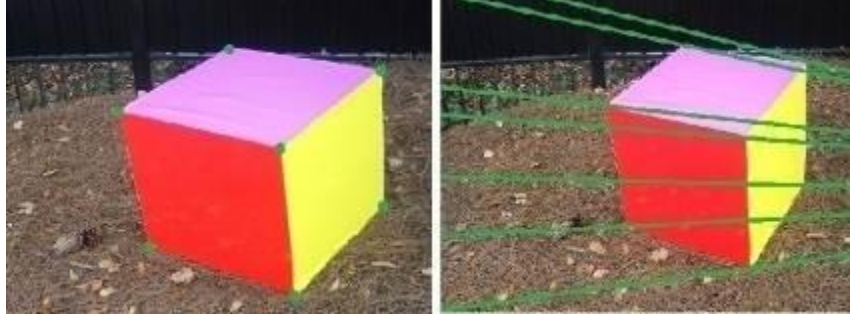


Figure 3.9: Corresponding corner points for the first view (left) are located on epipolar line in the next view (right)

2. Applying the color differences is the second criterion. We considered a rectangular window around each vertex. Since the motion of the camera between two consecutive key video frames is small, this research expects that the corresponding window in the other frame also contains similar color values. In other words, the best corresponding window is selected by following a differentiation and cross correlation approach between the color values of the two windows in two consecutive frames and calculating the similarity score as following (Rashidi, et al., 2011):

$$\text{Col - Diff (W, W')} = \sum_1^n \sum_1^m (|I_{xy} - I'_{xy}| + |R_{xy} - R'_{xy}| + |G_{xy} - G'_{xy}| + |B_{xy} - B'_{xy}|) dx dy \quad (3-11)$$

$$\text{Corr(W, W')} = \int_0^n \int_0^m |I_{xy} - I'_{xy}| + |R_{xy} - R'_{xy}| + |G_{xy} - G'_{xy}| + |B_{xy} - B'_{xy}| dx dy \quad (3-12)$$

$$\text{Similarity Score(W, W')} = \frac{\text{Core(W, W')}}{1 + (\text{Col - Diff (W, W')})} \quad (3-13)$$

Where R_{xy} , G_{xy} , B_{xy} and I_{xy} are the individual color channels and intensity values of the neighborhood pixels of the windows constructed around each vertex and n is the size of the window in pixels. W and W' refer to the first and second windows respectively.

It is necessary to emphasize that using fiduciary markers or more distinguishable patterns on the sides of cube would improve the performance of the detection algorithms; however, there are two reasons this solution was not chosen. First, it is more practical to keep the calibration object as simple as possible; and second, experiments indicate that the performance of the proposed algorithm for detecting the cube in current shape is very promising.

Step 3: 3D reconstruction of the cube's vertices along with other features of the scene

A standard 3D reconstruction pipeline was used, as introduced in (Rashidi et al., 2013), to reconstruct the vertices of the cube as well as other features of the scene. The Patch-Based Multi-view Stereo (PMVS) approach was chosen to reconstruct the entire scene and compute the PCD. Assuming that the dimensions of the cube are known, this research can scale up the entire PCD. As explained in the previous sections, the matches for the vertices come from using epipolar geometry + window search, while the others come from standard SURF matching algorithm. In order to calculate the absolute scale, at least three vertices of the cube should be successfully reconstructed. The vertices might be located on different faces so the distance between different vertices might vary. At least three vertices should be reconstructed to automatically extract all dimensions of the cube. Since the number of reconstructed vertices is usually more than three, a least square error (LSE) approach is applied to obtain a unique scaling factor for the entire scene as described below:

Assuming n is the number of reconstructed vertices, X_i is the i^{th} computed dimension with the actual length of Y_i ; the scale factor (S.F.) relates X_i and Y_i as:

$$Y_i = (S.F.) \times X_i + B \quad (3-14)$$

Where B is the computed error (in an ideal situation: B=0) and this research assumes that the distribution of errors in the 3D space is uniform. Considering the linearity assumption, the scale factor (S.F.) is calculated using the following regression-based equations (Montgomery, et al., 2012):

$$S.F. = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \quad (3-15)$$

$$B = \frac{\sum_{i=1}^n Y_i - S.F. \sum_{i=1}^n X_i}{n} \quad (3-16)$$

One important issue that needs to be taken into account is the drift problem. It is known that scaling a large infrastructure scene using a relatively small object is error prone (Botterill, et al., 2012). To address the issue, a weighting function has been added to the cost function of the Bundle Adjustment. The cost function of the Bundle Adjustment is the sum of the distance between detected points and projected points. This research set the weight of the cost function as 2 for corner points of the paper and vertex of the cube and kept the cost function weight of other points of the scene as 1; this way, this research gives priority to the important points of the scene (corner points and vertices) and reconstructs them more accurately. Another feasible solution to handle the drift problem is using multiple objects located in different parts of the scene. Using multiple objects would result in a more uniformed (instead of cumulative) distribution of errors. That being said, number, location and sizes of calibration objects play important roles in drift problem. This research work plans to focus more on this issue in future.

3.3.2. Automated absolute scale computation for indoor settings (proposed solution)

This research suggests for a proper object to use in indoor settings is a simple letter-size sheet of paper. Letter-size paper can be found in almost every indoor environment, including

homes and offices. The paper should be placed on a darker uniform surface to maximize detection (Figure 3.10).



Figure 3.10: Possible locations for the letter-size sheet of paper in indoor settings

The algorithm for detecting, matching and reconstructing the corners of the sheet of the paper is the same as those of the cube with the exception of the matching stage. All four corner points of the paper have almost the same color values; thus, it is not possible to effectively use the color differentiation criterion. The solution is straight forward: using only the four points of the corners of the paper, it suffices to implement the epipolar geometry constraint and taking note that the four corners in the first view and their correspondences in the second view are located based on a same clockwise order (Figure 3.11).

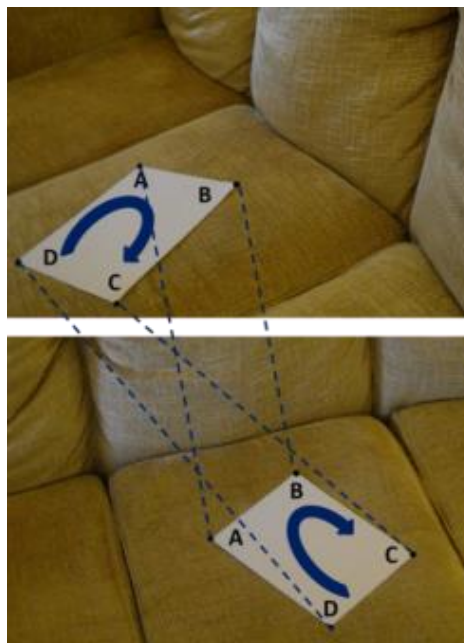


Figure 3.11: Locations of corner points of the sheet of paper follow the same clockwise order in different views

It is important to mention that using more distinctive objects such as printed sheets with elaborated patterns and codes might also leads to very accurate results but the advantage of this method lays on the simplicity of the chosen object, as well the sufficient accuracy of the results.

3.4. Point cloud data cleaning

As shown in Figure 3.12, the process of point cloud data cleaning mainly consists of three stages:

1. Removing outliers
2. Filling holes and gaps
3. Balancing the density of different parts of the point cloud using a plane recognition approach.

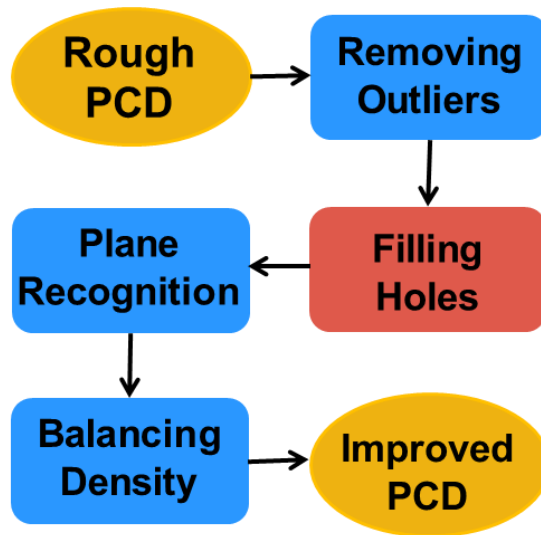


Figure 3.12: Overview of the proposed algorithm for point cloud data cleaning
Details of each stage are summaries below:

- **Removing outliers:**

In this research, outliers are considered as unattached points in the 3D space; thus, it is possible to remove them based on the distance between a point and its neighbor points. Based on a number of conducted experiments, outliers are considered as points that are located in cells with the density less than 5 points per cell. The average distance between the point and 5 closest

neighbor points is more than 5 cm and the density of the cell is less than 50 % of the average density of neighboring cells. It is also necessary to mention that the algorithm is not able to remove points belong to unwanted objects like trees and surrounding environment.

- **Filling holes and gaps:**

The innovative hole filling algorithm presented in this research is based on the following assumptions:

1. There might be some natural holes on surfaces of objects in the scene. The algorithm is not able to automatically differentiate between these real holes and artificial gaps and holes on surfaces of generated point cloud data. A manual inspection stage might be necessary so the user can verify the correct holes.
2. By definition, the algorithm is designed to detect and reconstruct holes whose longest dimension is between 5-100 mm. Holes smaller than the thresholds are not defined as insignificant and can be filled by implementing the plane recognition algorithm and filling large scale holes are out of scope of this research. Plus, there is a high chance that big holes are natural holes that should not be filled. The thresholds have been obtained based on experiments on real built infrastructure scenes.

The proposed approach for filling holes and gaps consist of the following stages (Figure 3.13):

1. Detecting the hole
2. Expanding the hole area
3. Detecting neighborhood edges
4. Restoring missing parts of the edges

- Implementing a surface reconstruction algorithm considering edges as extra constraints.

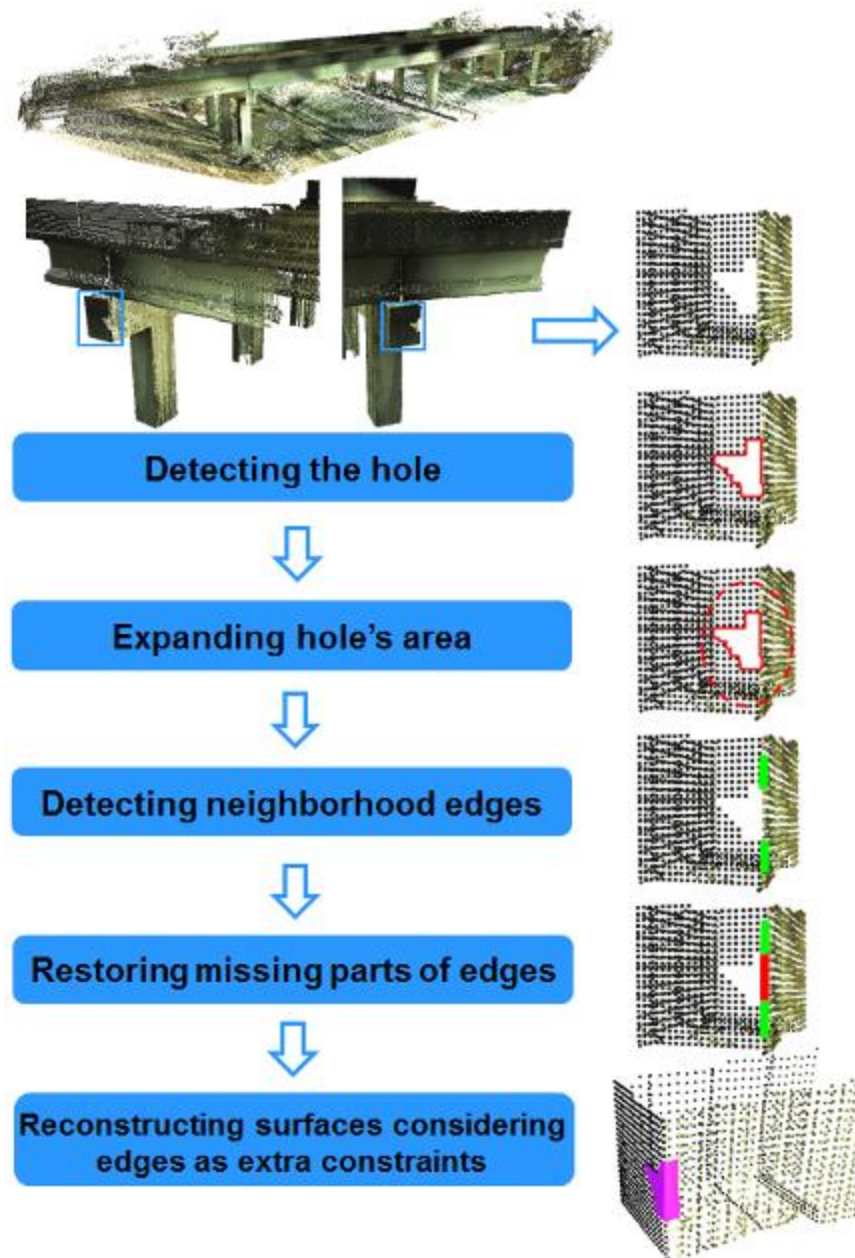


Figure 3.13: Different stages for 3D edge based reconstruction of holes

It is necessary to mention that based on Wang and Oliveira (2007), a hole consists of a loop of boundary edges (Figure 3.14). Also, sharp edges on surfaces of point cloud data are defined as the points with sudden changes in normal vector values (Figure 4.14). For regular

covering of gaps and holes, the ball pivoting algorithm was implemented. Details have been presented in previous chapter.

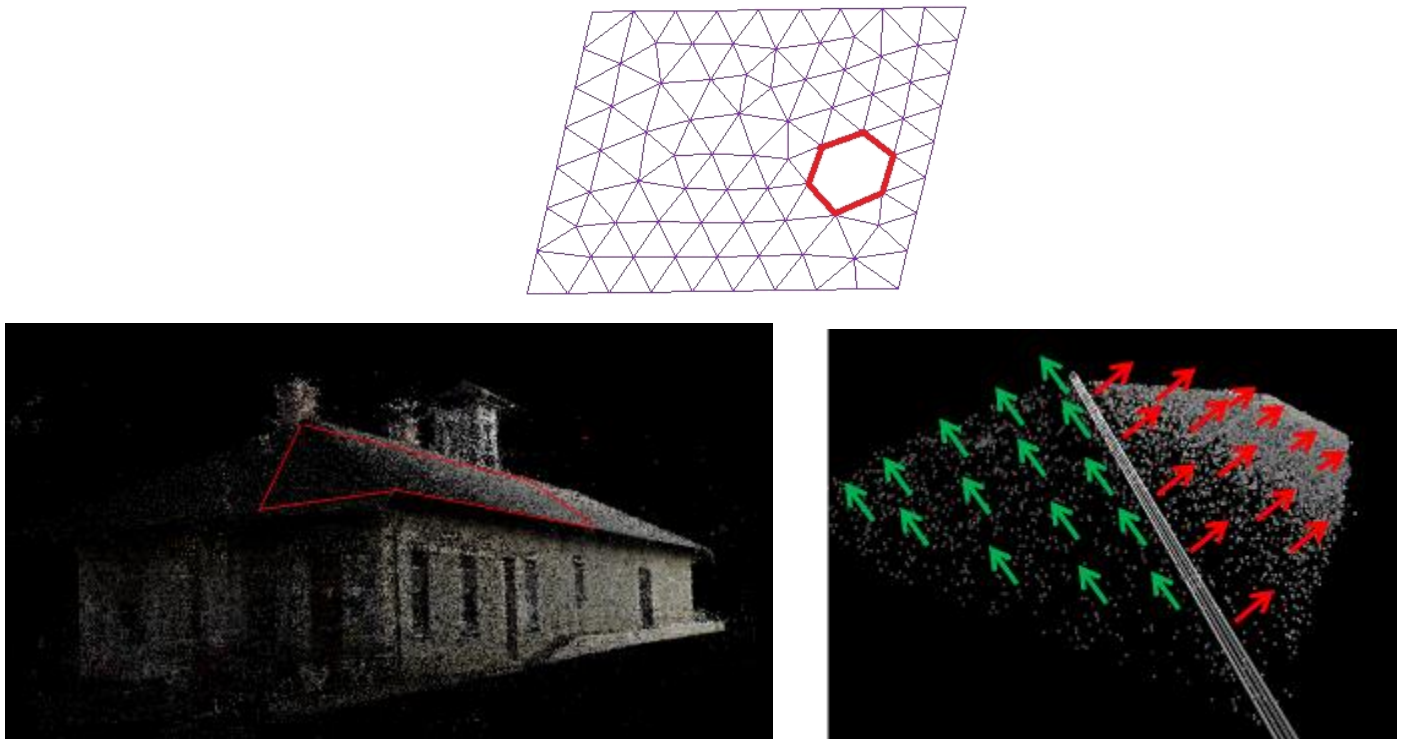


Figure 3.14: Definition of hole (top) and definition of 3D edges on point cloud data (bottom)

- **Balancing the density of point clouds**

Balancing the density of point clouds on surfaces is a pretty straight forward procedure. After implementing a plane recognition algorithm suggested by (Zhang et al., 2013), the detected planes can be divided into cells of a rectangular mesh. If the number of points in each cell is less than the assumed threshold, extra points should be added to the cell (up-sampling). However, if the number of points is more than the threshold, the extra points should be removed from point cloud data (down-sampling). The entire procedure is shown in the following figure:

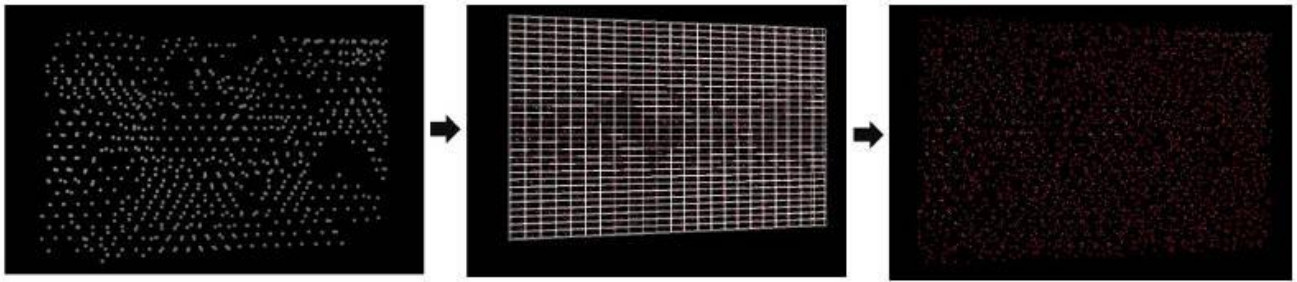


Figure 3.15: Different stages for balancing the density of point clouds

CHAPTER 4

SYNTHESIS

4.1. Prototype

A prototype has been created using Microsoft Visual C# and Windows Presentation Foundation (WPF) to implement the proposed frameworks. OpenCV (Intel® Open Source C++ Computer Vision Library) has been selected as its main image processing library and Emgu CV, a cross platform .Net wrapper to OpenCV, has been used to connect the platform and the library. They both are free and open source.

The prototype also provides a base to connect to any number of cameras through Ethernet network, USB and IEEE 1394b connection with real-time responsiveness. It also provides capability to perform basic operations on image streams read from the cameras such as image stream controls, buffering, image caching, video files encoding and communicating with the cameras.

4.2. Basic videogrammetric pipeline for generating dense 3D point clouds

This section briefly overviews the required steps for reconstructing a 3D dense point of civil infrastructure from a sequence of video frames. To achieve the maximum possible accuracy of the result, the camera calibration method proposed by (Zhang, 1999) was used for extracting the intrinsic parameters. As the next step, feature points on different frames are extracted and matched. In this research, considering the accuracy and efficiency in computation, the SURF method was used as feature detector and descriptor, as suggested by (Rashidi, et al., 2013). Note that unlike arbitrarily taken images, the sequential property of a video stream can be utilized to make the entire procedure more efficient. Information from each video frame can build upon the previous frame. Specifically, instead of matching feature points between all frames which is the

common case for image-based 3D reconstruction methods, it just needs to match the features between a limited numbers of consecutive frames. After extracting and matching feature points and knowing the intrinsic parameters of the camera and camera motion, i.e., translation and rotation of the camera, the 3D coordinates of the feature point could be computed through a procedure known as structure from motion. This research used the 5 point algorithm proposed by (Nister, 2004) to calculate the ego-motion of the camera (extrinsic parameters). In order to refine the entire process and minimizing the propagated errors, a global optimization technique of Sparse Bundle Adjustment is utilized. The final step of the solution is to generate the dense point cloud. Based on the literature review, the most effective algorithm to generate the dense point cloud is PMVS proposed by (Furukawa & Ponce, 2010). The main advantage of this method over other algorithms lies in its ability to reconstructing infrastructure components with fine surface detail despite low-texture regions, large concavities, and/or thin, high-curvature parts. This ability makes the method feasible for reconstructing civil infrastructure scenes since civil infrastructures mainly consist of texture-less areas (concrete surfaces of columns and deck of bridges), concavities (intersection of structural elements) and curvature elements (surfaces of vessels, reservoir or pipes).

4.3. Key frame selection: Implementation

Considering the variety in civil infrastructure senses, 25 video streams were captured from 8 different scenes: two highway bridges, three campus buildings, one residential building, one sport facility and one concrete water reservoir. The lengths of the video streams vary from 4-10 minutes. To validate the degeneracy cases, some video streams contain planar scenes like walls, and in some cases during the process of capturing videos, the video taper intentionally stops for a while and just rotates the camera without any translations. The implementation process takes place in three phases.

4.3.1. Identifying the thresholds

It is important to determine the threshold in order to apply the metric. The threshold is used to determine which frames should be removed. If the measured value of a frame by the metric is larger than the threshold, the frame is removed; otherwise, it is retained. In determining the threshold, the criterion is that a minimum level of frame quality required to obtain robust results is ensured. To achieve this, a sample set of satisfactorily high quality images are selected through human observations. Then these sample images are used to calculate their blur effect values for the estimation of the sample mean and sample standard deviation of the sample dataset. By applying the left-sided 95% confidence limit for normal distributions, an estimate of the threshold can be statistically determined.

4.3.2. Strategies for low quality frames filtering

In order to implement the key frame selection algorithms, values of thresholds depicted in equations 3 and 4 has to be identified. Figure 4.1 shows the relationship between corresponding ratio and number of frames for a number of captured video streams from various civil infrastructure scenes. As shown in Figure 4.1, rapid changes in R might happen in different points:

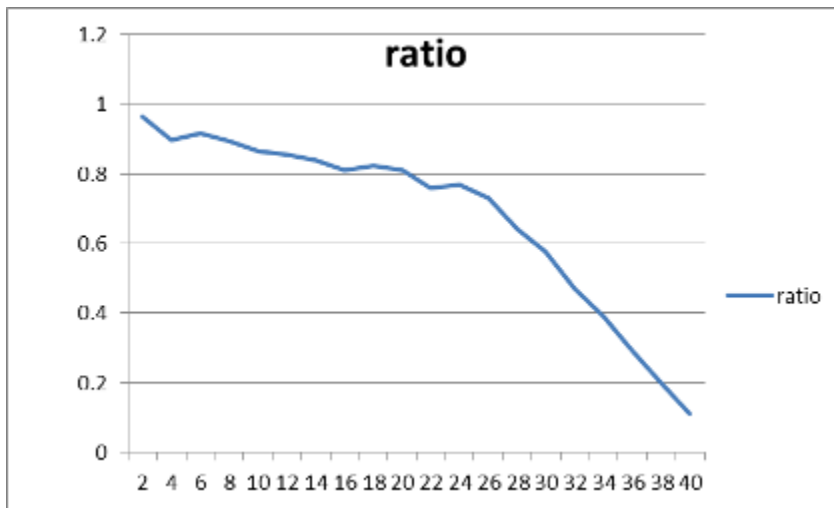


Figure 4.1: Samples of corresponding ratios for different video streams

Based on the experiments on civil infrastructure scenes, the changes in corresponding ratio mainly depend on two parameters:

1. Speed of camera movements, it is very difficult to measure the real speed of camera when it traverses around the job site. However, considering 25 fps as a common rate for a regular video camera, this research work has defined two categories for camera movement speed:
 - a) Normal movement, the video taper traverses with the hand held camera with the speed more than around 1 m/s.
 - b) Slow movement, the video taper traverses with the hand held camera with the speed less than around 1 m/s.
2. Uniformity of the scene, some of civil infrastructure scenes contain uniform texture like long span bridges. On the other hand, some other civil infrastructure scenes contain complex variable texture. Corresponding ratio declines rapidly in complex scenes where the texture of the scene is changing rapidly. However, in more uniform scenes, e.g., a long span bridge or a uniform wall, the changes are slower. Based on these observations, scenes are defined as either complex or uniform:

Ability in extracting sufficient numbers of feature points is the key factor for measuring the uniformity and complexity of the scene. Complex scenes contains multiple feature points while it is very difficult to extract sufficient feature points from uniform scenes (Figure 4.2)

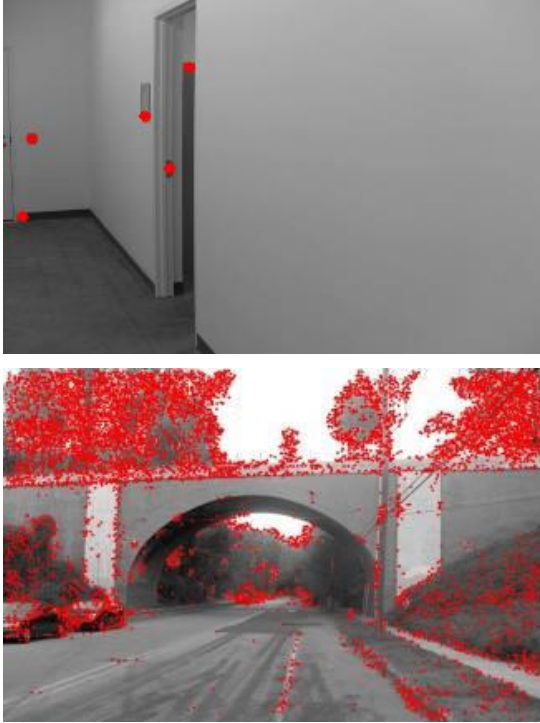


Figure 4.2: Uniform scenes with texture-less areas (top) and complex scenes with texture-full areas (bottom)

In this research, any scene with more than average of 100 feature points per 2 MP frames is considered as complex. For higher or lower resolutions, this threshold needs to be adjusted accordingly. The threshold has been defined based on experiments on 15 different indoor/outdoor built infrastructure scenes.

The other important factor in defining the upper and lower thresholds is the approximate number of required key frames. Infrastructure scenes could be reconstructed using various numbers of frames, which can be a trade-off problem. Processing more frames resulted in higher quality generated point clouds with more expensive computations. In order to find the optimum number of frames required for processing, different numbers of key frames were extracted and processed for different video streams captured from one specific scene, i.e. a highway bridge. The completeness factors of generated dense point clouds for different numbers of key frames were calculated using a method explained later. Results of measuring completeness factor for different numbers of key frames for one video sequence is illustrated in Figure 4.2. As shown in Figure 4.2, for a capturing rate equal to 30 fps, the maximum density of a point cloud is

achievable by processing a few portion of frames (5-15 percent), meaning that processing more frames is redundant. This fact was considered as another criterion in selecting corresponding ratios in order to achieve highest completeness of the dense point cloud while minimizing the number of key frames. It is possible to achieve higher percentages of completeness by increasing the resolution or videotaping the scene from more views; however, increase in the density would not be guaranteed. Plus, for many processing purposes, it is not always necessary to achieve the highest levels of completeness.

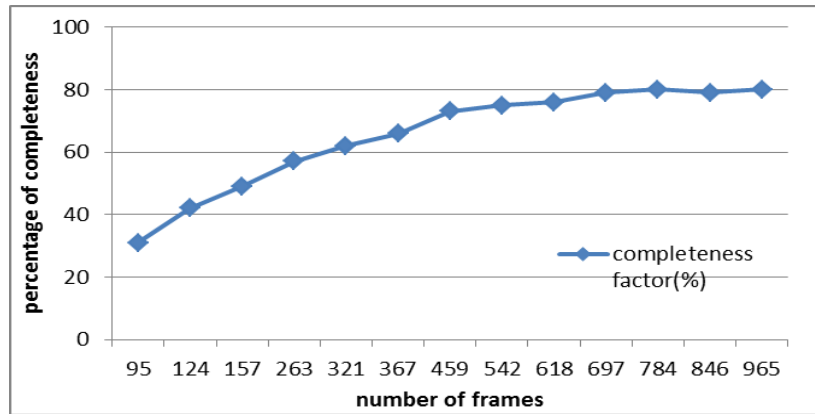


Figure 4.3: Completeness of generated dense point clouds versus number of processes key frames

Based on the above mentioned different scenarios, the ranges proposed in the table 6 are used as acceptable ranges for upper and lower thresholds.

Table 4.1: Upper and lower thresholds for corresponding ratios

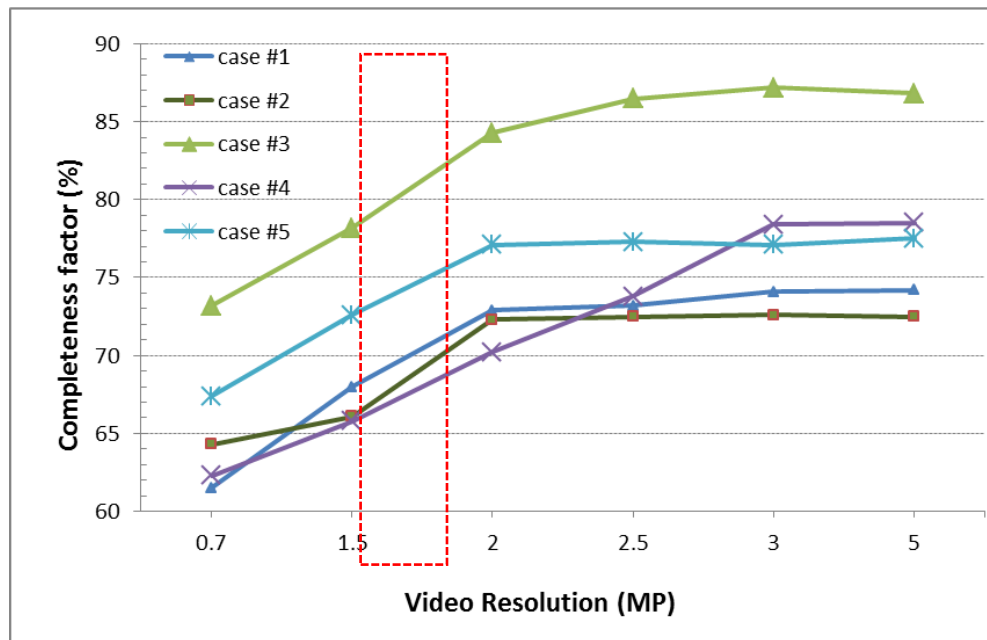
Complexity of the scene	Camera motion speed	Lower threshold (T ₁)	Upper threshold (T ₂)
Complex	Normal	0.85-0.75	0.60-0.5
Uniform	Normal	0.8-0.7	0.65-0.55
Complex	Slow	0.85-0.75	0.70-0.6
Uniform	Slow	0.9-0.8	0.70-0.6

*uniform scene: scenes with uniform texture such as bridges and concrete surfaces

*complex scene: scenes with complex surfaces like buildings with complex view

Another significant variable for experiments’ design is the proper resolution of cameras. It is known that by increasing the resolution it is possible to extract more feature points and as the result, generate denser point cloud data. However, this assumption is valid until reaching a

certain thresholds and beyond that, increasing the resolution is redundant and just makes the computing process more time consuming. In order to select the right resolution for videotaping different scenes, a number of experiments were conducted. Five outdoor and 4 indoor settings were selected and videotaped using a number of common resolution ranges (0.7-5 MP). Each captured video clip was processed and a dense point cloud for each video was generated. Completeness factors for a number of surfaces on each point cloud were calculated following the same procedure explained in the comparison study (chapter 2). The results are summarized in the following Figures:



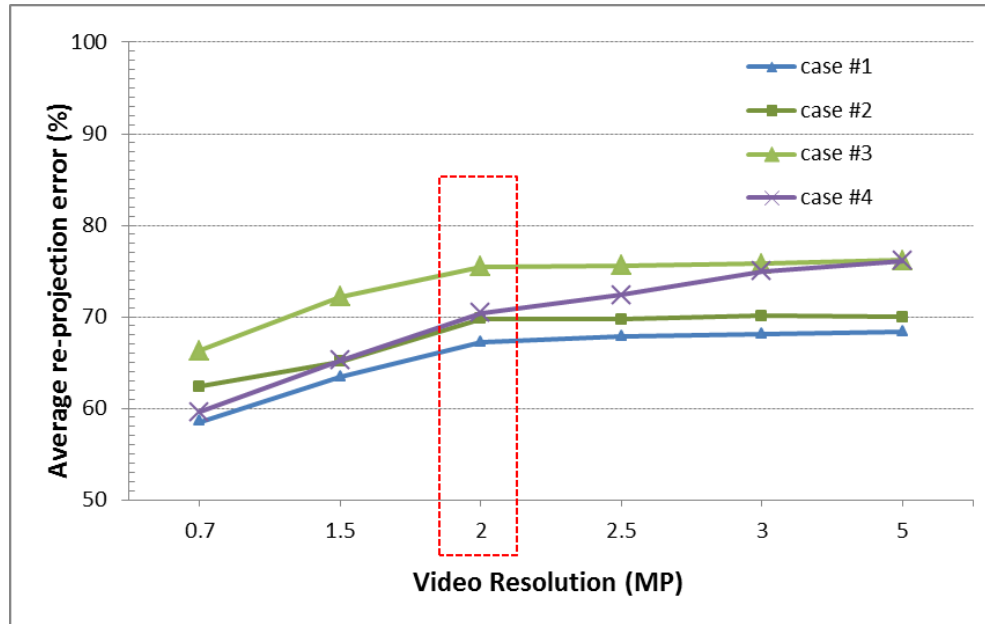


Figure 4.4: Proper resolution for outdoor (top) and indoor (bottom) settings

As presented in the graphs, 2MP is a proper choice as a common resolution for applications in AEC industry. Plus this resolution is easily achievable using off-the-shelf camcorders and smartphones.

4.4. Automated computation of absolute scale: Implementation

A C# based prototype was implemented to test the validity of the proposed algorithm. It was written in Visual Studio 2010 using Windows Presentation Foundation (WPF) and publicly available libraries such as OpenCV 2.0 (wrapped by EmguCV) for access to computer vision tools and DirectX 10 for the graphic display of results. The Open CV's image structure was the primary data structure. It removed the conversion needs of the image processing tools from that library, which significantly reduced the processing speed. The aim of the experimental setups is two folds: (1) identifying the thresholds for applying in the proposed algorithms, and (2) evaluating the performance of the implemented algorithms as well as the overall performance of the proposed method. Each step is explained in the following sections:

4.4.1. Identifying thresholds for the minimum acceptable area of the cube in images

As previously explained, if the areas of the cube surfaces in images were too small, i.e. the cube is located too far from the camera, the estimated errors in detecting and reconstructing the cube corner points would increase significantly. To tackle this issue, this research implemented a specific threshold as the minimum acceptable area of a surface of the cube, compared to the total area of the image. Frames including the cube surfaces smaller than the calculated threshold are removed from further processing.

In order to identify a proper threshold, this research conducted a number of experiments. Considering the variety in built infrastructure scenes, the cube and the sheet of paper were placed in 10 outdoor and 10 indoor built infrastructure scenes. The scenes were videotaped from different views with varying distance of the camera from the calibration object. As the first step, the video clips were processed and the surfaces of cubes were detected. The success rates of detecting the surfaces were measured using the precision and recall values as defined in the following equations:

$$Precision = \frac{TP}{TP + FP} \quad (4-1)$$

$$Recall = \frac{TP}{TP + FN} \quad (4-2)$$

In these equations, TP is the number of correctly detected cube surfaces' (paper) pixels, (TP+FP) is the number of detected cube surfaces' (paper) pixels, and (TP+FN) is the number of actual cube surfaces' (paper) pixels. Precision basically means the area of correctly recognized cube region divided by the total area of recognized cube regions and measures the “exactness” of the detection algorithm. Recall is known as the area of correctly recognized cube regions divided by the area of actual cube regions and shows the “completeness” of the detection algorithm.

The results of calculating precision and recall ratios for different sizes of the calibration objects compared to the entire size of the frames are illustrated in Figure 4.3.

In the next step, the corner points of the calibration objects were detected and reconstructed. The average errors in computing the 2D locations of the extracted corner points compared to the actual locations, as well as the re-projection errors for calculating the 3D locations of the corner points in the space were computed and demonstrated in Figure 4.4 .In this study, the 2D location error (%) was calculated by dividing the distance between the computed and actual locations of the vertex on the image to the length of the longer edge of the cube (paper) to where the vertex is located. The same approach, but in 3D, was implemented for computing the re-projection errors.

To determine the threshold, the minimum precision and recall rates set to 95% and 90% respectively. In addition, maximum allowable error in 2D location of corner points and re-projection error are considered as %2 and %1. As shown in Figures 9 and 10, the smallest ratio for achieving the above mentioned levels of accuracy is between 0.5-1 percentages. As the result, the minimum ratio of each component surface to the entire image surface was set to 0.005 (0.5%).

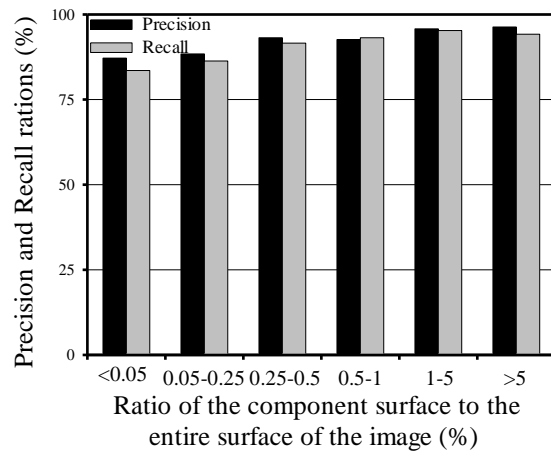
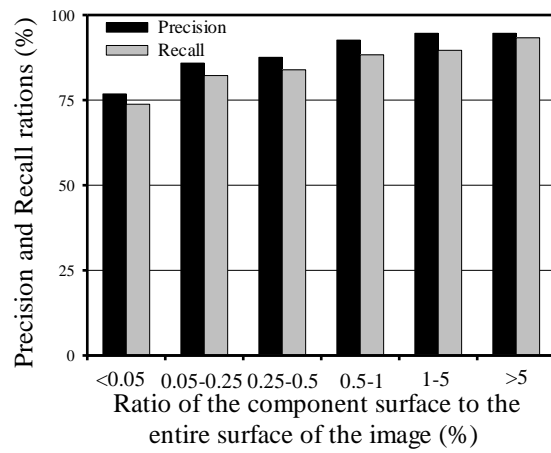


Figure 4.5: Precision and recall ratios for detection of the cube surfaces (top) and sheet of paper (bottom)

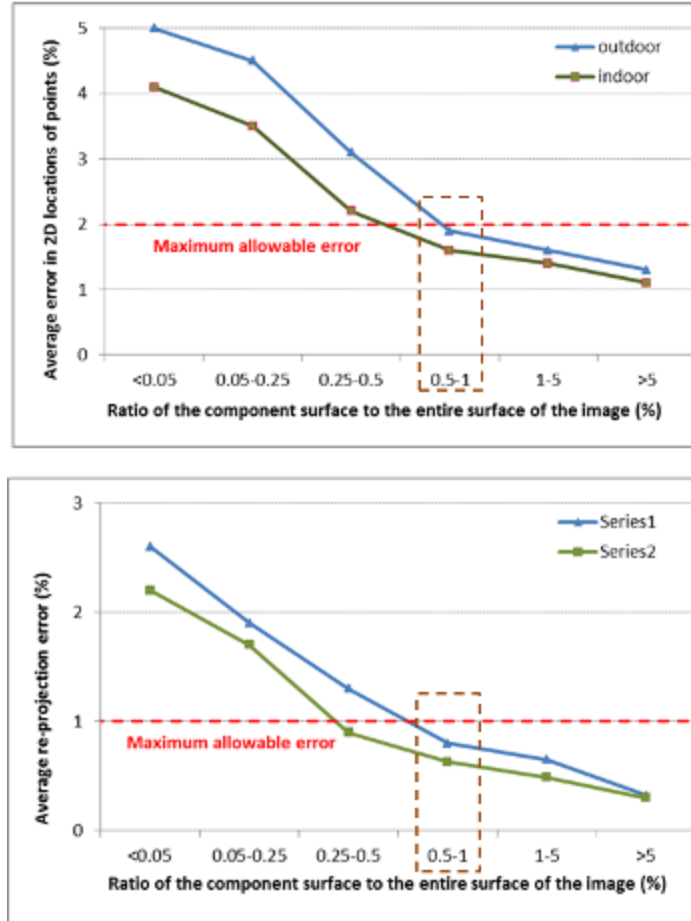


Figure 4.6: 2D location errors (top) and re-projection errors (bottom) for both indoor and outdoor settings

4.4.2. Identifying thresholds for the maximum and minimum roundness factors

Using the same video data as the previous section, the roundness factors for the cube surfaces (paper) in 437frames were computed. Upper and lower thresholds for the roundness factor can be identified by calculating the confidence intervals for this set of the measured roundness factors:

$$\text{upper and lower thresholds} = \left(\mu - 1.96 \frac{\sigma}{\sqrt{n}}, \mu + 1.96 \frac{\sigma}{\sqrt{n}} \right) \quad (4-3)$$

Where the confidence level is 95%, μ is the mean and σ is the standard deviation of the measured roundness factors. After plugging the observed values, the upper and lower thresholds were set to 0.85 and 0.1, respectively.


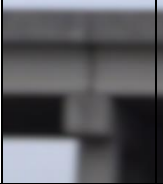


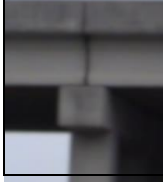

CHAPTER 5

VALIDATION

5.1. Validating the key frame selection algorithm

According to (Sheikh et al., 2005), an obvious way of measuring the quality of an image or video is to solicit opinion from human observers. Thus, 20 random frames from each sequence with visually satisfactory quality were observed and selected from each video clip to statistically estimate the threshold of the blur metric. Based on the threshold, the blur metric can be incorporated into the automatic process of the proposed method. Applied with the blur metric, 20 frames resulted in the sample mean 0.283, and the sample standard deviation 0.01 of the evaluation scores. The left-sided 95% confidence limit for a normal distribution is then used to determine the statistical estimate of the threshold as $0.283 + 1.64 \times 0.01 = 0.299$. This means that with 95% likelihood, any “satisfied” image would have a score measured by the proposed metric no more than 0.299. It is noted that the sample size is statistically significant to the threshold analysis in consideration of the expected accuracy level being in the order of 0.001 and the relatively small variation on the sample standard deviation of the resulting scores. In order to quantitatively evaluate the impact of blur on performance of reconstruction pipeline, those 20 frames were artificially blurred. Then, blurred frames were fed into the reconstruction pipeline. For each case, the BluM value, the number of extracted and matched feature points, and reprojection error obtained from structure and motion algorithm were computed. As an example, those values for 6 sample blurred frames obtained from a highway bridge sequence are presented in Table 5.1. As observed, in all cases blur has a significant effect on number of extracted feature points and reprojection error. In most cases, the number of feature points extracted from blurred frames was not even sufficient to robustly reconstruct the scene. The results verify the validity of the blur metric’s threshold accuracy.

Table 5.1: Impact of blur on number of extracted feature points and reprojection errors

	BluM value	Number of extracted feature points	Number of matches	Reprojection error		BluM value	Number of extracted feature points	Number of matches	Reprojection error
	0.259	2984	657	0.01025		0.622	1017	316	0.0375
	0.329	2562	579	0.0194		0.655	299	107	0.0461
	0.465	2077	474	0.0223		0.704	147	24	0.0492

In the next step, a number of key frames were extracted for each video sequence using this research work's key frame selection algorithm as well as the methods developed by (Pollefeys et al., 2004; Seo et al., 2008; Thormahlen et al., 2004). In addition, a number of frames equal to the number of key frames extracted by our method were selected at equally distributed intervals without using any key frame selection method. These extracted key frames were processed in the videogrammetric pipeline and percentages of failure cases as well as reprojection errors for successfully reconstructed cases for each method are computed. The obtained results are summarized in Table 5.2.

Table 5.2: Failure percentages and re-projection errors for different key frame extraction methods

Method	Average failure percentage (%)	Average reprojection error	Average number of extracted key frames
Uniformly extracted frames	54.54	0.0354	576
Pollefeys et al.	36.36	0.0112	854
Thormahlen et al.	27.27	0.00746	625
Seo et al.	45.45	0.0174	842

Our method	22.72	0.00975	576
------------	-------	---------	-----

As it can be observed from Table 5.2, the paper’s proposed method outperforms almost all other methods in terms of failure cases and average re-projection errors. Moreover, in this specific case, the number of extracted key frames is within the desired range.

5.2. Validating the proposed absolute scale calculation algorithm

The validation procedure took place in two steps:

Step 1: Validating the performance of the corner points’ detection and matching algorithm

To evaluate the performance of the corner points’ detection and matching algorithms, this research selected four indoor and five outdoor cases as our case studies. These case studies are different from the initial scenes which were used for computing different thresholds. The indoor cases include offices and different locations of homes, e.g. bathroom, living room and kitchen, while the outdoor cases cover a variety of civil infrastructure scenes including campus buildings, highway bridges, a train station building, a sport facility and an under-construction wall in a construction jobsite. Each scene was completely videotaped as much as possible. The user should traverse around the entire scene and videotape it from multiple viewpoints to minimize occlusions. An off-the-shelf Canon Vixia-HF S100 was utilized for data collection purposes. The corners point detection and matching algorithms were implemented for each captured video clip separately (figure 5.1) and the associated errors were measured in terms of precision and recall values for the surface detection algorithm, deviation between computed and actual 2D location of corner points for corner point detection algorithm and percentage of successfully corresponded corner points for the matching algorithm. The summary of the results are presented in Table 5.3.

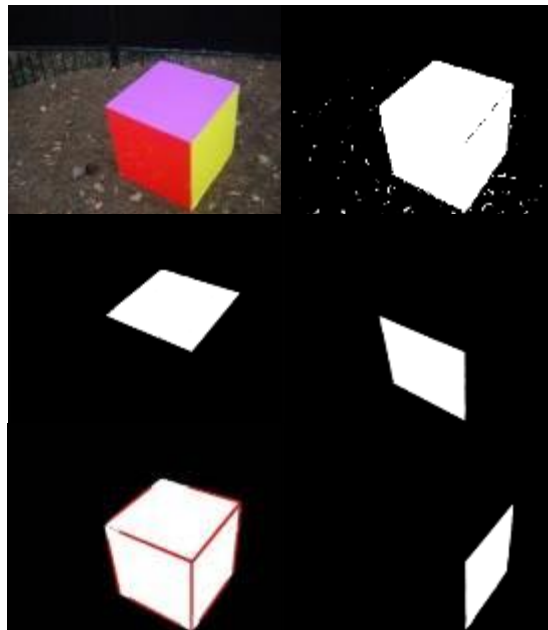
As shown in Table 5.3, the performance of the detection algorithm was the best for yellow surfaces. It is necessary to highlight that we do not need to detect and reconstruct all the cube

vertices in all frames. It is only sufficient to successfully detect and reconstruct three vertices of the cube for the entire video clip.

Table 5.3: Summary of the results obtained from implementing the corner detection and matching algorithms

Experimental setting		Average error in 2D corner points detection algorithm*		Average error in 2D corners point detection algorithm *	Average accuracy of 2D matching algorithm (%)
		Precision (%)	Recall (%)		
Outdoor setting (cube)	Surface			0.03	98.7
	Red	92.1	90.8		
	Yellow	96.5	94.1		
	Purple	91.8	89.9		
Indoor setting (sheet of paper)		98.3	92.1	0.01	100

*error is calculated as Δ/l where Δ is the deviation between actual and computed 2D location of the corner point (in pixel) and l is the longest associated vertex.



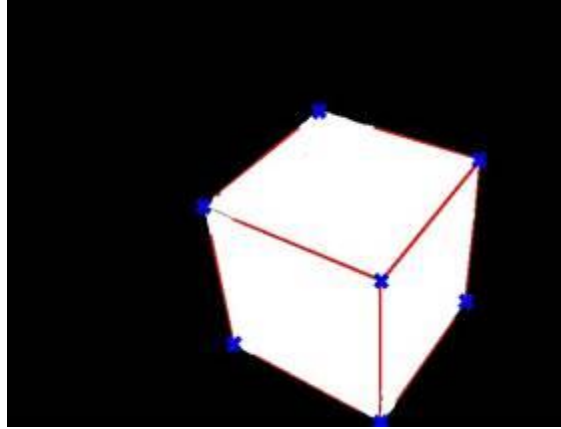


Figure 5.1: Sample of the implementation results for the cube corners detection algorithm: from top left: the original image of the cube-the result of filtering the image based on HSV thresholds- Detected red, yellow and purple surfaces- detected lines based on the improved Hough transform - and the intersections of the cube edges as the final result

Step 2: Validating the overall performance of the proposed algorithm for computing the absolute scale PCD of the scenes

To validate the overall performance of the proposed methods, the captured video clips were processed and the absolute scale PCD for each built infrastructure scene was generated following the procedures explained in the methodology section. For each case study, this research considered the deviation between a number of real dimensions and computed dimensions of the scene as the metric for measuring the accuracy of the presented methods. For each scene, several dimensions and distances were identified and measured by a TC805 total station for outdoor cases and a Leica DISTO D5 Laser measurer for indoor environment (Figure 5.2). The average measuring time for each dimension measurement of the outdoor setting is around 15 minutes. This time includes traversing between different locations within the jobsite, setting up and adjusting the total station, conducting measurements, converting the files into the computer, manually finding the corresponding dimensions on the PCD and calculating the scale factor.

Samples of generated PCD for both indoor and outdoor case studies are presented in Figures 5.3 and 5.4.

The results of computing the accuracy of the proposed methods in measuring different dimensions within built infrastructure case studies are summarized in Table 5.4.



Figure 5.2: Actual distance measurements and preparation of ground truth: Leica TC805 total station (left) and Leica DISTO D5 Laser measurer (middle and right)

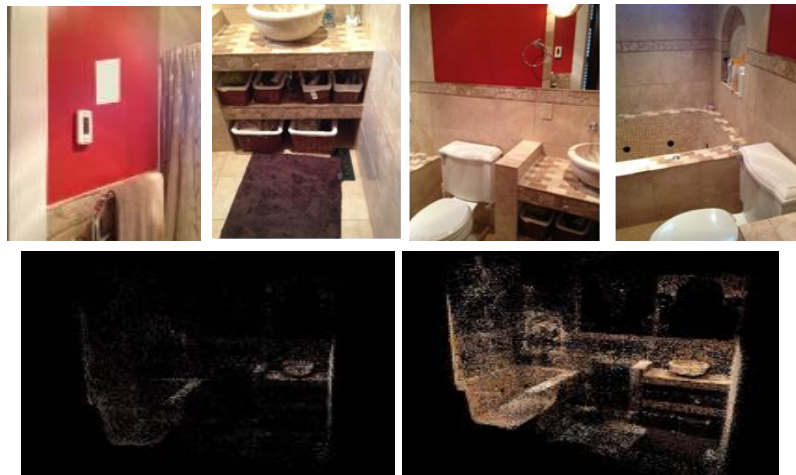


Figure 5.3: A sample of the generated PCD for indoor settings: bathroom- Sparse PCD generated by SfM (left) and PCD generated by PMVS (right)



Figure 5.4: Samples of the generated PCD for outdoor settings: Campus building (top) and construction wall (bottom)

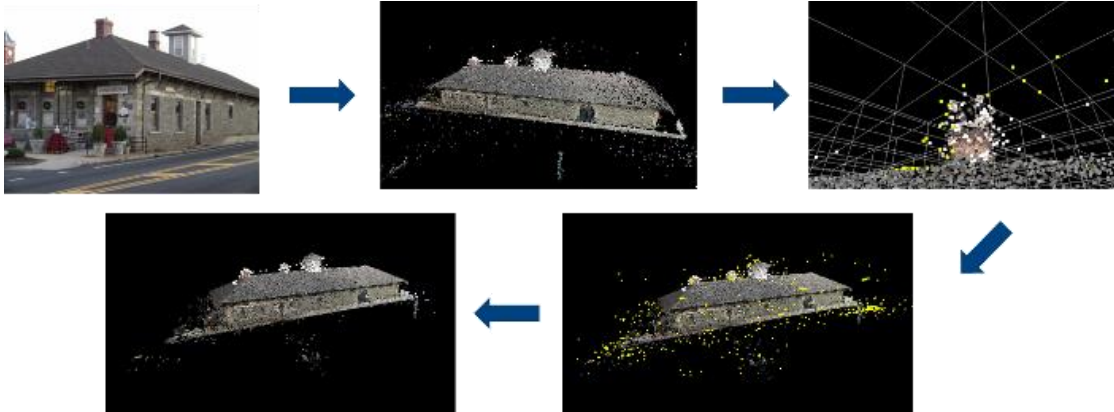
Table 5.4: Summary of the results obtained from evaluating the overall performance of the proposed method

Experimental setting	Indoor	Outdoor
Average number of measurements for each case study	107	281
Average error* (mm per meter)	1.1	2.3
Maximum error (mm per meter)	3.2	6.5
Standard Deviation	0.4	1.1

*error is measured based on the ratio of computed dimensions to actual dimensions per unit of length (meter)

5.3. Validating the proposed point cloud data cleaning algorithm:

To evaluate the performance of the three major algorithms for filling gaps/holes, 9 different case studies were selected and included mainly sharp edges and corners. Two major hole filling algorithms, as well as the proposed edge based hole filling algorithm are then implemented on the generated PCD to cover the existing gaps. Percentages of completeness and number of points before and after implementing the algorithms each algorithm are measured.



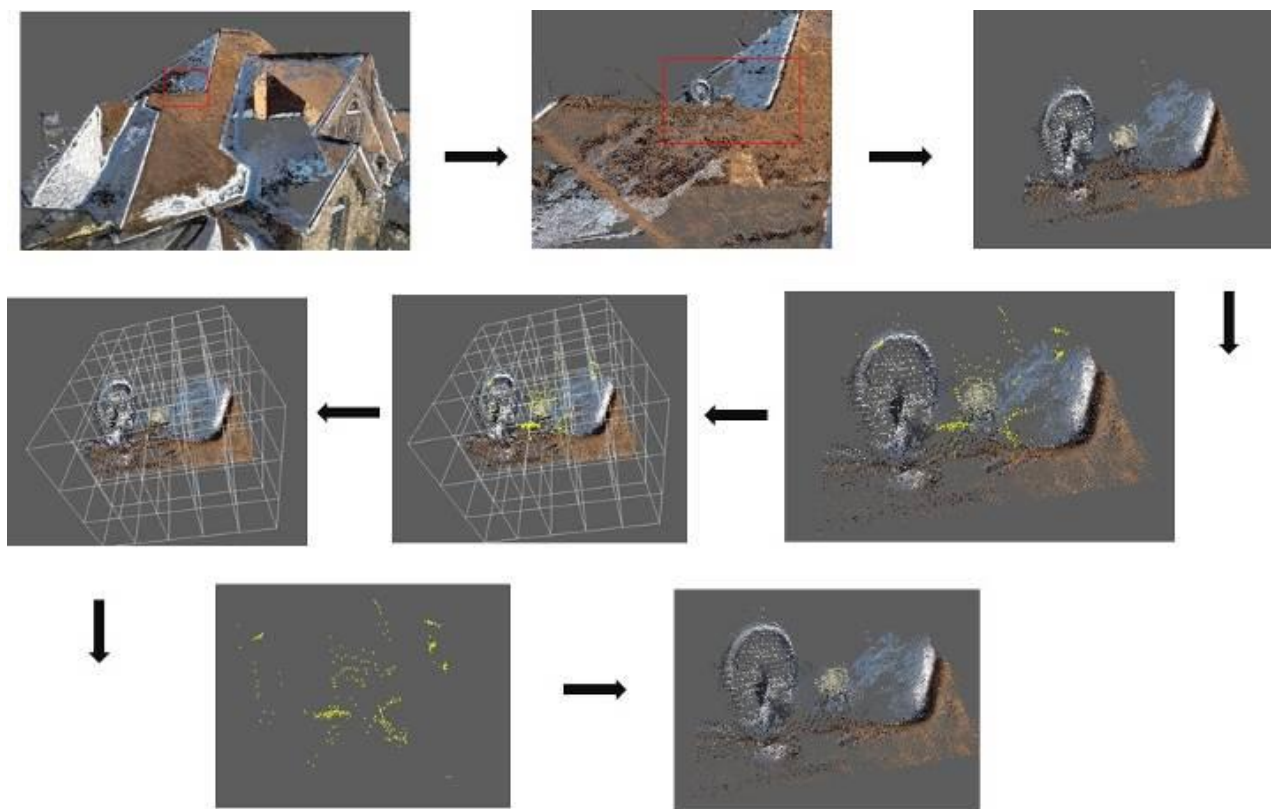


Figure 5.5: Different stages for detecting and removing outliers



Figure 5.6: Implementing the hole filling algorithm for one dataset as a case study

The summary of the comparison results are presented in table 5.5. As shown in the table, the proposed method of this research outperformed the existing methods in terms of successfully filling the gaps. In particular, the results are very promising for points which are located close to sharp edges and corners (Figures 5.7.A- 5.7.D). In order to achieve the maximum reliability, the operator needs to exactly measure the locations of points on hole/gap areas using a total station.

Table 5.5: Summary of comparison results for 5 different datasets.

Dataset #	Method	Number of points before implementing the algorithm	Number of points after implementing the algorithm	Completeness before implementing the algorithm (%)	Completeness after implementing the algorithm (%)
1	Our method	459'412	512'307	82.74	95.15

	MLS *	459'412	483'251	82.74	91.32
	VDM **	459'412	491'775	82.74	89.43
2	Our method	2'578'209	2'876'292	79.21	88.10
	MLS *	2'578'209	2'793'251	79.21	84.49
	VDM **	2'578'209	2'655'439	79.21	81.98
3	Our method	782'891	841'257	73.65	86.21
	MLS *	782'891	804'219	73.65	84.54
	VDM **	782'891	812'593	73.65	80.22
4	Our method	1'021'544	1'261'004	85.74	93.24
	MLS *	1'021'544	1'110'736	85.74	92.73
	VDM **	1'021'544	1'008'703	85.74	88.41
5	Our method	678'325	697'841	86.29	94.91
	MLS *	678'325	692'610	86.29	91.65
	VDM **	678'325	683'541	86.29	92.81

*MLS: Moving Least Square method

** VDM: Volumetric Diffusion Method

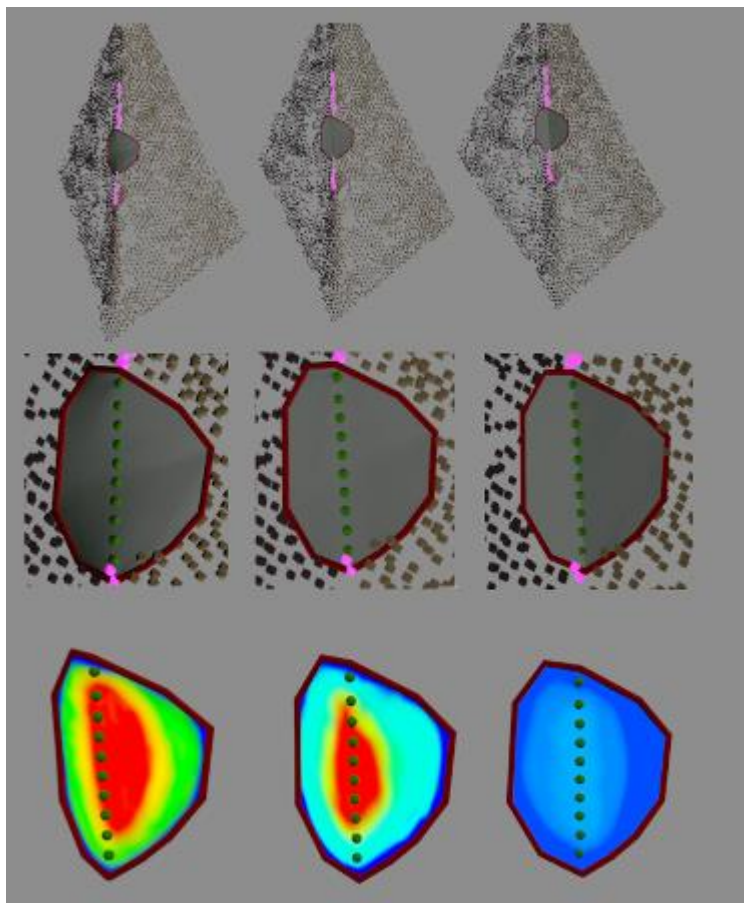
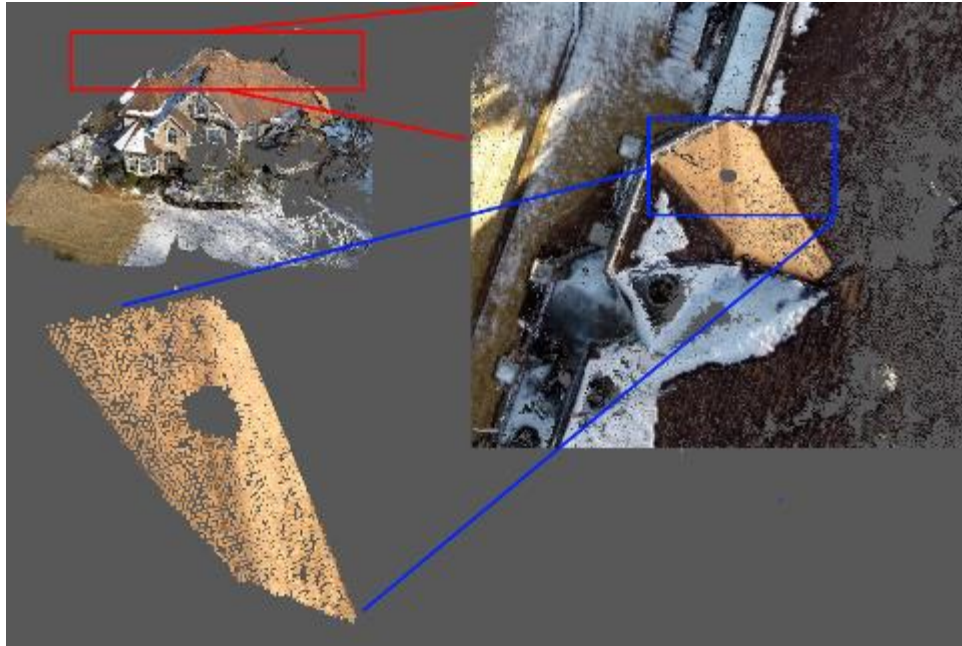


Figure 5.7.A: Different approaches for filling holes on surfaces of point cloud data
 MLS (left), VD (middle) and 3D edge based (right)

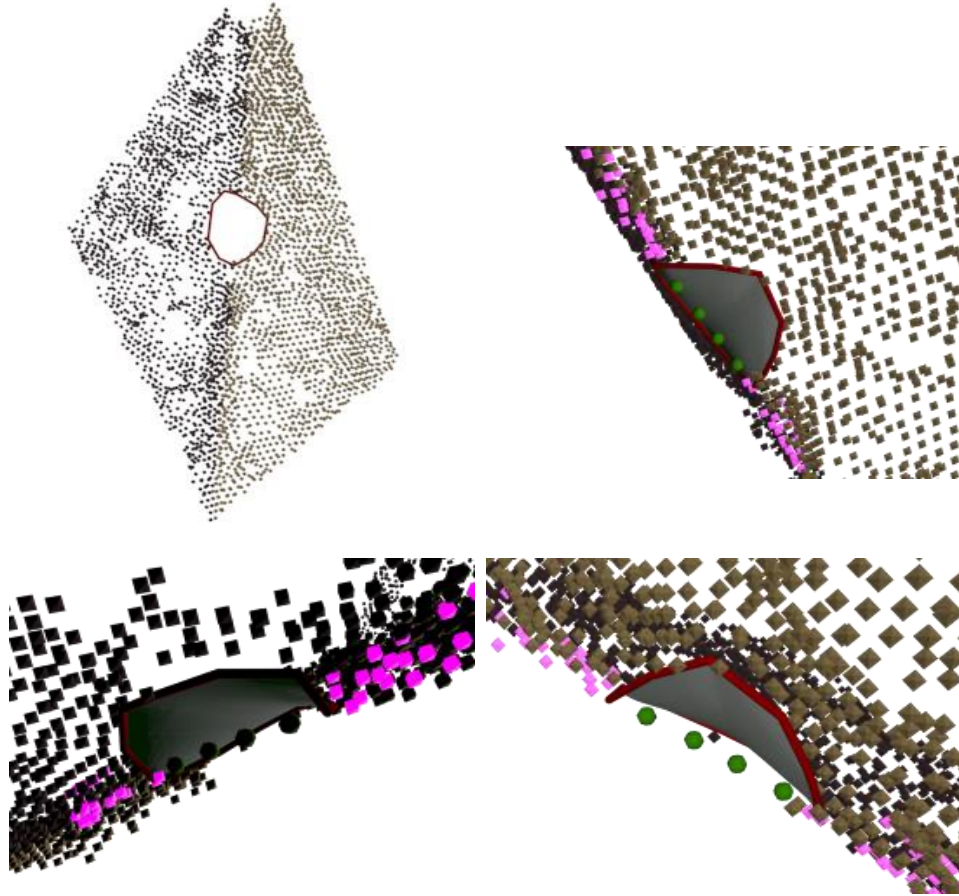


Figure 5.7.B: Results of filling the hole using MLS method

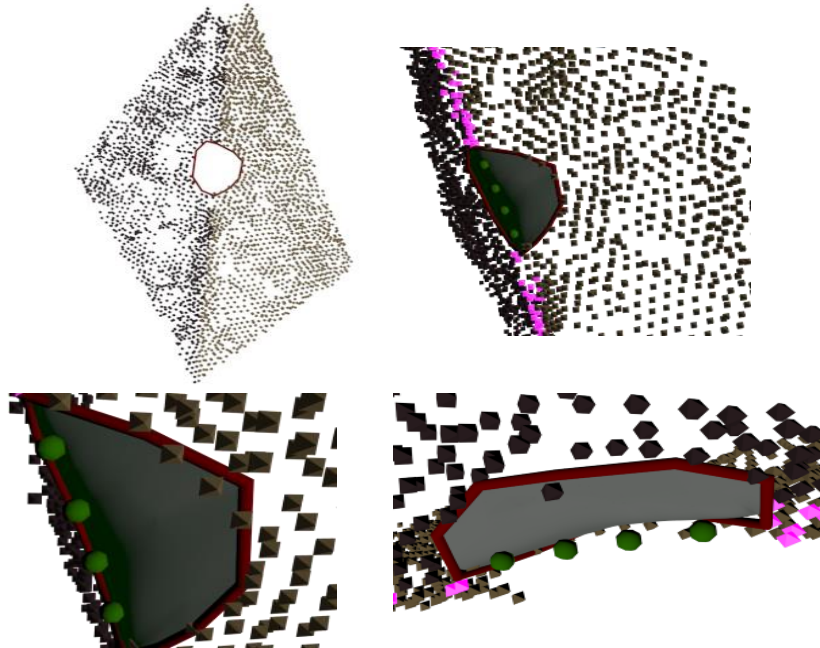


Figure 5.7.C: Results of filling the hole using Volumetric Diffusion method

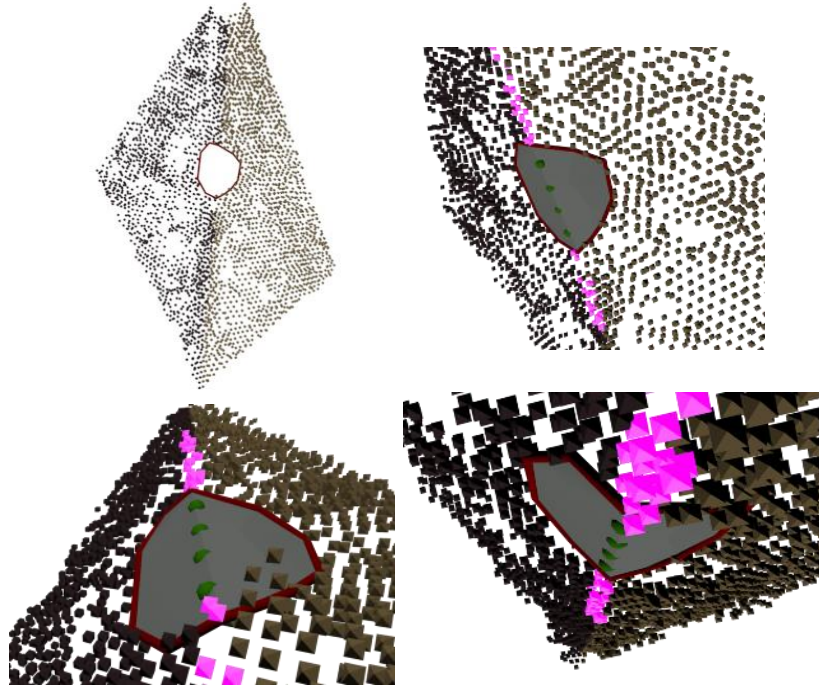


Figure 5.7.D: Results of filling the hole using 3D edge based method

5.4. Validating the improved videogrammetric pipeline for as-built documentation of built infrastructure

The last stage of validation is evaluating the performance of the improved videogrammetric pipeline as a whole. To this end, 4 indoor and 5 outdoor settings were selected. The selected case studies have been used for validating different stages of the research previously. In this stage, the collected video clips from these 9 case studies have been processed considering two scenarios.

The first scenario is implementing the three components of the improved prototype all together: key frame selection, automated calculation of absolute scale and point cloud data cleaning.

The second scenario is implementing the basic videogrammetric pipeline without the proposed components. Instead of the key frame selection algorithm, a uniform number of frames equal to the number of selected key frames in the first scenario were processed. Instead of automatically calculating the absolute scale, a laser measurer (for indoor settings) and a total

station (for outdoor cases) were used to measure one dimension of the scene and scale up the entire built infrastructure. Same as the previous validation stages, accuracy and completeness (uniform density) were considered as the evaluation metrics. More details about the case studies are presented here:

Outdoor case study#01: steel girder-box bridge:

This bridge is located close to the CRC (Campus Recreation Center) building at Georgia Tech. The bridge contains very high quality, smooth surfaces on concrete frames so it is an excellent case study for measurement purposes.



Figure 5.8: Outdoor case study #01: steel bridge with concrete columns

Outdoor case study#02: Concrete arch bridge:

This case study has been used for the comparison studies previously. Since it contains curvatures and arch surfaces, it can effectively be used for validating hole filling and plane recognition algorithms (Figure 5.9)



Figure 5.9: Outdoor case study #02: concrete arch bridge: point cloud data before improvement (middle) and after improvement (bottom)

Outdoor case study#03: Masonry wall in a construction site:

This masonry wall was part of an ongoing construction project on Georgia tech campus during spring 2012. Existence of planar and texture-full surfaces makes it a suitable case study for validating purposes.



Figure 5.10: Outdoor case study #03: masonry wall in a construction site

Outdoor case study #04: campus building:

There was a main reason for choosing this particular building as one of the case studies. Due to access limitations, there are limitations in covering the entire scene from different views;

thus, the generated point cloud contains several gaps and hole and makes it a good case for evaluation purposes (Figure 5.11)



Figure 5.11: Outdoor case study #04: campus building

Outdoor case study #05: Country house

The selected case study contains several planar surfaces, sharp edges and corners (Figure 5.12).



Figure 5.12: Outdoor case study #05: country building

All case studies were videotaped and the generated video clips were processed following the mentioned procedure. A summary of obtained results plus useful information regarding number of generated points, number of conducted measurements and numbers of surfaces used for validation purposes are summarized in Table 5.6:

Table 5.6: Summary of evaluation results for the 5 different outdoor datasets

Case study #	1	2	3	4	5
Number of generated points before improvement	1'765'432	1'234'107	2'345'453	2'650'421	3'106'876

Number of generated points after improvement	1'893'659	1'521'447	2'398'054	2'993'141	3'235'719
Average error (mm) before improvement	27	17	14	35	38
Average error (mm) after improvement	21	14	9	30	29
Number of conducted measurements	48	21	17	39	52
Number of surfaces used for validation purposes	7	5	2	12	16
Percentage of truly detected and reconstructed holes	73.2	79.7	81.4	73.2	69.4
Completeness ratio (%) before improvement	78.2	81.4	75.5	80.2	72.3
Completeness ratio (%) after improvement	83.5	86.3	82.8	79.3	75.5
Number of detected outlier	54'321	13'345	9,651	31'056	27'431

The indoor case studies contain a kitchen, a bedroom, a bathroom and an office corridor which are all common spaces for facility managers and interior designers:

Indoor case study #01: Kitchen



Figure 5.13: Indoor case study #01: kitchen

Indoor case study #02: Bathroom



Figure 5.14: Indoor case study #02: bathroom

Indoor case study #03: Bedroom



Figure 5.15: Indoor case study #03: bedroom

Indoor case study #04: Office corridor

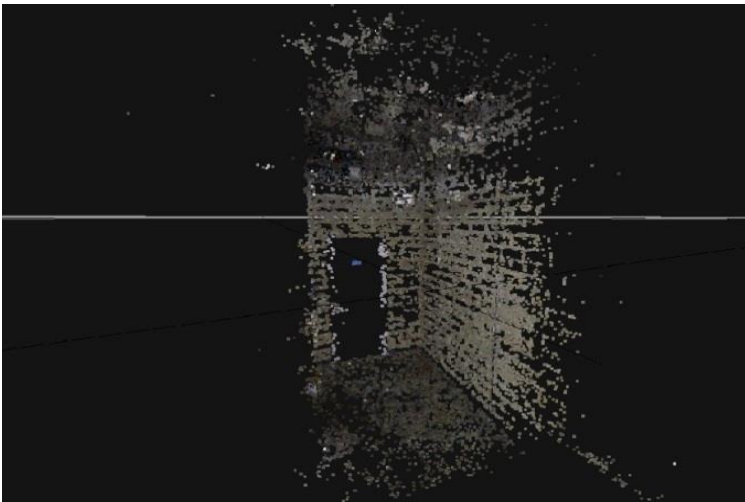


Figure 5.16: Indoor case study #04: office corridor

Following the same procedure as outdoor cases, obtained results are summarized in Table

5.7:

Table 5.7: Summary of evaluation results for the 4 different indoor datasets.

Case study #	1	2	3	4
Number of generated points before improvement	1'765'432	1'234'107	2'345'453	2'650'421
Number of generated points after improvement	1'893'659	1'521'447	2'398'054	2'993'141
Average error (mm) before improvement	27	17	14	35
Average error (mm) after improvement	21	14	9	30
Number of conducted	48	21	17	39

measurements				
Number of surfaces used for validation purposes	7	5	2	12
Percentage of truly detected and reconstructed holes	73.2	79.7	81.4	73.2
Completeness ratio (%) before improvement	78.2	81.4	75.5	80.2
Completeness ratio (%) after improvement	83.5	86.3	82.8	79.3
Number of detected outlier	54'321	13'345	9,651	31'056

In order to identify potential applications for the improved videogrammetric pipeline, minimum required accuracy of different AEC/FM applications are presented in the following table. From observing accuracy ranges, it is concluded that the proposed videogrammetric pipeline has the applicability potential in several areas including insurance industry, quantity take off, layout design and interior design.

Table 5.8: Required levels of accuracy for multiple applications in AEC/FM domain (Dai et.al, 2013)

Application:	Required accuracy	Does videogrammetry meet the accuracy requirements?
Structural quality control	<1mm	no
Surveying	0.5-2 mm	no
fabrication	1-5 mm	no
Quantity take off	5-20 mm	yes
Rough estimations for insurance companies	10-50 mm	yes
Job site layout design	20-100 mm	yes
Home and offices interior design	20-50 mm	yes

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1. Conclusion and Remarks

Monocular video/photogrammetry is an easy to use, cost effective technique for 3D reconstruction of civil infrastructure. Unlike applications in computer vision domain, where in most cases results are demonstrated only for visual purposes, civil engineers need to extract accurate dimensions and measurements from the point clouds. It is well known that results obtain from monocular video/photogrammetry are up to scale and to calculate absolute scale, at least one dimension of the scene should be known. To address this issue, in this research, a practical solution based on reconstructing predefined simple 3D objects is presented and validated.

As the second part, this research presented the work of creating a novel method for extracting the high quality, informative frames from a video stream. The resulting frames can be fed into the videogrammetric pipeline to effectively generate dense point clouds of civil infrastructure. The proposed algorithm automates the processes of removing frames with blur effects and selecting a number of frames in a fashion that computational efficiency is achieved and common degeneracy cases are minimized. The experiments results reveal that the proposed key frame selection algorithm eclipses the existing methods in higher successful rate of the 3D reconstruction while maintaining the best reprojection accuracy. Moreover, this research validated the proposed videogrammetric solution in terms of its accuracy and completeness of the resulting 3D point clouds under various settings such as different types of camera, resolution configurations, and data collection distances.

By applying the proposed key frame selection algorithm and the videogrammetric pipeline, up to 1.82 cm accuracy as well as 81% completeness is achievable. Proposing an innovative algorithm for filling the holes existing on surfaces of the point clouds is the other

aspect of this research. This innovative algorithm is mainly based on the geometry and visual properties of specific civil engineering scenes. The achieved level of accuracy is sufficient for several applications in the AEC/FM domain including rough estimations for insurance industries, quantity take-off, interior design of homes and offices and job site layout design of construction job sites.

6.2. Plans for Future Research

As explained before, 3D reconstruction of built infrastructure and computing point cloud data is the first stage of the entire as built modeling documentation procedure. The next stage of this research would be processing those point cloud data and extracting informative, object oriented shapes.

6.3. Achieved Contribution

This research could have several contributions both in the civil engineering and computer vision communities. First, the proposed algorithm for optimized selection of key video frames significantly improves the accuracy and processing time for generating point clouds. Second, the proposed method for automated extraction of absolute scale from generated point clouds enables users to accurately extract Euclidian distances between different interesting points. It allows the user to measure different dimensions of the scene, which is an essential part of every 3D as-built documentation technology. Finally, the proposed method for filling the gaps and holes on the surfaces of point clouds is able to provide more realistic results by maintaining sharp edges and corners.

REFERENCES

- Ahmed, M.T., Dailey, M.N., Landabaso, J.L., & Herrero, N., 2010. Robust key frame extraction for 3D reconstruction from video streams. In proceedings of the VISAPP.
- Amenta N., Bern, M., & Kamvyselis, M., 1998. A new voronoi-based surface reconstruction algorithm, In proceedings of the SIGGRAPH'98, Orlando, USA.
- Anati, R., Scaramuzza, D., Derpanis, K., & Daniilidis, K., 2012. Robot localization using soft object detection. In proceedings of the IEEE International Conference on Robotics and Automation, Minnesota, USA.
- Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., & Taubin, G., 1999. The Ball-Pivoting Algorithm for surface reconstruction. IEEE Transactions on Visualization and Computer Graphics, 5(4), pp.349-359.
- Botterill, T., Mills, S., & Green, R., 2012. Correcting scale drift by object recognition in single-camera SLAM. IEEE Transactions on Systems, Man, and Cybernetics, 43 (6), pp. 1767-1780.
- Brilakis, I., Fathi, H., & Rashidi, A., 2011. Progressive 3D reconstruction of infrastructure with videogrammetry. Automation in Construction, 20(7), pp. 884-895.
- Cai, J.F., Ji, H., Liu, C., & Shen, Z., 2009. Blind motion deblurring using multiple images. Journal of Computational Physics, 228(14), pp. 5057-5071.
- Carozza, L., Tingdahl, D., Bosché, F., & Van Gool, L., 2012. Markerless vision-based augmented reality for urban planning. Computer-Aided Civil and Infrastructure Engineering, Early View (Online Version of Record published before inclusion in an issue).
- Castellanos, J.A., Montiel, J.M.M., Neira, J., & Tardós, J.D., 2000. Sensor influence in the performance of simultaneous mobile robot localization and map building. Proc., In proceedings of the 6th International Symposium on Experimental Robotics, Sydney, Australia.
- Chen, M.J., & Bovik, A.C., 2009. No reference image blur assessment using multiscale gradient. In proceedings of the 1st International Workshop on Quality of Multimedia Experience (QoMEX).

- Chi, S., & Caldas, C.H., 2011. Automated object identification using optical video cameras on construction sites. *Computer-Aided Civil and Infrastructure Engineering*, 26(5), pp. 368-380.
- Cho, S., & Lee, S., 2009. Fast motion deblurring. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia*, 28(5).
- Cho, Y. and Gai, M. 2014. Projection-Recognition-Projection (PRP) Method for Rapid Object Recognition and Registration from a 3D Point Cloud. *ASCE Journal of Computing in Civil Engineering*, doi: 10.1061/(ASCE) CP.1943-5487.0000332 (in press)
- Chung, Y., Wang, J., Bailey, R., Chen, S., & Chang, S., 2004. A non-parametric blur measure based on edge analysis for image processing applications. In proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, Singapore.
- Coaker, L.H., 2009. Reflector-less total station measurements and their accuracy, precision and reliability. Dissertation, USQ Project, University of Southern Queensland, Australia.
- Crete, F., Dolmiere, T., Ladret, P., & Nicolas, M., 2007. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In proceedings of the SPIE, eds. B. E. Rogowitz, T. N. Pappas, and S. J. Daly, 6492, article id. 64920I.
- Dai, F., & Lu, M., 2010. Assessing the accuracy of applying photogrammetry to take geometric measurements on building products. *Journal of Construction Engineering and Management*, American Society of Civil Engineers, (136)2, pp. 242-250.
- Dai, F., Rashidi, A., Brilakis, I., & Vela, P., 2013. Comparison of image-based and time-of-flight-based technologies for three-dimensional reconstruction of infrastructure. *ASCE Journal of Construction Engineering and Management*, 139(1), pp. 69-79.
- Davis, J., Marschner, S., Garr, M. & Levoy, M., 2002. Filling holes in complex surfaces using volumetric diffusion. In proceedings of the first international Symposium on 3D Data Processing, Visualization, Transmission, pp. 428-438.
- Dias, J.M.S., Bastos, R., Correia, J., & Vicente, R., 2006. Semi-automatic 3D reconstruction of urban areas using epipolar geometry and template matching. *Computer-Aided Civil and Infrastructure Engineering*, 21(7), pp. 466-485.

- Edelsbrunner, H. & Muehe, E.P., 1994. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1), pp.43-72.
- Eudes, A., Lhuillier, M., Naudet, S., & Dhome, M., 2010. Fast odometry integration in local bundle adjustment-based visual SLAM. In proceedings of the 20th International Conference of Pattern Recognition (ICPR), Istanbul, Turkey.
- Fathi, H., & Brilakis, I., 2011. Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features. *Journal of Advanced Engineering Informatics*, 25(4), pp. 760-770.
- Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., & Freeman, W.T., 2006. Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 25(3), pp. 787–794.
- Furukawa, Y., & Ponce, J., 2010. Accurate, dense, and robust multi-view Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), pp. 1362-1376.
- Ganci, G., & Brown, J., 2008. Developments in Non-Contact Measurement Using Videogrammetry. Boeing large scale metrology seminar, Melbourne, FL.
- Gibson, S., Cook, J., Howard, T., Hubbard, R., & Oram, D., 2002. Accurate camera calibration for off-line, video-based augmented reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, Darmstadt, Germany.
- Golparvar-Fard, M., Peña-Mora, F., & Savarese, S., 2009. Application of D4AR – A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of Information Technology in Construction*, 14(Special Issue), pp. 129-153.
- Golparvar-Fard, M., Peña-Mora, F., & Savarese, S., 2012. Automated progress monitoring using unordered daily construction photographs and IFC-based building information models. *Journal of Computing in Civil Engineering*, [http://dx.doi.org/10.1061/\(ASCE\)CP.1943-5487.0000205](http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000205).
- Gopi, M. & Krishnan S., 2000. A fast and efficient projection based approach for surface reconstruction. *International Journal of High Performance Computer Graphics, Multimedia and Visualization*, 1(1), pp. 1-12.

- Greenwood, J., 1999. Large component deformation studies using videogrammetry. In proceedings of the 6th International Workshop on Accelerator Alignment (IWAA 99), Grenoble, France.
- Gutierrez-Gomez, D., Puig, L., & Guerrero, J.J., 2012. Full scaled 3D visual odometry from a single wearable omnidirectional camera. International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal.
- Hansen, M.H., Hurwitz, W.N., & Madow, W.G., 1953. Simple random sampling. Sample Survey Methods and Theory, Volume I Methods and Applications, John Wiley & Sons, New York, Chap. 4, pp. 126-129.
- Hartley, R., & Zisserman, A., 2004. Multiple view geometry. Cambridge, UK: Cambridge University Press.
- Jahanshahi, M.R., Masri, S.F., Padgett, C.W., & Sukhatme, G.S., 2013. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. Machine Vision and Applications, 24(2), pp. 227-241.
- Joshi, N., Kang, S.B., Lawrence Zitnick, C., & Szeliski, R., 2010. Image Deblurring using inertial measurement sensors. ACM Transactions on Graphics, 29(4) pp. 29:30:1-30:9.
- Jung, M.J., Myung, H., Hong, S.G., Park, D.R., Lee, H.K., & Bang, S.W. 2004. Structured light 2D range finder for simultaneous localization and map-building (SLAM) in home environments. In proceedings of the IEEE 4th International Symposium on Micro-NanoMechatronics and Human Science.
- Kien, D.T., 2005. A review of 3D reconstruction from video sequences. Intelligent sensory information systems, Department of Computer Science, University of Amsterdam, Netherlands.
- Kim, H., & Kano, N., 2008. Comparison of construction photograph and VR image in construction progress. Automation in Construction, 17(2), pp. 137-143.
- Klein, L., Li, N. & Becerik-Gerber, B., 2011. Imaged-based verification of as-built documentation of operational buildings. Automation in Construction, (21), pp. 161-171.
- Kneip, L., Martinelli, A., Weiss, S., Scaramuzza, D., & Siegwart, R., 2011. A closed-form solution for absolute scale velocity determination combining inertial measurements and a

- single feature correspondence. In proceedings of the IEEE International Conference on Robotics and Automation (ISRA), Shanghai, China.
- Kuhl, A., Wöhler, C., Krüger, L., d'Angelo, P., & Groß, H.M., 2006. Monocular 3D scene reconstruction at absolute scales by combination of geometric and real-aperture methods. *Lecture Notes in Computer Science*, 4174, pp. 607-616.
- Li, H., Huang, Y., Ou, J., & Bao, Y., 2011. Fractal dimension-based damage detection method for beams with a uniform cross-section. *Computer-Aided Civil and Infrastructure Engineering*, 26(3), pp.190-206.
- Lourakis, M.A., Argyros, A.A., 2009. SBA: A Software Package for Generic Sparse Bundle Adjustment, *ACM Transactions on Mathematical Software (TOMS)*, 36(1), pp. 2:1-2:30.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), pp. 91-110.
- Marziliano, P., Dufaux, F., Winkler, S., & Ebrahimi, T., 2002. A no-reference perceptual blur metric. In proceedings of the International Conference on Image Processing, Rochester, NY.
- Memon, Z.A., Majid, M.Z.A., & Mustaffar, M., 2005. An automatic project progress monitoring model by integrating AutoCAD and digital photos. In proceedings of the International Conference on Computing in Civil Engineering, L. Soibelman and F. Peña-Mora, eds., American Society of Civil Engineers (ASCE), Cancun, Mexico.
- Mencel R., 1995. A graph-based approach to surface reconstruction, *Proceedings of EUROGRAPHICS'95. Computer Graphics Forum*, 14(3), pp. 445-456.
- Morse, B.S., Yoo, T.S., Chen, D.T., Rheingans, P., & Subramanian, K.R., 2001. Interpolating implicit surfaces from scattered surface data using compactly supported radial basis functions. In proceedings of the International Conference on Shape Modeling & Applications, IEEE Computer Society Press, Genova, Italy.
- Nistér, D., 2004. Automatic passive recovery of 3D from images and video. Invited Paper, In proceedings of the Second International Symposium on 3D Data Processing, Visualization & Transmission (3DPVT04), Thessaloniki, Greece.

- Nistér, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(6), pp. 756-770.
- Nützi, G., Weiss, S., Scaramuzza, D., & Siegwart, R., 2011. Fusion of IMU and vision for absolute scale estimation in monocular SLAM. *Journal of Intelligent and Robotic Systems (JIRS)*, 61(1-4), pp. 287-299.
- Pappa, R.S., Black, J.T., Blandino, J.R., Jones, T.W., Daney, P.M., & Dorrington, A.A., 2003. Dot-Projection Photogrammetry and Videogrammetry of Gossamer Space Structures. In proceedings of the 21st International Modal Analysis Conference (IMAC), Kissimmee, FL, USA.
- Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., & Koch, R., 2004. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3), pp. 207-232.
- Pollefeys, M., et al., 2008. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 78(2-3), pp. 143-167.
- Quiñones-Rozo, C.A., Hashash, Y.M.A., & Liu, L.Y., 2008. Digital image reasoning for tracking excavation activities. *Automation in Construction*, 17(5), pp. 608-622.
- Rashidi, A., Dai, F., Brilakis, I., & Vela, P., 2011. Comparison of camera motion estimation methods for 3D reconstruction of infrastructure. In proceedings of the ASCE International Workshop on Computing in Civil Engineering, Miami, FL, USA.
- Rashidi, A., Fathi, H., & Brilakis, I., 2011. Innovative stereo vision-based approach to generate dense depth map of transportation infrastructure. *Transportation Research Record, Journal of the Transportation Research Board*, 2215, pp. 93-99.
- Rashidi, A., Dai, F., Brilakis, I., & Vela, P., 2013. Optimized selection of key frames for monocular videogrammetric surveying of civil infrastructure. *Journal of Advanced Engineering Informatics*, 27(2), pp. 270-282.
- Scaramuzza, D. Fraundorfer, F. Pollefeys, M., & Siegwart, R., 2009. Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints. In proceedings of the IEEE International conference on computer vision, Kyoto, Japan.

- Scarloff, S., & Pentland, A., 1991. Generalized implicit functions for computer graphics. In proceedings of the SIGGRAPH'91.
- Scharstein, D., & Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1), pp. 7-42.
- Seo, J., Kim, S., Jho, C., Hong, H., 2003. 3D estimation and keyframe selection for match move. In proceedings of the International Technical Conference on Circuits Systems, Computers and Communications (ITC-CSCC).
- Seo, Y.H., Kim, S.H., Doo, K.S., & Choi, J.S., 2008. Optimal keyframe selection algorithm for three-dimensional reconstruction in uncalibrated multiple images. *Journal of the Society of Photo-Optical Instrumentation Engineers*, 47(5), pp. 53201-53400.
- Shan, Q., Jia, J., & Agarwala, A., 2008. High-quality motion deblurring from a single image. *ACM Transactions on Graphics*, 27(3), pp. 73:1–73:10.
- Sheikh, R.H., Bovik, C.A., & de Veciana, G., An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12), pp. 2117-2128.
- Snavely, N., Seitz, S.M., & Szeliski, R., 2008. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, 80(2), pp. 189-210.
- Tang, P., Huber, D., Akinci, B., Lipman, R., and Lytle, A. 2010. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction*, 19(7), 14.
- Teizer, J. and Vela, P.A. 2009. Personnel Tracking on Construction Sites using Video Cameras. *Advanced Engineering Informatics, Special Issue, Elsevier*, 23(4), pp. 452-462.
- Teizer, J. 2008. 3D Range Image Sensing for Active Safety in Construction. *Journal of Information Technology in Construction, Sensors in Construction and Infrastructure Management*, 13 (Special Issue), pp. 103-117.
- Terzopoulos, D., & Metaxas, D., 1991. Dynamic 3D models with local and global deformations: deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7), pp.703-714.

- Thormahlen, T., Broszio, H., & Weissenfeld, A., 2004. Keyframe selection for camera motion and structure estimation from multiple views Thorsten, In proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic.
- Torr, P.H.S, Fitzgibbon, A.W., & Zisserman, A., 1998. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In proceedings of the 6th international conference on computer vision, Bombay, India.
- Tribou, M., 2009. Recovering scale in relative pose and target model estimation using monocular vision. M.S. thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Varadarajan, S., & Karam, L.J., 2008. An improved perception-based no-reference objective image sharpness metric using iterative edge refinement. In proceedings of the 15th IEEE International Conference on Image Processing, San Diego, CA, USA.
- Wang J. & Oliveira, M.M., 2002. Improved scene reconstruction from range images. In proceedings of the EUROGRAPHICS'02.
- Wang, J., & Oliveira, M.M., 2007. Filling holes on locally smooth surfaces reconstructed from point clouds. *Image and Vision Computing*, 25(1), pp. 103-113.
- Zhang, Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations. In proceedings of the 7th IEEE International Conference on Computer Vision, Kerkrya, Greece.
- Zhang, G., & Wang, Y., 2013. Optimizing coordinated ramp metering: a preemptive hierarchical control approach. *Computer-Aided Civil and Infrastructure Engineering*, 28(1), pp. 22–37.
- Zhu, Z., & Brilakis, I., 2009. Comparison of optical-sensor-based spatial data collection techniques for civil infrastructure modeling. *Journal of Computing in Civil Engineering*, 23(3), pp. 170-177