Large-scale Range Data Acquisition and Segmentation

Reyhaneh Hesami

Submitted in total fulfilment of the requirements for the degree of Doctor of Philosophy

Faculty of Engineering and Industrial Sciences
Swinburne University of Technology
2009
Abstract

The research reported in this thesis aims to devise a cost-effective, robust technology and technique for accurately measuring and segmenting geometric details embedded in the exterior surfaces of large buildings. A diverse range of applications become viable or significantly enhanced by capturing the accurate geometric data of significant buildings and extracting the fine details of such data. Motivations for this thesis stem from the facts that commercially available large-scale data acquisition systems are expensive and existing algorithms for 3D data processing are yet to address the processing complexity associated with the 3D segmentation of large outdoor objects.

The contributions of this thesis are threefold. First, a low-cost versatile large-scale rangescanner, capable of capturing range data up to 300 metres, is designed and implemented. An innovative method for system calibration and the data fusion of the intensity and range measurements has been developed. A number of experiments are also conducted to evaluate the performance of the proposed rangescanner system and the data fusion technique. The range data obtained by the rangescanner system has been verified using two methods of verification — application checking and equivalence checking. Example results of the verification, presented in this thesis, shows that the laser rangescanner device is capable of providing adequate accuracy and resolution for large-scale civil application with minimum complexity and cost. The design provides portability, flexibility and ease of operation. Secondly, problems associated with range data acquisition and processing of large building exteriors are studied. Key challenges of processing the range data of large buildings, including significant disparities in the size and depth and the existence of substantial construction error in historical buildings, have been identified and their effects for the segmentation task are examined. Thirdly, a computationally effective and robust segmentation technique, capable of extracting geometric details of large building
exteriors, is developed. The segmentation algorithm, titled *Hierarchical Robust Segmentation (HRS)*, uses a high breakdown, robust estimator in a hierarchical coarse-to-fine approach. This algorithm is then tested on several range data sets obtained by different laser rangescanners. The experimental results show that the proposed algorithm overcomes most of the current problems of large-scale data segmentation presented in this thesis by extracting both coarse and fine details from range data of large building exteriors in a relatively short period of time.
Acknowledgement

"If I have seen further, it is by standing on the shoulders of giants."

-Isaac Newton

First and foremost, I would like to express my deepest gratitude and appreciation to my supervisor, Associate Professor Alireza Bab-Hadiashar for his mentoring and friendship throughout my candidature. He introduced computer vision to me and without his constant support I could not have finished this thesis. I thank my first co-supervisor Dr. Hanes Van Der Walt who was always cheerful and encouraging. I also thank Dr. Reza Hosseinnezhad who stepped in as my co-supervisor midway through this project. We had fruitful discussions through the last stage of my research.

I would also like to thank Professor Ray Jarvis, Professor David Suter and his research team at Monash University, Melbourne, Australia, for providing access to their Riegl laser rangescanner and several sets of range data captured from Monash University buildings. Thank you is also due to Professor Peter Allen and Mr Paul Blaer of Columbia University, USA, for providing me with range data of Notre Dame Church in France. I thank Mr Babak Majidi for his help in data acquisition of Melbourne Exhibition Building.

There are also technical and support staff members at Swinburne University who have supported this study, particularly Mr Walter Chetcuti, Mr Warren Gooch and Ms Nancy Moncrieff. I also thank Ms Dionne C Eagleson for her proofreading services.

My sincere thanks and appreciations go to my teachers, my friends and family. They made me what I am. Specially, to my devoting parents who have been my source of inspiration and strength, I am forever indebted for their countless support and sacrifices. I would like to particularly thank my mum, my first teacher, who started teaching me maths and alphabets
when I was as young as three. I thank my younger brothers Ali, Ehsan and Erfan for their encouragement.

When one takes on a project of this size, it is going to affect those people who are near and dear. I owe the completion of this dissertation to the support, camaraderie, and encouragement of my husband Dr. Payam Ghadirian over the years of my study. To Pouya, my son, who has been grown up with this project, for his patient and understanding as a very young boy, when mum was busy working late on her PhD!
Dedicated to:

My parents, Pouya and Payam
Declaration

This dissertation is submitted to Swinburne University of Technology in fulfilment of the requirements for the degree of Doctor of Philosophy. I hereby declare that this dissertation is entirely my own work and contains no material previously published or written by another person, except where otherwise mentioned. Parts of the work described in Chapter three, four and five have previously appeared or currently under review in the following conference or journal papers:


**Hesami R.**, Bab Hadiashar A., Gheissari N., “*Large Object Range Data Acquisition, Fusion and Segmentation*”, Digital Image Computing Techniques and Application, DICTA’05, December 6-8, Cairns, Australia

Reyhaneh Hesami
2009
Table of Content

Abstract........................................................................................................................i
Acknowledgement ..................................................................................................... iii
Declaration................................................................................................................. vi
Table of Content....................................................................................................... vii
List of Figures............................................................................................................ xi
List of Tables ......................................................................................................... xviii
Chapter 1 Introduction......................................................................................... 1
  1.1 Background and motivation................................................................. 1
  1.2 What particularly is a complicated problem? ................................. 4
  1.3 Aim ........................................................................................................... 6
  1.4 Methodology ........................................................................................... 7
  1.5 Contributions of the thesis ................................................................. 8
  1.6 Overview of the study ....................................................................... 9
Chapter 2 3D Measurement ........................................................................ 11
  2.1 Introduction............................................................................................ 11
  2.2 3D measurement systems............................................................... 12
     2.2.1 Triangulation based rangescanner systems................................. 13
         2.2.1.1 Passive triangulation (stereovision) ........................................... 14
         2.2.1.2 Active triangulation............................................................... 15
2.2.2 Time-of-flight based rangescanner systems .................................................. 16
   2.2.2.1 Non-scanning 3D TOF cameras ............................................................ 19
   2.2.2.2 Scanning 3D TOF cameras (3D laser scanners) .................................. 20
   2.2.3 Triangulation vs. laser TOF based rangescanners .................................. 24

2.3 Range and intensity data integration ............................................................... 30
   2.3.1 Feature-based range and intensity data fusion ....................................... 31
   2.3.2 System-based range and intensity data fusion ....................................... 35

2.4 Chapter review ............................................................................................... 37

Chapter 3 Range Image Segmentation .................................................................. 39

3.1 Introduction ..................................................................................................... 39

3.2 Range segmentation ....................................................................................... 40

3.3 Taxonomy of range segmentation algorithms ............................................... 43
   3.3.1 Edge detection based range segmentation algorithms ........................... 44
   3.3.2 Region growing based range segmentation algorithms .......................... 45
   3.3.3 Parametric fitting-based range segmentation algorithm ....................... 46
      3.3.3.1 Modified Selective Statistical Estimator (MSSE) ............................. 51
   3.3.4 Combined methods .................................................................................. 55

3.4 Range segmentation of large man-made objects .......................................... 56

3.5 Chapter review ............................................................................................... 60

Chapter 4 Laser Rangescanner System ............................................................... 62

4.1 Introduction ..................................................................................................... 62

4.2 Laser rangescanner system design ............................................................... 62
   4.2.1 Hardware architecture ............................................................................ 63
   4.2.2 Software architecture ............................................................................ 66
   4.2.3 Automatic system calibration ................................................................. 67
   4.2.4 Range and intensity data fusion .............................................................. 71
      4.2.4.1 Camera models .................................................................................. 71
      4.2.4.2 Camera calibration .......................................................................... 76
      4.2.4.3 System calibration verification for fusion ....................................... 79
4.3 Laser rangescanner system experimental results................................. 80

4.4 Laser rangescanner system characterisation ........................................ 85
  4.4.1 Rangefinder errors................................................................................ 85
    4.4.1.1 Temperature effect on the laser hardware system.............................. 86
    4.4.1.2 Surrounding temperature effect........................................................ 86
    4.4.1.3 Range accuracy at different distances................................................. 88
    4.4.1.4 Mixed Pixel effects ............................................................................ 90
    4.4.1.5 Target reflectance properties.............................................................. 92
  4.4.2 PTU Errors ............................................................................................. 93

4.5 Verification of the functionality of the 3D scanner system ................. 95
  4.5.1 Application checking verification of 3D measurement system ............. 96
  4.5.2 Equivalence checking verification of 3D measurement system ............ 100

4.6 Chapter Review ....................................................................................... 105

Chapter 5 Range Data of Large Building Exteriors ............................... 109

5.1 Introduction............................................................................................ 109

5.2 Characteristics of range data of building exteriors ............................ 109
  5.2.1 Existence of moving objects ................................................................. 110
  5.2.2 Disparity in size ................................................................................... 111
  5.2.3 Existence of very fine details ............................................................... 114
  5.2.4 High uncertainty due to the construction errors............................ 116

5.3 Chapter review ....................................................................................... 118

Chapter 6 Hierarchical Robust Range Segmentation ............................ 120

6.1 Introduction............................................................................................ 120

6.2 From robust estimation to robust segmentation .................................. 121

6.3 Hierarchical Robust Segmentation scheme ....................................... 122
  6.3.1 How does the hierarchical scheme reduce computation costs? ........ 126

6.4 Experimental results............................................................................... 127
  6.4.1 Synthetic data....................................................................................... 127
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.2 Real data</td>
<td>132</td>
</tr>
<tr>
<td>6.5 Chapter review</td>
<td>139</td>
</tr>
<tr>
<td>Chapter 7 Conclusions and Future Directions</td>
<td>140</td>
</tr>
<tr>
<td>7.1 Conclusions</td>
<td>140</td>
</tr>
<tr>
<td>7.2 Future directions</td>
<td>142</td>
</tr>
<tr>
<td>Bibliography</td>
<td>144</td>
</tr>
<tr>
<td>Appendices</td>
<td>166</td>
</tr>
</tbody>
</table>
List of Figures

Chapter 1
Figure 1.1: A typical computer vision system ................................................................. 2

Figure 1.2: Royal Melbourne Exhibition Building, Melbourne, Australia: cars, passengers and vegetation are obstacles that can not be avoided at the time of data collection. .......... 4

Figure 1.3: Examples of disparity in the size of various features. Different colours show different features. ..................................................................................................................... 5

Figure 1.4: Effect of a construction error on the range segmentation of a large building: (a) Walls of the front side of the Shrine of Remembrance building (coloured in blue) seem co-planar. (b) Accurate ground truth measurement prove they are two separate planar surfaces. It is shown in the hypothetical top view of the building. ........................................... 6

Figure 1.5: Steps of the research ...................................................................................... 8

Chapter 2
Figure 2.1: A classification of 3D measurement techniques based on [16] and [17]. .......... 12

Figure 2.2: Passive triangulation (Stereovision) ............................................................... 14

Figure 2.3: Active triangulation ....................................................................................... 15

Figure 2.4: A typical active triangulation–based rangescanner ....................................... 16
Figure 2.5: Non-scanning 3D TOF camera: (a) Principle (Adapted from [33]) (b) Commercialised, manufactured by MESA imaging (http://www.mesa-imaging.ch) .......................... 19

Figure 2.6: Principle of 3D laser scanner operation manufactured by Riegl (www.riegl.com) ........................................................................................................................................... 20

Figure 2.7: An example of laser rangescanner based on pulse TOF (Adapted from [44]) ..... 21

Figure 2.8: Laser rangescanner adapted from [49]. The scanner consisted of (a) an 8-foot vertical truss, (b) a 3-foot horizontal arm, (c) a pan (d) and a tilt assembly (d) and a scan head (e) that mounted on the pan-tilt assembly. ........................................................................ 24

Figure 2.9: Laser footprint shape and size changes with beam incident angle, \( \alpha \), and divergence, \( \beta \). .................................................................................................................................................. 28

Figure 2.10: Mixed pixel effect........................................................................................................ 29

Figure 2.11: An example of feature base range and intensity data fusion. This method has been used for 3D modelling of large-scale scene (adapted from [63]) ......................... 32

Figure 2.12: An example of feature-based range and intensity data fusion (adapted from [25]) ...................................................................................................................................... 34

Chapter 3

Figure 3.1: Range image of a polyhedral object (right) and its intensity image (left) .......... 40

Figure 3.2: Segmentation of range data illustrated in Figure 3.1: (left) correct segmentation, (middle) under-segmentation and (right) over-segmentation. ................................................. 43

Figure 3.3: Range segmentation into planar and quadric, adapted from [112] (a) pre-processing and (b) surface extraction .......................................................... 49

Figure 3.4: Robust segmentation using MSSE. The process has two parts of estimation and segmentation ........................................................................................................... 52

Figure 3.5: The front view of the Royal Melbourne Exhibition Building, Melbourne, Australia: (top) intensity image, (bottom) colour-coded range image. ........................................ 57
Figure 3.6: Semantic net model for the architectural scene (adapted from [142]). Nodes represent the model entities (e.g. walls, roofs, floors) and are linked by architectural relationships (e.g. parallelism and orthogonality).

Chapter 4

Figure 4.1: Laser rangescanner system: (top) Overall system architecture (middle) System alignment ($r>>d$, $O_c$ is the centre of camera coordinate system and $O_l$ is the centre of LRF coordinate system) (bottom) Experimental system set-up.

Figure 4.2: Laser rangescanner system: Software architecture.

Figure 4.3: System set-up for automatic system calibration procedure.

Figure 4.4: Automatic system calibration.

Figure 4.5: Intrinsic camera parameters in a pinhole camera model. $\Phi$ and $R$ are focal and image planes, respectively. $c$ is the camera principle point and $f$ is the camera focal length. $(O_c, X_c, Y_c, Z_c)$ is the coordinate system of camera and $(c, u, v)$ is the coordinate system in the image plane.

Figure 4.6: Pinhole camera vs. real camera (adapted from [152]).

Figure 4.7: Plot of calibration validation result in $y$ direction.

Figure 4.8: System works!

Figure 4.9: Intensity image and colour-coded range image captured from the exterior of AR building (Swinburne University of Technology, Melbourne, Australia).

Figure 4.10: Royal Melbourne Exhibition Building: (a) Intensity image (720×576 pixels) (b) Colour-coded range image (715×400 pixels) (c) Range and intensity data fusion (720×400 pixels). The areas coloured in white by the fusion algorithm shows that algorithm could not find any valid range or intensity data.

Figure 4.11: The Shrine of Remembrance, Melbourne, Australia: (a) Intensity image (720×576 pixels) (b) Colour-coded range image (715×400 pixels) (c) Range and intensity
data fusion (720×400 pixels). The areas coloured in white by the fusion algorithm shows that algorithm could not find any valid range or intensity data. ................................................... 84

Figure 4.12: Measurement fluctuations over time due to temperature effect on laser rangefinder electronics.......................................................... 86

Figure 4.13: Range measurements vs. temperature at distances of nine and nineteen meters87

Figure 4.14: Comparison of the variance and RMS error for five different distances........ 90

Figure 4.15: Range image of the front side of the Shrine of Remembrance building, Melbourne, Australia: (a) edges appears as sawtooth due to the mixed pixel effect, (b) mixed pixel effect in the part of the three cross sections of the range image, between columns 250 and 500 (row 240 appears in red, row 250 appears in cyan and row 260 appears in green).. 91

Figure 4.16: Comparison of target property effects on range measurement.................... 92

Figure 4.17: System set-up to determine errors in PTU measurement: (left) horizontal angular resolution (HAR) and (right) vertical angular resolution (VAR). In this experiment, the distance from target (D) measures 1520 centimetres by LRF............................................. 94

Figure 4.18: Laser rangescanner system verification....................................................... 95

Figure 4.19: Application checking verification of laser rangescanner system .................... 96

Figure 4.20: The Royal Melbourne Exhibition Building: a) colour-coded range image (715×400) pixels, b) intensity image (715×400) pixels, and c) range segmentation result. Different colour shades in grey-scale shows each segment. ...................................................... 98

Figure 4.21: The Shrine of Remembrance: a) colour-coded range image (715×400) pixels; b) intensity image (715×400) pixels; and c) range segmentation result. Different colour shades in grey-scale show each segment. ............................................................... 99

Figure 4.22: Equivalence checking verification of laser rangescanner system............... 100
Figure 4.23: Range images of the front side of the Shrine of Remembrance, Melbourne, Australia: (a) obtained by the Riegl laser rangescanner (b) obtained by the experimental laser rangescanner ........................................................................................................................ 101

Figure 4.24: Range images of the front side of the Royal Melbourne Exhibition Building, Melbourne, Australia: (a) obtained by the Riegl laser rangescanner (b) obtained by the experimental laser rangescanner .......................................................................................... 101

Figure 4.25: Corresponding features of the range images of the Shrine of Remembrance, Melbourne, Australia obtained by: (a) Riegl laser rangescanner (b) Experimental laser rangescanner ........................................................................................................................ 102

Figure 4.26: Corresponding features of the range images of the Royal Melbourne Exhibition Building, Australia obtained by: (a) Riegl laser rangescanner (b) Experimental laser rangescanner. ....................................................................................................................... 103

Figure 4.27: Fine details embedded in the pediment and frieze of the Shrine of Remembrance building. ..................................................................................................................................................... 107

Chapter 5
Figure 5.1: Intensity and range images of the Royal Melbourne Exhibition Building. Cars, passengers and vegetation are obstacles that are unavoidable at the time of data collection. Moving objects appear as straight lines in the range image................................................. 110

Figure 5.2: The front view of the Shrine of Remembrance building. Large disparities in the size of features are highlighted in different colours. This building is obviously composed of a collection of large, medium and very small structures................................................................. 111

Figure 5.3: a) A sample of the synthetic data used to demonstrate the effect of disparity in size for large-scale range data segmentation. b) A plot representing the percentage of success in segmenting both small and large structures with a robust estimator versus the ratio of the size of small structure to the size of whole population for different values of K. ............... 113

Figure 5.4: Fine details in the façade of the Shrine of Remembrance building................. 114
Figure 5.5: Plot representing the number of required random samples versus the minimum relative size of the smallest desired detail for different probabilities of success in 3D. The number of required samples enormously increases when the size of desired structure is less than 10 per cent.

Figure 5.6: a) Sample of the synthetic data used for segmentation analysis of distant co-planar surfaces. b) Likelihood of detecting co-planar surfaces as one segment vs. distance of structures for cases where K=0.1.

Chapter 6
Figure 6.1: Inputs and outputs of MSSE as segmentor.

Figure 6.2: Cascade of Hierarchical Robust Segmentation (HRS). K is the relative size of smallest structure to be segmented. σ is range accuracy of measurement system and is varied by instrument.

Figure 6.3: Example of synthetic data with multiple structures. In this example, total number of data including gross outliers is 2,270 points; number of gross outliers is 30 points and size of the smallest structure is 40 points (1.7 per cent of the entire data).

Figure 6.4: Segmentation result for dataset presented in Figure 6.1 where P is set to 0.85 and K (size of smallest structure) is set to 0.05.

Figure 6.5: First level of robust segmentation.

Figure 6.6: Second level of robust segmentation.

Figure 6.7: Third level of robust segmentation.

Figure 6.8: Hierarchical range segmentation of the front side of the Shrine of Remembrance, Melbourne – Australia. Range data of the building is captured by Riegl (LMS-Z210) laser rangescanner: a-left) Intensity image, a-right) Range image, b) First level of segmentation, c) Final result of hierarchical segmentation. d) All possible detail (planar and decorative) of the pediment is successfully segmented. This figure is best viewed in colour.

Figure 6.9: Hierarchical segmentation of the back side of the Melbourne Exhibition Centre – Australia. Range data of the building is captured by the experimental laser rangescanner.
explained in Chapter 3: a) Intensity image of the building. b) Fist level of segmentation (coarse segmentation). c) Final result of hierarchical segmentation. This figure is best viewed in colour. .............................................................................................................................. 137

Figure 6.10: Hierarchical segmentation of the south view of Notre-Dam Church – France. Range data of the church captured by Leica HDS2500 laser rangescanner. a) Intensity image of the church [172]. b) First level of segmentation (coarse segmentation). c) Final result of hierarchical segmentation d, e, f) All possible details (planar and decorative) of different parts of building are successfully segmented. This figure is best viewed in colour. .......... 138
List of Tables

Table 2.1: Commercial laser rangescanners based on triangulation ................................. 26
Table 2.2: Examples of commercial laser rangescanners based on time-of-flight .............. 27
Table 4.1: Technical specification of LRF (http://www.mdl.co.uk) .................................... 65
Table 4.2: Technical specification of PTU (http://www.dperception.com/) ....................... 65
Table 4.3: System calibration parameters ........................................................................ 70
Table 4.4: Camera calibration parameters: estimated vs. expected ............................... 77
Table 4.5: Calibration validation result for eight distances ............................................. 80
Table 4.6: Range accuracy at different distances and statistical values of range measurements ......................................................................................................................... 88
Table 4.7: PTU angular errors ......................................................................................... 94
Table 4.8: Comparison of estimated corresponding line segments of range image obtained by Riegl and experimental laser rangescanner (image of the Shrine of Remembrance) ........ 104
Table 4.9: Comparison of estimated corresponding line segments of range image obtained by Riegl and experimental laser rangescanner (image of the Royal Melbourne Exhibition Building) ............................................................................................................................ 104
Table 6.1: Outcome of direct implementation of the robust range segmentation algorithm for the range data of the Shrine of Remembrance with different values of K (size of the smallest structure) .............................................................................................................................. 134

Table 6.2: Outcome of hierarchical implementation of robust segmentation algorithm for the range data of the Shrine of Remembrance .............................................................................................................................. 135
Chapter 1

Introduction

1.1 Background and motivation

The field of computer vision is vast and multi-disciplinary. Computer vision brings together subjects such as Computer Science, Computer Graphics, Artificial Intelligence, Robotics, Geomatics, Mathematics, Statistics, Physics and Psychology. The main task of computer vision is to devise theories and algorithms for machines to see and understand as effectively as human sight and vision. While the components of a computer vision system are highly application dependant, a typical system can be modelled and is shown at Figure 1.1. The main components of the computer vision system include:

► **Image acquisition:** the task of this section is to produce the necessary visual data by one or more optical sensor(s). Depending on the type of sensor, data can be captured in the form of single or multiple photos (ordinary 2D image), video sequences, three-dimensional point clouds (range image), etc.

► **Low-level processing:** the data obtained by the image acquisition system is raw and needs to be refined to be useable for further processing. In this stage, data can be:

---

1 Sight is the process where the retina focuses on images and vision is the ability of the brain to give these images meaning.
filtered in order to reduce noise, as sensor or environment noise may introduce false information;

re-sampled in order to correct the image coordinate system or simply reduce the size of data. This will increase the performance of further processing;

re-scaled in order to enhance the image structure at locally appropriate scales; and/or

processed for hole-filling in order to produce data for unseen part of scanned scene or object.

Object/feature/activity extraction: depending on the application, this module may function separately or as part of the segmentation section to provide an understanding for the system. Examples are separating building from vegetation in data of outdoor scene (object extraction); lines, edges and corners of the structured
objects (feature extraction); facial expression, footballers’ motion and the behaviour of moving cars (activity extraction).

► Segmentation: at this step, regions with specific information of interest are extracted from the data. Examples are planar surfaces from the structured objects, fine architectural details from exteriors and interiors of the historical buildings or brain parts from Magnetic Resonance Images (MRI).

► High-level processing: further processing of the segmented data very much depends on the application of the computer vision system. For instance, it may involve estimating the specific parameters (e.g. object pose or size) or verifying the application specific assumptions.

There are a wide and growing range of applications for computer vision systems. The applications include, but are not limited to, visualisation, automated inspection, machine vision, medical imaging and the navigation of autonomous vehicles. In recent years, with the considerable improvement in vision sensors and computation technology, a number of civic and large-scale applications have become feasible or significantly enhanced. These applications are mainly based on the 3D modelling of the urban environment and have increasingly become part of the visualisation tasks undertaken by architects, planning engineers and multimedia developers. Examples of such applications are augmented reality [1], archaeology [2], architectural reproduction [3], surveying [4] and GIS, production of automated urban models of whole buildings and streetscapes ([5], [6] and [7] to cite a few) and 3D modelling and preservation of historical and cultural sites ([8], [9], [10], [11] and [12] to cite a few).

Traditionally, producing three-dimensional models of the environment entails a lot of manual work using Computer-Aided Design (CAD) packages and is therefore a painful and time-consuming task. As technology has progressed, computer vision scientists are motivated to improve and invent data acquisition systems to capture accurate data quickly and develop algorithms to automatically generate such models in a relatively short period of time. The end user of these models can be, for instance, an archaeologist who is concerned with preserving historical buildings, an architect who is interested in visualising buildings in
3D and changing the arrangement of the kitchen and the style of the windows, or a planning engineer who would like to know how extending a shopping centre can affect the area by creating and visualising the extension model. To perform such tasks automatically and accurately, a set of data needs to be captured, and the detailed components of the model need to be identified and segmented by a vision algorithm.

1.2 What particularly is a complicated problem?

Production of range data (as a form of image data), especially from an outdoor scene, is unavoidably affected by moving objects such as birds, pedestrians and vehicles, as shown in Figure 1.2. This is widely referred to as the occlusion problem in the literature.

In addition, man-made objects in outdoor scenes appear in different sizes and geometric complexities. For example, building of the Shrine of Remembrance in Melbourne, Australia, shown in Figure 1.3 contains large objects such as walls and ceilings; medium-size objects such as staircases and statues; and smaller objects such as detailed decorative artefacts embedded in the façade of the building. Discrepancies between the size and distance of the objects in an outdoor scene highly complicate automatic separation of fine

Figure 1.2: Royal Melbourne Exhibition Building, Melbourne, Australia: cars, passengers and vegetation are obstacles that can not be avoided at the time of data collection.

Figure 1.3: Shrine of Remembrance, Melbourne, Australia: large, medium and small objects.
details from large structures. For instance, in the existence of high level of disparities, segmentation procedure of small size objects often lead to the over-segmentation of large objects while an algorithm used for large object segmentation are not generally able to detect small details.

![Figure 1.3: Examples of disparity in the size of various features. Different colours show different features.](image)

Besides, the construction of man-made objects often includes construction errors. These errors are measurable by state-of-art precise imaging systems and therefore affect the segmentation of data obtained by those systems. For instance, a visual inspection of the front view of the Shrine of Remembrance in Melbourne, Australia, shown in Figure 1.4 (a), suggests that left and right side walls, coloured in blue, are co-planar surfaces, however based on an accurate ground truth measurement, left and right side walls are considered as two planar surfaces. This issue is illustrated in the hypothetical top view of the building in Figure 1.4 (b). Any accurate segmentation algorithm could detect left and right side walls as two different structures (planar surfaces).

Furthermore, for most imaging systems such as laser rangescanners developed for the large-scale environment, range data of shiny or clear objects are hard to capture. Such systems also suffer from a *mixed pixel effect* problem on the edge of the objects due to the limitation of measurement technology.
Chapter 1-Introduction

Figure 1.4: Effect of a construction error on the range segmentation of a large building: (a) Walls of the front side of the Shrine of Remembrance building (coloured in blue) seem co-planar. (b) Accurate ground truth measurement prove they are two separate planar surfaces. It is shown in the hypothetical top view of the building.

1.3 Aim

This thesis aims to develop a cost-effective and robust technology and technique for the measurement and segmentation of geometric details embedded in the exterior of large buildings. Such a technology and technique facilitates the automatic generation of 3D...
models for the fine features of outdoor environments demanded by the emerging multimedia and engineering applications.

1.4 Methodology

Figure 1.5 illustrates the major steps of this research. At first, to establish the theoretical background for the research, a review of current range acquisition systems (step 1-1) and range segmentation methods (step 1-2) was conducted. In the second step, an experimental laser rangescanner system was designed and implemented based on the information provided in step one. The experimental system provides several sets of range data and helps to gain a better understanding for the 3D measurement of large-scale objects from a practical point of view. In addition, an existing range segmentation algorithm was applied to the range data of building exteriors that was captured by the prototype rangescanner. These experiments not only evaluated the performance of the laser rangescanner system, but also identified problems associated with processing of captured data (step three). Furthermore, to formulate the identified problems in a mathematical form, the range segmentation algorithm was applied to a set of synthetic data that was generated to simulate the range data of large buildings.

The knowledge earned by completing the first three steps provided the required resources and background to develop a new range segmentation technique (step four). This method uses a robust coarse-to-fine approach, designed to automatically extract all possible geometric features embedded in the three-dimensional data of large building exteriors. Several range data sets, captured by the experimental laser rangescanner and two other commercial systems (Manufactured by Leica and Rieggl), were used to examine the performance of the new segmentation technique (step 5).
Figure 1.5: Steps of the research

1.5 Contributions of the thesis

The major contributions of this research are threefold:

1. A new low-cost laser rangescanner system is designed and implemented to capture 3D and 2D data of the long distance environment (up to 300 metres). This system uses a novel automatic calibration process which facilitates accurate range and intensity data fusion. The implemented rangescanner system provides an in-depth understanding of range measurement systems and the problems associated with range data acquisition and processing, particularly those used for large-scale applications.

2. Problems associated with the range segmentation of building exteriors are formulated, based on the special characteristics of the geometric data of such objects.

3. A novel and robust hierarchical parametric (model-based) range segmentation technique is designed and developed. This method is capable of detecting fine details embedded in the façade of large building exteriors in a relatively short period of time. The new technique significantly reduces the computational cost and achieves the level of accuracy required for segmenting fine details.
Chapter 1 - Introduction

1.6 Overview of the study

The rest of this thesis is as follows.

Chapter two — 3D Measurement: provides an overview of three-dimensional data acquisition methods. This chapter includes a comparative analysis of the advantages and shortcomings of each approach for long-distance measurement and lays the background for the 3D measurement system that is built and used in this project. The chapter also provides a brief background on range and intensity data fusion as part of the rangescanner systems.

Chapter three — Range Data Segmentation: presents a background on the range segmentation problem. The chapter introduces the taxonomy of existing range segmentation techniques and identifies the robust estimator based techniques as the best candidates meeting the performance requirements for data segmentation of complex structures.

Chapter four — Laser Rangescanner System: describes the steps of design, development and verification of a novel low-cost laser rangescanner system for large-scale applications. The stages of system development consist of hardware and software design, calibration, range and intensity data fusion and system characterisation. The experimental system is verified using two methods — equivalence checking and application checking. The implementation of the system is presented in [13]. The data obtained by the experimental device and two commercial systems are used for this project.

Chapter five — Range Data of Large Building Exteriors: addresses the main issues exist in the geometric segmentation of complex man-made objects (particularly building exteriors). In this chapter a robust range segmentation algorithm is applied to the synthetic data in order to identify how, and where, problems influence the performance of the algorithm.

Chapter six — Hierarchical Robust Range Segmentation: proposes a solution for the computational complexity of the range segmentation of large building exteriors.
This chapter describes a robust hierarchical scheme for automatically recovering fine and coarse details embedded in the geometric data of large-scale objects. The chapter also demonstrates how the proposed segmentation scheme is able to effectively segment range data of historical buildings containing small parts, substantial noise (of different types) and large disparities in size. Most of these works are presented in [14].

Chapter seven — Conclusions and Future Directions: summarises the research and discusses the strengths and weakness of the proposed range measurement technology and segmentation technique. This final chapter also provides some suggestions and directions for future work.
Chapter 2

3D Measurement

2.1 Introduction

Generally the first stage of a computer vision system is image acquisition. Such images are usually performed by a digital camera, a camcorder or a 3D scanner system. In recent years there has been enormous progress in the development of rangescanner systems. The capability of generating, presenting and manipulating 3D digital representation of the real-world environment and objects can play a significant role in a variety of applications. These applications are as diverse as anthropometry, manufacturing, space robotics; and heritage conservation and documentation. The input for the computer vision system addressed in this research is range data of large building exteriors. This chapter provides an overview of the common 3D measurement techniques used in computer vision and discusses the suitability of those methods for long distance applications. The chapter also provides an introduction to the range and intensity data fusion as used in three-dimensional vision systems.
2.2 3D measurement systems

3D measurement technology has progressed considerably in the last two decades. A fine review on the range measurement techniques was first presented by Jarvis in 1983 [15]. Several comprehensive overviews of state-of-the-art range measurement systems, covering both hardware and software aspects are also available in the recent research papers (e.g. [16] and [17]). The taxonomy of the most common 3D measurement systems is shown in Figure 2.1.

![Figure 2.1: A classification of 3D measurement techniques based on [16] and [17].](image-url)
Based on this classification, rangefinding techniques are broadly divided into contact and non-contact methods. An important subclass of the latter, for computer vision applications, is optical range measurement. Optical 3D measurement is further subdivided into active and passive. Examples of optical distance measurement methods are triangulation (e.g. [18, 19]), time-of-flight (e.g. [15, 20]), depth-from-defocus (e.g. [21]) and light-section [22, 23].

The classification of three-dimensional imaging systems are based on the technology of the range sensor they utilise. Rangescanners used in the domain of large-scale real-world applications can be divided into two main categories ─ Triangulation and Time-Of-Flight (TOF) [15, 17, 24-26].

### 2.2.1 Triangulation based rangescanner systems

Simple triangulation as one of the first scientific methods of distance (or range) estimation has been used for centuries. For instance, the Greek philosopher Thales (624 BC) used triangulation to measure the height of the pyramids and distance to the ships at sea. The Iranian great scientist Abu Rayhan Biruni (d. 1048) applied the triangulation method to measure the Earth and distance between cities and the Dutch cartographer Gemma Frisius (d.1555) introduced triangulation technique to the map-making.

In the computer vision context, triangulation is the task of finding coordinates of a point in three-dimensional space, where its projections onto two or more images are known. In order to solve this problem, the camera parameters need to be known. Theoretically, each point in an image (2D space) corresponds to a line in 3D space so that all points on the line are projected to that 2D point in the image. If the corresponding points in two or more images can be found, they must be the projection of a common point \( P \) in 3D space. It means a set of lines in 3D space must intersect at \( P \). The geometric coordinate of \( P \) can be computed from the triangulation formula, as is described next. In triangulation measurement, at least two viewpoints are required to compute the coordinate of the 3D point \( P \). This method is called passive when both viewpoints are passive components such as CCD cameras or active when one of the viewpoints is a light projection device such as laser.
computer vision literature, passive triangulation is also referred as stereovision. Passive and active triangulation techniques are elaborated in the following subsections:

### 2.2.1.1 Passive triangulation (stereovision)

The idea of the stereovision-based data acquisition method is similar to the human visual system and the way human perceive depth. Figure 2.2 shows the basics of this method in which geometric information of the objects is calculated from two overlapping photos of object(s) or scenes that are simultaneously taken from two viewpoints ($C_1$ and $C_2$). The 3D point $P$ (which is in the field of view of both cameras) projects onto the two image plane as $P_1$ and $P_2$. Coordinates of point $P$ is then calculated from the linear relationship between $P_1$, $P_2$ and the fundamental matrix of the camera [27, 28]. This matrix is driven from camera parameters and can be obtained through a camera calibration procedure. Chapter four revisits camera calibration in more detail.

![Figure 2.2: Passive triangulation (Stereovision)](image)

- $C_1$: Camera 1
- $C_2$: Camera 2
- $C_1C_2$: Base line
- $P$: Point on the object
- $F_1$: Projection of $P$ into image plane of camera 1
- $F_2$: Projection of $P$ into image plane of camera 2
2.2.1.2 Active triangulation

The configuration of a simple active triangulation range sensor is shown in Figure 2.3. A light source (e.g. laser) projects a single point on the surface of object and a receiver (e.g. camera) detects the projected point, either directly or using a filter.

\[ D = B \tan \theta \]  

(2.1)

By knowing the distance between the light source and receiver (baseline) and the angle between receiver and baseline, the range \( D \) can be determined by the formula:

Unlike the passive triangulation method, sensor configuration for active triangulation restricts the range measurement aptitude to one single point. To provide a complete 3D view of the environment, a precise mechanical equipment (such as steering or rotating mirrors) is required to move the laser light around controlled directions (e.g. [29, 30]). An example of a rangescanner system based on active triangulation method is shown in Figure 2.4.
2.2.2 Time-of-flight based rangescanner systems

Time-Of-Flight (TOF) based 3D scanning is one of the many techniques [17] for generating dense 3D data of environment. In the TOF range measurement method, a transmitted optical or acoustic wave travels through the air towards the target and is reflected back to the receiver. The total travel time is measured by the electronic components and hence the distance of the target from the transmitter is calculated by the following formula:

$$D = \frac{v\Delta t}{2} \quad (2.2)$$

where $v$ is the speed of light or sound wave, and $t$ is the optic or sonar travelling time from the emitter to receiver. The accuracy of measurement is defined by:

$$\Delta D = \frac{v\Delta t}{2} \quad (2.3)$$

The above accuracy is determined by the characteristics of electronic components of the measurement system. For instance, to achieve 10cm accuracy, the electronic components must have a minimum 1.5 GHz bandwidth. When compared with optical TOF, sonar TOF
has less accuracy because sound is emitted as a cone and, depending on the target, the sound may be totally scattered.

The most generic device that is used as a distance measurement device in TOF based rangescanner systems is a laser rangefinder. A laser rangefinder is an instrument which uses a laser beam to measure the distance. In general, measurement in laser rangefinder devices are based on pulse time delay, amplitude modulation, frequency modulation, hybrid detection and self-mixing diodes. The common techniques of laser TOF are briefly described here.

In the pulse time delay technique, when receiver detects the reflected laser pulse, the repetition pulse generator is triggered. The receiver is usually an avalanche photodiode that provides high sensitivity and fast recovery with a high bandwidth. In this method, range $D$ is calculated by:

$$D = \frac{c}{2f_r}$$  \hspace{1cm} (2.4)

where $c$ is the speed of light and $f_r$ is the frequency of the laser pulse.

Generally, a pulse TOF provides a resolution of 0.5 to 1 cm but the speckle effect,\(^1\) which is related to the surface irregularity of the target, is the major limitation of this technique.

*Amplitude modulated (AM) phase shift* technique is another distance measurement method based on TOF. The distance is obtained from the phase shift between an emitted, and detected, coherent laser beam. The phase shift is defined as:

$$\Delta \theta = 2\pi f \Delta t$$  \hspace{1cm} (2.5)

where $f$ is the modulation frequency and $\Delta t$ is:

\(^1\) A consequence of the highly coherent nature of laser light is that the beam is scattered by different parts of the target and causes interference between radiations. Speckle is a mottled pattern that arises when laser light falls on a non-specular reflecting surface.
\[ \Delta t = \frac{2D}{c}. \]  

(2.6)

Hence the range \( D \) is calculated from:

\[ D = \frac{\lambda \Delta \theta}{4\pi} \]  

(2.7)

In comparison with pulse time delay method, this technique provides a highly accurate range resolution of 3 to 5mm. However, the only limitation of this method is range ambiguity which occurs when the distance value exceeds \( \frac{\lambda}{2} \times (\sin (\alpha t + \pi) = -\sin \alpha t) \).

The frequency modulated (FM) technique employs TOF to measure range using a transmitted coherent laser beam which is modulated linearly over a given frequency, \( f_m \). Inside the detector, the reflected signal is mixed with a reference signal produced by an internal oscillator. The signal produces a beat frequency, \( f_b \), which is proportional to the distance. Range is then determined from the following formula [20, 32]:

\[ D = \frac{c f_b}{4 f_m \Delta f} \]  

(2.8)

where \( \Delta f \) is the frequency deviation. While ambiguity is not a problem in FM measurement technique, it suffers from the speckle effect (signal fluctuation) [31, 32].

Three-dimensional imaging systems based on TOF range sensors can be divided into two categories ─ scanning and non-scanning 3D TOF cameras.
2.2.2.1 Non-scanning 3D TOF cameras

Non-scanning three-dimensional pulse TOF cameras were invented in the 1990s [33, 34]. In this method, instead of point-wise range scanning, the range data of the entire scene is measured parallel with a modulated surface. A modulated infrared light source emits the light pulses to the object. The reflected light travels back to the camera, where their precise time of arrival is measured locally for each pixel of a custom image sensor. The camera electronic controls the data acquisition, processes the depth image and transfers the range and intensity image to the computer. Figure 2.5 (a) and (b) show the principle [33] and an example of commercially available [35] 3D cameras, respectively.

![Figure 2.5: Non-scanning 3D TOF camera: (a) Principle (Adapted from [33]) (b) Commercialised, manufactured by MESA imaging (http://www.mesa-imaging.ch/)](http://www.mesa-imaging.ch/)
2.2.2.2 Scanning 3D TOF cameras (3D laser scanners)

Scanning 3D TOF cameras often consist of a laser rangefinder and a scanning mechanism, commonly called 3D laser scanners or laser rangescanners. The scanning device can be rotating (e.g. [32, 36]) or oscillating mirrors, pan-tilt units [37], combinations of mirrors and pan-tilt units [38, 39], opto-mechanical heads (e.g. [40-42]) or micro-mechanical mirrors (e.g. [43]). Figure 2.6 shows a typical state-of-the-art commercial laser rangescanner system (manufactured by Riegl).

![Figure 2.6: Principle of 3D laser scanner operation manufactured by Riegl](www.riegl.com)

1. Laser rangefinder electronics
2. Laser beam
3. Polygon laser mirror
4. Optical head
5. Ethernet interface
6. Laptop
7. Data acquisition software

One of the earliest instances of laser rangescanner systems, with its design based on amplitude modulation laser TOF technique, was introduced by Nitzan et al [36], in 1977. The system comprised an amplitude modulated transmitter that emitted the laser beam, a receiver and a scanning mirror. The modulated beam was deflected by the scanner to the objects. A fraction of the laser beam was then reflected back along a coaxial direction with the incident beam and deflected into the receiver through the scanning mirror. In the receiver part a photomultiplier converted the received laser beam to a sinusoidal current. The range was then estimated by the measurement of the phase shift of the transmitted and received signal. The laser reflectance strength of the object was also obtained from the amplitude of the received signal.
In turn, one of the first rangescanners based on pulse modulation TOF method was presented by Moring et al. [44], in 1986. As shown in Figure 2.7, the system consists of a laser rangefinder and two scanning mirrors driven by stirring iron galvanometers where the diversion angles of the mirrors are controlled by the servo amplifiers. The maximum scanning rate of the system was 20 line/s and the maximum range image size was 512×512 pixels. The maximum horizontal and vertical field of view of the system was 35°. The 3D scanner system was designed for laboratory-size applications; however, same principle can be used to develop laser scanner system to measure longer distances.

The 3D system presented by Hancock et al. [45] in 1997 is one of the examples of early attempts to design and build high speed and highly accurate laser rangescanner systems. It used Z+F laser for range measurement based on amplitude-modulation method to provide a high speed measurement with a pixel rate of 500 kHz. The scanning component consisted of a gold-coated scanning mirror, a yoke assembly to allow the mirror to rotate vertically, a spindle assembly to pivot the yoke assembly horizontally, electronic circuits to...
establish the position of the mirror and a computer to collect and store range data. The scanning part was designed to provide high accuracy with maximum resolution of approximately 0.06 degrees in both vertical and horizontal scan. The maximum measured range varies, depending on the type and power of the laser has been used. For example the range can be varied from 0.10m to 400m.

Another effort to design and develop a 3D scanner for capturing dense range data of environment was made by Nyland [46] in 1998. The hardware of the system consisted of a TOF laser rangefinder (AccuRange 4000-LIR manufactured by Acuity Research), a line sweeping mirror, a panning unit (PTU-46-70 manufactured by Directed Perception) and a digital camera (EOS D2000 manufactured by Canon). A software system was designed and developed to control the hardware parts through a conventional personal computer. The maximum distance that could be measured with this system is approximately 15m and the measurement speed could be set up to 50 kHz. The vertical scanning resolution was adjustable between 5-20 pixels per degree and the horizontal scanning resolution could be set at 0.771 arc minutes (about 0.01 degrees) at its best resolution. This scanner could acquire dense, accurate and high resolution range data of real world environment; however the laser that was used for distance measurement was not eye-safe. Replacing the laser rangefinder with an eye-safe and long range laser can improve the safety and range of the measurement system.

In 1999, Zhao [47] presented a laser rangescanner for urban applications. In that system, a TOF laser rangefinder was set together with a CCD camera on a programmable rotating platform. Both range measurement and rotation of the scanning platform were controlled through a serial port and the CCD camera was controlled with a capture board. The acquisition system was designed to take a sequence of overlapping intensity image patches to generate a spherical image of the environment as well as range data. Each image patches were then mapped onto the corresponding range data captured by the laser rangefinder. This system was inexpensive with wide field of view and long range measurement (up to 300m) appropriate for outdoor application. In addition, the direct range and intensity of data integration made urban scene reconstruction more accurate and with less post-processing.
In 2000, Haverinen and Röning [48] has presented a laser rangescanner system, based on active triangulation principle, to simultaneously capture geometry and intensity data of environment as part of a robot architecture. The Colour-Range Scanner, called CRS, utilises a 256×256 element MAPP2200 sensor to extract the projected laser profile from the surfaces of the object. The colour values for each 3D point are acquired using a standard colour CCD camera. A beam splitter is used to provide a common optical axis for those two sensors in order to provide accurate colour mapping. For the purpose of the camera-to-camera calibration task, they have developed a camera calibration toolbox for MATLAB. The scanning component of the system consists of a mirror that rotates the laser sheet using a galvanometer. The scanning controller is part of the host computer and is able to control CRS via RS232 as well as Common Object Request Broker Architecture (CORBA). This scanner is designed to be simple, compact and portable, however the size and resolution of the captured range image is not adjustable and is limited to the size of the MAPP2200 sensor.

In the same year, Levoy et al. [49] designed a hardware and software system for the range and colour data acquisition of large delicate objects (e.g. Michelangelo’s David) under non-laboratory conditions. The system hardware consists of a laser strip triangulation rangefinder, a motorized gantry and a 512×480 pixel CCD camera to provide the desired accuracy, robustness and portability for the large-object applications. The rangefinder used in this project is a 5mW, 660-nanometer laser diode manufactured by Cyberware Inc. Unlike the rangescanner designed by Haverinen and Röning [48], this scanner system uses a fixed triangulation angle of 20° rather than 30°, and the sensor views the laser sheet from only one side rather than combining views from both sides using a beam splitter. Therefore, the baseline of the triangle is smaller, which in turn cause to reduce the size and weight of the scan head. The scanning part of the system design is based on a rotational motion to produce 100° of tilting and 100° of panning motion. The system software has been designed to acquire, align, merge and visualise scanned data. Figure 2.8 shows the detail of the rangescanner system set-up.
Chapter 2- 3D Measurement

Figure 2.8: Laser rangescanner adapted from [49]. The scanner consisted of (a) an 8-foot vertical truss, (b) a 3-foot horizontal arm, (c) a pan (d) and a tilt assembly (d) and a scan head (e) that mounted on the pan-tilt assembly.

The above rangescanner provides accurate range data but the speed of range acquisition is relatively low due to the usage of the motorized gantry. In addition, the acquired scans require a lengthy post-processing procedure to produce a polygon mesh associated with intensity value at each mesh vertex.

2.2.3 Triangulation vs. laser TOF based rangescanners

Rangescanner systems based on triangulation and TOF range sensors each have their own advantages and limitations that make them suitable for different purposes. The principle of passive triangulation is simple and building a rangescanner based on this method is not expensive, however there are three major drawbacks with this method: First, finding corresponding points in two images requires a great deal of computation and the intensity cameras need to be calibrated accurately. Second, the accuracy of point matching is inversely proportional to the baseline of the system \( C_1C_2 \) in Figure 2.2) and again drops off with large baselines. As a result this method is not able to generate dense and accurate 3D data of large-
scale scans or objects located in long distances. Lastly, the quality of the range sensing is fairly sensitive to the ambient light. Therefore, the usage of this method is generally limited to indoor/short distance measurement where lighting can be controlled.

Similar to the passive triangulation method, one of the major disadvantages of active triangulation technique is the complexity and the cost of computation [50]. This issue is mainly because a new image has to be taken for each position of the single point. In addition, a fairly large baseline is required in order to measure a large distance and, as a result, this method is also limited to short distance measurement. Table 2-1 presents the maximum measurable distance and accuracy for some of the commercially available laser rangescanner systems, based on triangulation method. As shown in this table, triangulation rangescanners have a limited range of a few metres yet their accuracy for close distances is relatively high, in the order of micrometers.

Three-dimensional rangescanners, based on the laser TOF distance measurement techniques, are fast and compact with a tight focus. Also, unlike other rangescanners such as stereo vision, they are less depending on the time of the day as they can also work at night. In addition, all data is collected is three-dimensional, so CAD or solid models can be built from the raw data with less processing costs. Moreover, they are capable of operating over very large distances, in the order of kilometres. These devices are therefore suitable for scanning large objects such as building exteriors, however the accuracy of laser TOF scanners are relatively low (compared to triangulation-based scanners), in the order of millimetres. This accuracy is mainly due to the limitation of the electronic devices inside the rangescanner systems. Table 2.2 shows the maximum range and accuracy of some of the most popular commercial laser rangescanners.
### Table 2.1: Commercial laser rangescanners based on triangulation

<table>
<thead>
<tr>
<th>Company</th>
<th>Maximum Range</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taicaan (Xyris 2000)</td>
<td>2.5 mm</td>
<td>250 nm</td>
</tr>
<tr>
<td>Virtek Vision International</td>
<td>25 mm</td>
<td>100 µm</td>
</tr>
<tr>
<td>Metris</td>
<td>100 mm</td>
<td>15 µm</td>
</tr>
<tr>
<td>Kreon</td>
<td>50 – 100 mm</td>
<td>10-25 µm</td>
</tr>
<tr>
<td>Scantech</td>
<td>100 – 400 mm</td>
<td>100-200 µm</td>
</tr>
<tr>
<td>Steintek – 3D – SCAN</td>
<td>100 – 600 mm</td>
<td>20 – 300 µm</td>
</tr>
<tr>
<td>Acuity Research – AR600</td>
<td>Up to 0.5 m</td>
<td>Up to 50 mm</td>
</tr>
<tr>
<td>Minolta – Vivid Systems (Vi – 9i)</td>
<td>0.6 – 1m</td>
<td>0.05mm</td>
</tr>
<tr>
<td>Mensi – SOISIC</td>
<td>Up to 25m</td>
<td>0.2-0.6mm</td>
</tr>
<tr>
<td>Cyberware</td>
<td>Varied</td>
<td>50 – 300 µm</td>
</tr>
</tbody>
</table>
### Table 2.2: Examples of commercial laser rangescanners based on time-of-flight

<table>
<thead>
<tr>
<th>Company</th>
<th>Maximum Range</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis Software Inc.-SurPhaser</td>
<td>5m</td>
<td>25 (\mu)m</td>
</tr>
<tr>
<td><a href="www.surphaser.com">www.surphaser.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rdTech</td>
<td>0.3-12m</td>
<td>7.5 mm @ 12m</td>
</tr>
<tr>
<td><a href="www.3rdtech.com">www.3rdtech.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metric Vision</td>
<td>2-24m</td>
<td>50-300 (\mu)m</td>
</tr>
<tr>
<td><a href="www.metricvision.com">www.metricvision.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acuity Research – AR4000</td>
<td>50m</td>
<td>2.5 cm</td>
</tr>
<tr>
<td><a href="www.acuityresearch.com">www.acuityresearch.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoller + Froehlich GmbH</td>
<td>53.5m</td>
<td>(\leq 5) mm</td>
</tr>
<tr>
<td><a href="www.zofre.de">www.zofre.de</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQSun Gmbh</td>
<td>80m</td>
<td>3 mm</td>
</tr>
<tr>
<td><a href="www.iqsun.com">www.iqsun.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyra-The Cyrax System</td>
<td>100m</td>
<td>4 mm</td>
</tr>
<tr>
<td><a href="www.cyra.com">www.cyra.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trimble</td>
<td>200m (standard)</td>
<td>7mm @ 100m</td>
</tr>
<tr>
<td><a href="www.trimble.com">www.trimble.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDL – LaserAce@IM</td>
<td>5 – 300m</td>
<td>10 mm</td>
</tr>
<tr>
<td><a href="www.mdl.co.uk">www.mdl.co.uk</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riegl - LMS - Z210</td>
<td>400m</td>
<td>15 mm</td>
</tr>
<tr>
<td><a href="www.riegl.co.at">www.riegl.co.at</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I–Site 3D Laser Imaging</td>
<td>400m</td>
<td>50 mm</td>
</tr>
<tr>
<td><a href="www.isite3d.com">www.isite3d.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optech</td>
<td>800m</td>
<td>10 mm</td>
</tr>
<tr>
<td><a href="www.optech.on.ca">www.optech.on.ca</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
One of the inherent properties of TOF rangescanners that strongly affects the measurement accuracy is the beam-width. The apparent location of the range measurement is along the centreline of the emitted beam. The actual location, however, cannot be estimated as it may lie within the beam footprint [51]. To estimate this uncertainty for a laser TOF rangefinder, it can be assumed that the level of the laser power across the beam-width diameter is as a symmetrical Gaussian distribution, where 100 percent of the laser energy is within the footprint (as in [52]). Based on this model, the footprint diameter can be approximated by ±3σ (~99.7%). As shown in Figure 2.9, the laser footprint can be modelled as an ellipse that is shaped by the intersection between a cone and the local tangent plane with normal \( n \) (target area).

![Figure 2.9: Laser footprint shape and size changes with beam incident angle, \( \alpha \), and divergence, \( \beta \).](image)

In Figure 2.9, the cone origin denotes as \( O \), laser direction as \( l \) and beam divergence as \( \beta \). Once the main axes (major and minor axis of ellipse are \( a \) and \( b \)) of the footprint are computed in 3D using cone canonicals, the footprint can be decomposed into its maximum
horizontal and vertical error components. Assuming the Gaussian power distribution is symmetrical, the approximate positioning error due to the scanning geometry can be calculated from the following equations [52]:

\[
\sigma_{xy}^{\text{geom}} = (1/3) \cdot \max \left\{ \begin{bmatrix} a_x \\ a_y \end{bmatrix}^T \begin{bmatrix} b_x \\ b_y \end{bmatrix} \right\}
\]

and

\[
\sigma_{z}^{\text{geom}} = (1/3) \cdot \max \left\{ \begin{bmatrix} a_z \end{bmatrix}^T \begin{bmatrix} b_z \end{bmatrix} \right\}
\]

The uncertainty in beam-width causes imprecise measurement in two ways: first, when the emitted beam bridges a depth discontinuity; and second when the emitted beam points to the target area with an angle of incident more than 90°.

The limitation of TOF rangescanners in depth discontinuities is named the mixed pixel effect [53] [32] [54]. This artefact occurs when light source (e.g. laser beam) is reflected off the edge of the target (Figure 2.10). In this case, the estimated range is a weighted average of the distances to the two surfaces where the laser footprint lies, and has no physical meaning. As the footprint of light is a function of distance and beam divergence, the narrower or closer the beam results in a more accurate measurement.

*Figure 2.10: Mixed pixel effect*
A light beam with an angle of incidence not equal to 90° leads to inaccuracy in range measurement as the laser footprint is also a function of the incident angle [52]. As a result, narrower beams result in a more accurate measurement.

In addition, range measurement using TOF technique is colour-dependant, meaning that distance measurement jumps at reflectance and colour boundaries. Therefore, very dark or highly reflective objects produce highly imprecise measurements as dark objects reflect less energy to the detector and reflective objects saturate the range detector.

In order to build a photorealistic model of the environment, it is desirable to construct 3D models with additional attributes such as colour information. In fact, by assigning (mapping) an intensity to each range value, a six dimensional data \((X_{ij}, Y_{ij}, Z_{ij}, R_{ij}, G_{ij}, B_{ij})\) can be achieved, where \(R_{ij}, G_{ij}, B_{ij}\) are the intensity values of each pixel. With the additional information provided by photometric attributes, the accuracy and robustness of 3D modelling applications is increased. In the literature of computer vision, the task of combining range and photometric (/texture) information refers to data fusion (e.g. [55, 56]), 3D to 2D registration (e.g. [57]) and data integration (e.g. [58]).

Fusion of range and intensity image is fairly straightforward when the data is obtained using the same sensor, such as range from passive triangulation, range and reflectance obtained by laser rangefinder or range from structured light. The task, however, becomes complex when data is acquired by two different sensors such as a laser rangefinder for range data and a camera for intensity data. In the first approaches, points belonging to the two sets of data are directly assigned from 2D to 3D to obtain an accurate and fast result. In contrast, the latter requires a fine initial estimation and calibration to recover the rigid transformation between the sensor poses. Both methods can be mixed to obtain maximum accuracy and reliability as a result of the fusion process.

### 2.3 Range and intensity data integration

Based on the available rangescanner systems, the process of combining shape (geometry) and appearance (colour) of the objects can be classified into two main categories ─ feature-
based and system-based. Each category is elaborated upon here:

2.3.1 Feature-based range and intensity data fusion

Generally, in feature-based range and intensity data fusion methods, 3D attributes – e.g. lines, planar patches or quadratic surfaces – are extracted from range image and 2D attributes, for example edges, corners and vanishing points, are detected from intensity image. A fusion algorithm is then designed to match the corresponding features obtained from two data sets. Examples of feature-based fusion techniques, designed for real-world applications, are set out below.

In 1996 Umeda et al. [59] presented a range and intensity data fusion, based on the features such as planar and cylindrical surfaces. In this method, planar and cylindrical regions are first extracted from range data. Corresponding planar regions, from range image and edges from intensity image, were then fused using 3D edge measurements. Next, parameters of the cylindrical surfaces, such as radius, centre and direction of central axis of the cylinder, were measured from range data to project the cylindrical surfaces onto the corresponding edges in intensity image.

In the same year, Weckesser et al. [60] presented an approach to merge range and intensity information to apply mobile robot navigation. In this method, linear segmentation was used to extract edges from data sets obtained from the trinocular vision system and the laser line scanner. The edge segments extracted from stereo measurement are represented by their midpoint, half-length and direction vector. The *iterative end-point fit* algorithm was used to segment edges from data set obtained by line scanner. The line segments were then represented by their normalised direction vector to match with the line segments from stereo reconstruction. The line segments were then transformed from the robot coordinate system to the world coordinate system using rotation and translation. In order to find corresponding edge segments in both data sets, the *Mahalanobis distance* were applied as the matching criterion and the *Kalman filter* was used for feature matching. To improve this method, in 1999 Neira et al. [61] integrated range and intensity data to obtain map-based localisation for the path planning of a mobile robot using SPmodel [62] for edge-extraction. SPmodel uses
the concept of symmetry to present a common representation for different types of geometric features obtained by different sensor.

In 2001 Stamos and Allen [63, 64] introduced an automatic feature-based method to accurately integrate range and intensity data sets of urban environment to produce a realistic CAD model of building exteriors. In this method, the inherent properties of building exteriors, such as parallelism and orthogonally, were used to match corresponding rectangular and quadrangular structures extracted from 3D and 2D data sets, respectively. In order to find the optimal match between range and intensity data sets, the RANSAC technique was employed. This technique was used to search the space from the set of 3D and 2D lines to the set of 3D and 2D rectangles. Figure 2.11 shows the principle of this method.

Figure 2.11: An example of feature base range and intensity data fusion. This method has been used for 3D modelling of large-scale scene (adapted from [63])
Chapter 2 - 3D Measurement

The advantage of this method is the flexibility of data acquisition and the automatic processing of 3D modelling. In addition, unlike most other methods, the pre-camera calibration process is not necessary. However, using the RANSAC technique when data is contaminated by a large number of outliers and pseudo-outliers is computationally an expensive task. It is due to the fact that if one wants to ensure, with probability close to one, that at least one of the random selections corresponds to a valid match, a large number of samples are needed. In order to solve this problem, in 2005, Liu and Stamos [65] developed a new approach for 3D and 2D data fusion. Similar to the previous method, the range data are abstracted into sets of parallel lines, followed by three clustering steps to extract 3D features (rectangles). For the corresponding 2D image, features are produced through vanishing point extraction, camera calibration and rectification steps. Then, an optimised transformation between the 2D and 3D data sets is computed. Unlike previous methods, this algorithm is based on a match of 3D with 2D features that maximises an overlap criterion.

In 2001, Dias et al [25, 66] presented a 3D and 2D data integration method to improve 3D modelling of the real world scenes, where they used data features in different way from the previous methods as illustrated in Figure 2.12. This method first resized the video image to match the dimension of the reflectance image. To this purpose, a Canny edge detector and the distance transform were applied to the video image. The reflectance edges were then translated, scaled and rotated (see [27, 28] for methods) over the distance transform image until a minimum distance between the edges was found. In the next step, a corner detector was employed to find some control points (corners) of the reflectance image. The corresponding points in the intensity image were then obtained. Finally, the matched images were utilised to estimate intrinsic and extrinsic camera parameters based on Tsai camera calibration technique [67]. The camera parameters then used to reproject intensity image onto reflectance image.

---

2 The random sampling problem will be discussed in Chapter 3 in more detail.
A weakness of this method is the difficulty to find the fully reliable corresponding points from the reflectance and intensity images. This weakness affects the goodness of camera calibration, which in turn affects the accuracy of the fusion. To improve this method, in 2003 the same research group [68, 69], used passive triangulation of the intensity data to refine the initial camera calibrations and ensure a more reliable fusion of the range and intensity data sets.

When range measurement is obtained by a laser rangefinder, in most measurement systems, the laser reflectance strength can also be obtained for each range data point. As the reflectance image is attached to the range image, and laser reflectance image and intensity image of an object are highly correlated, texture mapping of such datasets is simplified to the 2D-2D registration problem [36]. The fusion method extended by Kurazume et al. [70], in
2001, is an example of colour and reflectance data fusion. In this technique, edges of the objects are extracted from colour and reflectance images using Canny operator [71]. The corresponding edges were then determined through the iterative pose estimation of each sensor using a robust M-estimator. In 2005 Sagawa et al. [72] introduced a method of texture mapping using the same principle. In their work, range and reflectance data is first needed to be merged. Then, the edges from colour image are extracted using Canny operator. The occluding edges and reflectance edges are also extracted from 3D data. In order to accomplish texture mapping, the posture of the camera is iteratively computed.

2.3.2 System-based range and intensity data fusion

In system-based range and intensity data fusion, mapping intensity data into volumetric data is based on the characteristics of the rangescanner system such as the method of rangescanning and the relative position of range and intensity sensors (e.g. [73], [74] and [75]). In this method, the accuracy of the fusion task is directly related to the accuracy of the sensors calibration. The task of range and intensity fusion using the system-based method is straightforward where rangescanning is based on triangulation method. In this case, the rangescanner first captures intensity image(s) from one or more viewpoints (depending on the passive or active triangulation set-up) and the distances are then calculated by finding the corresponding points. In this case, there is an intensity value attached to a calculated range and registration is relatively precise. Some examples of system-based fusion method are as follows:

In 1998 Zhao and Shibasaki [76] presented a system-based range and intensity data fusion for urban modelling applications. In this method, a translation algorithm was employed to align range and intensity data. Panoramic CCD images were then projected onto the triangular meshes using inverse projection of range data into intensity data.

In 2004, Haala et al. [77] have developed a range and intensity data fusion for forestry applications to recognise tree species. In this method, both data sets are represented in a cylindrical coordinate system by using a second order polynomial to simplify the process of assigning intensity data to range data. To determine parameters of the polynomial the measurement of several corresponding points in intensity and reflectance images is required.
This method is simple but the accuracy of fusion is highly related to the accuracy of estimated parameters.

In the same year, Umeda et al. [78] also presented a precise range and intensity data fusion based on the simultaneous range and reflectance data acquisition. This method utilises the reflectance image and uses the optical flow constraint between the reflectance image and the texture image to form range and intensity data fusion. In order to simplify the data fusion process, the intensity sensor is carefully calibrated using the least squares method.

Scheibe et al. [56] have also developed a system-based data fusion technique to integrate geometry data, captured by laser rangescanner system and panoramic images obtained from several camera scans. This method employs traditional ray-tracing algorithm to find the corresponding points in both data sets. The algorithm initially assumes that the centre point of the laser rangescanner and the optical projection centre of the camera are identical. In this case, the laser points can be directly assigned onto the image pixels. However, in real world applications, this common way is not accurate mainly for two reasons: existence of a distance between the laser rangescanner and camera; and the large number of data. As a result, the assignment of the corresponding points should be within the local neighbourhoods that can be achieved through the triangulation and meshing of the range data.

To produce a 3D model of wide area outdoor environment, in 2005 Asai et al. [79] employed a range and colour integration method. In their proposed method, range data acquired by an omnidirectional laser rangescanner system and colour information captured by an omnidirectional multi-camera system, were both aligned and installed on a vehicle. Multiple range images are first registered by minimising the distance between corresponding points in the different range images. The initial position and orientation values of registration were obtained by a GPS and Gyroscope. As the geometrical relationship between the rangescanner and camera is known and fixed, the correspondence between a 2D point in intensity image and a 3D point in range image can be estimated and assigned.
2.4 Chapter review

This chapter is comprised of two parts. The first part discussed the principles of 3D measurement systems and presented the latest technology in range imaging. A comparison of range sensors shows that although triangulation-based methods provide affordable and simple system design, they are computationally expensive and fairly sensitive to environmental conditions. Since in this method the range resolution is proportional to distance, for longer distances, the range data is relatively coarse. As a result, the usage of triangulation-based rangescanner systems is generally limited to indoor (short distance) measurements where cameras can be located in close proximity to each other and lighting can be controlled.

On the other hand, laser time-of-flight based rangescanners are preferred systems for large object range measurements for two main reasons: first, they provide a direct range computation which reduces the computational cost and acquisition time. Second, as the accuracy of laser time-of-flight measurement is not very sensitive to the ambient illumination, the imaging (measurement) process is rather straightforward. This is more important in the outdoors where the ambient light is not easily controllable (or predictable). The inherent drawback of this method of imaging is that the image does not represent the real scene of any specific instant as it is not obtained in an instantaneous shot. Also, the mechanical parts used for scanning consume additional energy.

The second part presented background and a review of range and intensity data fusion methods with an emphasis on large-scale real-world applications. Based on the review of literature, data integration algorithms can be categorised into two techniques: feature-based and system-based. In general, feature-based range and intensity data fusion methods work well on the data sets obtained from the structured man-made environment. This method of fusion is independent from the data acquisition devices. However the feature extraction — such as edge detection — techniques can be easily trapped in a local minimum on surfaces with rich textures.

System-based range and intensity data integration has been widely used in systems which recreate photorealistic models of the environment. In these systems, the range and image integration problem is generally solved by fixing the relative orientation and position
of the intensity camera, with respect to the range sensor by rigidly attaching two sensors on
the same platform. In spite of lacking 2D image acquisition flexibility, this approach
provides a relatively inexpensive and accurate range and intensity data fusion for large scale
applications.
Chapter 3

Range Image Segmentation

3.1 Introduction

In the previous chapter 3D imaging techniques and technologies were introduced. The output of such systems is a number of range measurements, often called range data or range image. Range data in itself is not generally useable for most applications and therefore needs to be interpreted into useful spatial information.

The task of obtaining a useful and meaningful representation of the range data sets is known as range segmentation. A ubiquitous application of range segmentation is 3D object recognition [80-84]. Most 3D object recognition methods segment various surfaces of the objects before applying a recognition algorithm. Another applications of range segmentation are reverse engineering [85], archaeological visualisation [86] and 3D reconstruction [87, 88]. Figure 3.1 shows range image of a polyhedral object and its intensity image.
This chapter provides the required theoretical background for range segmentation, particularly for the parametric range segmentation of large and complex objects. In Section 3.2, the problem of range segmentation is defined and the method of Hoover et al. [89] in evaluating a range segmentation algorithm is described as this evaluation scheme is later used to evaluate the performance of the proposed segmentation algorithm. A taxonomy of the range segmentation algorithm is presented in Section 3.3. Then, the range segmentation methods applicable to the range image of complex large objects are discussed in Section 3.4.

### 3.2 Range segmentation

The range segmentation task is to divide a set of range data into meaningful, homogeneous regions with no overlap where the union of the segments is the entire data set. In mathematical terms, if \( R \) is the whole range data and \( S_1, S_2, \ldots, S_k \) are the segmented regions, then range segmentation can be defined as follows [20, 90]:

1. \[ \bigcup_{i=1}^{k} S_i = R \]  
   (This is true when there is no outlier. In the presence of outliers this formula can be defined as \[ \bigcup_{i=1}^{k} S_i + \text{outliers} = R \) 
2. \( S_i \) is a homogenous, \( i = 1, 2, \ldots, k \)
3. \( S_i \cap S_j = \emptyset \) when \( i \neq j \)
4. \( P(S_i) \) is True for \( i = 1, 2, \ldots, k \)

5. \( P(S_i \cup S_j) \) is False when \( i \neq j \)

where \( P(S_i) \) is a logical predicate over the points in set \( S_i \).

The main characteristics of an effective range segmentation algorithm are [91] that:

- the segmented regions have to be uniform and homogeneous;
- the segmented regions should contain as few holes as possible and
- the boundaries of regions have to be accurately preserved. This characteristic is especially important to some of the applications such as reverse engineering and historical building preservation.

The performance of a range segmentation technique can be evaluated by the well-known method of Hoover et al. [89, 92]. This method assumes that the ground truth image contains \( N \) regions, while segmentation algorithm detects \( M \) regions. Let \( S_n \) and \( S_m \) denote a region in the ground truth and segmentation result, respectively. The number of data points in \( S_n \) and \( S_m \) are equal to \( P_n \) and \( P_m \), in that order. Let \( O_{mn} = S_m \cap S_n \) be the number of pixels belonging to the intersection of regions \( S_m \) and \( S_n \). Therefore the fraction of \( S_n \) where the intersection of \( S_n \) and \( S_m \) covers is \( O_{mn}/P_n \). Similarly, \( O_{mn}/P_m \) represents the fraction of \( S_m \) that the intersection of \( S_n \) and \( S_m \) covers. According to the evaluation method introduced by Hoover et al., the decision for the classification of range segmentation result is made based upon a threshold \( T \). This threshold refers to the percentage of the \( S_n \) and \( S_m \) coverage and can be set to reflect the degree of exactness of the desired definition. The result of a range segmentation task can then fall into one of the five categories described below:

- **Correct segmentation:** The region is correctly segmented, if more than \( T \) percent of data points in the segmented region (\( S_m \)) are correctly assigned. Figure 3.2 (left) shows the correct segmentation of a polyhedral object presented in Figure 3.1.

- **Over-segmentation:** A region \( S_n \) in ground truth image and a set of regions \( S_m^1 \ldots S_m^i \) (where \( 2 \leq x \leq M \)) in the segmentation result are considered an instance of over-segmentation if:

\[
\forall i \in \{2, \ldots, M\}, O_{mn} \geq T \times P_m
\]  

\((3.1)\)
The above equations show that there are more than \( T \) percent of data points in an over-segmented region that should have been assigned to other regions. Figure 3.2 (right) depicts an example of an over-segmentation of the polyhedral object presented in Figure 3.1.

- **Under-segmentation:** A region \( S_m \) in the segmentation result and a set of regions \( S_{n_1}, \ldots, S_{n_i} \) (where \( 2 \leq x \leq N \)) in the ground truth image are considered an instance of under-segmentation if:

\[
\forall i \in \{2, \ldots, N\}, O_{mn} \geq T \times P_n \tag{3.3}
\]

and

\[
\sum_{i=1}^{x} O_{mn} \geq T \times P_m \tag{3.4}
\]

Based on the equations 3.3 and 3.4, an under-segmented region contains more than one identical surface. Figure 3.2 (middle) shows that the region in purple in the segmentation result of the polyhedral object presented in Figure 3.1 is under-segmented. The yellow region in the segmentation result belongs to the purple region.

- **Missing region:** If a region \( S_n \) in the ground truth image does not exist in any instance of correct detection, there is an over-segmentation or under-segmentation in the segmentation result and the region is missed. Figure 3.2 (middle) shows that there are two regions missing in the segmentation result of a polyhedral object presented in Figure 3.1.

- **Noise:** Noise is those data points in the segmentation result that do not belong to any region of correct segmentation, over-segmentation or under-segmentation.
Chapter 3 — Range Image Segmentation

Figure 3.2: Segmentation of range data illustrated in Figure 3.1: (left) correct segmentation, (middle) under-segmentation and (right) over-segmentation.

As suggested by Hoover et al., the threshold $T$ is considered to be $0.5 \leq T \leq 1$. In this range any region can contribute to at most three classifications being correct segmentation, over-segmentation and under-segmentation. To explain this, firstly, consider that based on the definition of correct segmentation, more than $T$ percent of a $S_n$ (a ground truth region) must overlie some $S_m$ (a segmented region). This means that only $1.0 - T$ percent of the $S_n$ can overlap any other $S_m$ region. It is clear that $1.0 - T$ cannot be greater than $T$ and $T > 0.5$, therefore no other $S_m$ region can overlay the $S_n$ region sufficiently to produce another correct segmentation. This argument is correct for any segmented region in a correct segmentation.

Secondly, with the same argument and based on the definition of over-segmentation, if the union of the set of $S_m$ regions overlaps the ground truth region by at least $T$ percent, then once again there is not enough left of the $S_n$ region to use in assigning another over-segmentation. Lastly, reversing the direction of the previous argument between $S_n$ and $S_m$ proves the same result for an under-segmentation assigning.

### 3.3 Taxonomy of range segmentation algorithms

There is a number of three-dimensional data segmentation methods reported in computer vision literature. These methods can be widely classified into three main categories: edge detection, region growing and parametric fitting algorithms.
3.3.1 Edge detection based range segmentation algorithms

In these methods, discontinuities are extracted using an edge detection operator. Discontinuities in range images refer to the surface (jumps and boundaries) and orientation (creases and roofs). Once boundaries are extracted, edges with common properties are clustered together. An example of edge detection based range segmentation algorithms is presented by Fan et al. [93]. The segmentation procedure first detects discontinuities using zero-crossing and curvature values to obtain a preliminary segmentation. In the next step, the initial segmentation is refined by fitting quadratics where the coefficients are calculated, based on the Least Squares method.

Another example of the range segmentation method based on edge detection is introduced by Pajdla and Hlaváč [94] with the emphasis on roof or crease edge detection, based on polynomial approximation. In this method, range data was first locally approximated by second order bivariate polynomial. Then, the 2D problem of surface discontinuity strength estimation was converted into a 1D problem along the direction of maximal normal curvature at each point on the surface. Lastly the surface discontinuity estimation was computed by comparing the polynomial approximation of data and polynomial approximation of the discretised model. The range segmentation by using this method is robust on sharp edges; however the strength of smooth edges or edges contaminated by noise can not be robustly estimated as the local character of the surface severely affected by the quantisation.

In general, a drawback of edge-based range segmentation algorithms is that although they produce well defined boundaries between regions in a clean data set, they tend to produce gaps between boundaries. Another problem of these methods is that the result of the range segmentation is quite sensitive to the window size of the edge operator. In addition, the effectiveness of these methods is greatly reduced when data is noisy. When the data points detected by an edge operator are not continuous and contaminated by noise, it is very difficult to link those data points as the one edge. Moreover, these data points are not able to detect coplanar surfaces that are separated by an occluding object and so the segmentation of multi structure and complex scenes using this method entails a significant amount of post processing work.
Region growing based range segmentation algorithms

Region growing range segmentation methods are more popular than edge detection based techniques. These algorithms rely highly on the existence of spatial coherency in the range data. Generally, region-growing based segmentation methods first estimate the feature vectors of each pixel. Then they combine the pixels that have similar feature vectors and simultaneously separate the pixels whose feature vectors are not alike [83, 95, 96]. The estimation of feature vectors can be obtained using different methods such as Seeded Region Growing (SRG) [97].

The BJ algorithm presented by Besl and Jain [20] in 1988, is a classic example of these techniques. This method first labels each data point based on its depth. The largest continuous region consisting of data points with the same label is then selected. An erosion operator is applied to the separated data points to ensure there is no outlier in the selection. In the next step, the algorithm considers that all data points belong to a planar surface. Based on this hypothesis, the region is grown by data points which can either leave or enter, based on the similarity of their labels and the goodness of their fit. If the hypothesis is not true, the algorithm considers all points to be on a quadratic surface and extends the seed using the region-growing method. This method is very sensitive to noise and requires a large initial region for optimal performance. To suppress the effect of noise in the BJ algorithm, Yang [98] proposed a similar method in a multi-resolution framework. In Yang’s technique, input data is at the lowest level of resolution and the resolution for each of the next levels are increased, incorporating a pyramid.

The UE algorithm introduced by Fitzgibbon et al. [99] in 1995, and the USF algorithm proposed by Hoover et al.[89] in 1996, are instances of region growing range segmentation methods. The first step of the UE method is to calculate the surface normal of each pixel, and depth discontinuity based on a predetermined normal and depth threshold. The Gaussian ($H$) and mean ($K$) curvature for each data point is then estimated. The sign of $H$ and $K$ together can determine the type of the surface (e.g. flat, pit, peak, valley, etc). Each data point and its eight neighbouring data with a similar label are grouped together to form an initial region. In the next stage, a region-growing method is performed to extend the preliminary segment. In this case, a data point can leave or enter the region, based on the angle of its normal vector and the perpendicular distance between the data point and the grown region. At the end, the surface is refitted and a boundary refinement is performed to preserve the boundaries.
The USF algorithm generally performs in a similar way to the UE algorithm. This method however tends to segment planar surfaces, so it does not require an H-K based segmentation to obtain curvatures and the boundary refinement process as with the UE algorithm.

A general drawback of region growing range segmentation techniques is that they rely on global thresholds which need to be predetermined for different applications and different levels of noise. In addition, the choice of preliminary seeds greatly affects the performance of these methods. In particular, when the initial seeds are located on a boundary or noisy data, the segmentation result will be corrupted. Moreover, range segmentation algorithms based on a region growing often produce distorted boundaries because the segmentation usually carried out at a regional level instead of a pixel level. The last two problems however are significantly improved by the UE algorithm. Furthermore, similar to the edge detection methods, they are not able to overcome occlusion problem that exists in multi structural scene; therefore the final segmentation stage involves heavy post-processing work.

3.3.3 Parametric fitting-based range segmentation algorithm

Parametric fitting-based range segmentation algorithms are generally designed to segment range images of multi-structural man-made objects that are well represented by geometric objects. An often adopted approach to this technique of range segmentation is to first fit surface models to range data and then estimate geometric parameters, such as surface normal and curvature [20]. In this context, the range segmentation problem is similar to robust statistical regression and can be called robust segmentation. The robust estimators used as the core of these segmentation methods are either adopted from the statistics community (e.g. Least Median of Square (LMedS) [100]), or innate to the computer vision field (e.g. Hough Transform (HT) [101], RANdom SAmple Consensus (RANSAC) [102]). There are two main differences between the robust regression algorithms in statistics and computer vision. The robust regression techniques in the statistics community are usually designed to apply to the data which has only one target distribution with an absolute majority [103, 104]. On the other hand, robust regression techniques used for computer vision tasks generally apply to the data that contain a number of populations with different distributions. In this
case, every population of data will be an outlier to other populations, called pseudo-outliers [105]. In addition, a population with absolute majority may not actually exist.

The Least Median of Squares (LMS) (1984) is probably the most well known example of robust estimators in the statistics community [100] and a foundation for most of the robust estimators that have been developed in computer vision [106]. This method itself is a modification to the Least Square (LS) estimator in that it replaces the sum, in the least sum of squares, with the median. The objective function that the LMS estimator optimises is described by the following formula:

$$\hat{\theta} = \arg \min_i (\text{med} r_i^2)$$

where $\hat{\theta}$ stands for the regression coefficient and $r$ is the residuals. There are however some limitations of directly applying this method to most computer vision problems [105, 107]. The main limitation of this method is the fact that its breakdown point is at most 50 per cent. This limitation means that the structure to be segmented should occupy the majority of the data set, which is more than 50 per cent of data points. In addition, this method assumes that the correct fit corresponds to the fit with the least median of squared residuals. This assumption is not true when the variance of the inliers is large and data includes multiple structures and clustered outliers.

To overcome the problems of the LMS estimator, a number of other estimators have been developed to solve computer vision problems. Minimum Unbiased Scale Estimator (MUSE) (1996) presented by Miller and Stewart [108] and Adaptive Least $K_{th}$ Order Squares (ALKS) (1998), proposed by Lee et al. [109] are good examples of these algorithms. The ALKS estimator minimises the $K_{th}$ order of the squared residuals and the MUSE estimator minimises an unbiased scale estimate of the ordered residuals. Both approaches are based on Least $K_{th}$ Order Squares (LKS) [104]. Starting with the whole data points, they first find the size of the biggest minority population through minimising some scaled measure of the variance of residuals. This scale is obtained by an initial LKS fit. Then again the LKS

---

1 Cluster outliers refer to those data points that are captured from an unwanted object in a scene. For instance, data points captured from a standing passenger in front of a large building produce a cluster of outliers in the data set, when the data of interest are those from the building exterior.
technique is used to find the correct structure. In order to solve the objective function of LKS, both methods rely on random sampling as it is used in LMS and Random Sample Consensus (RANSAC) (1981) [102] estimators. Both methods of ALKS and the MUSE estimator have a breakdown point of greater than 50 per cent so they can perform better than LMS at small scale discontinuities. Both approaches do however share the same limitations: The optimisation process for finding the optimum value for $K$ in both methods implicitly relies on having an accurate estimate of $K$ [110]. In addition, the scale estimation using ranked order statistics (as in LKS) assumes only one underlying distribution for the data. In this case, the scale cannot be accurately estimated for data of multi-structural objects.

In order to solve the above issues associated with ALKS and MUSE, Bab-Hadiashar and Suter proposed the Selective Statistical Estimator (SSE) (1998) [110, 111]. The SSE method is a simplified LKS-based estimator whereby the value of $K$ is not determined by a complex optimisation routine. In other words, the value of $K$ is defined by the user as the minimum acceptable size of the population one is interested in. When $K$ is determined, the rest of estimation procedure follows the LKS method. It means that the algorithm uses a random sampling technique to find the sub-sample that has the smallest $K$th order least squares. This will randomly detect one of the structures, which has a population of at least $K$ data points. In order to find all data points belonging to the selected structure, the residuals are scaled with respect to this structure using $K$th order statistics of the residuals. Although this method returns a good fit, similar to ALKS and MUSE, it does not necessarily produce a good inlier/outlier classification in a multi-structure data set. To overcome this problem, the authors of SSE presented the Modified Selective Statistical Estimator (MSSE) (1999) [111]. The MSSE method uses the same principal as SSE for finding the first fit. This method then finds the scale from the unbiased scale estimate without using the rank ordering statistics. Besides the cost of random sampling for large data sets, the other drawback of this algorithm is that it assumes the underlying model of every surface is known a priori therefore its application for the segmentation of range images of different surfaces is limited. The MSSE method is explained next in more detail as this method is used in the next three chapters as a rangescanner system evaluation tool (Chapter 4), an experimental range segmentor to describe challenges of large-scale range segmentation (Chapter 5) and as a robust estimator for the new range segmentation scheme introduced in this thesis (Chapter 6).
In order to improve surface (in particular, planar and quadratic) detection and the extraction from range data of a multi-structure scene, Gotardo et al. (2004) [112] have introduced a parametric range segmentation method. The robust estimator designed for the proposed segmentation is an extension of M-estimator Sample Consensus (MSAC) (2000), presented by Torr and Zisserman [113] (which is, in turn, an improvement on RANSAC). This segmentation method has two main stages: pre-processing and surface extraction. In the pre-processing stage, local orientation at each pixel is estimated and a set of coarse segments are generated using step and roof edge detection. Each pixel is then roughly classified as planar, curved or undetermined, to be used as surface model selection in the next stage. In the surface extraction stage, the previous segmentation is iteratively refined and then the surface model is selected, based on the previous classification. The robust estimator is then used to effectively identify inliers to be as seeds of region growing segmentation, which is in turn, identify the sub-region corresponding to the surface being extracted. Unlike MSAC, this estimator uses genetic algorithm to accelerate the optimisation process of robust surface fitting. The steps of this algorithm are illustrated in Figure 3.3.

Figure 3.3: Range segmentation into planar and quadric, adapted from [112] (a) pre-processing and (b) surface extraction
Wang and Suter have presented the Maximum Density Power Estimator (MDPE) (2003) \cite{114, 115} algorithm. The MDPE is a robust estimator to use as part of an effective range segmentation algorithm. This estimator applies density gradient estimation and nonparametric density estimation techniques for model fitting. This method optimises an objective function to measure both the density distribution of the data points in residual space as well as the size of the residual that corresponds to the local maximum of the density distribution. The MDPE is claimed to be able to tolerate more than 85 per cent of outliers.

In recent years (2003), a new class of projection based M-estimators (PbM-estimators) has been introduced by Meer et al. \cite{116-118}. These methods are based on finding the modes of kernel density estimates for the projections of data points on different directions. In this approach, the scale parameter in the cost function of the M-estimator is replaced by the band-width of the kernel density estimator, which is claimed to have a much weaker influence on the final result.

Most recently, Hosseinnezhad and Bab-Hadiashar (2007) \cite{119} presented a computationally efficient High Breakdown M-estimator (HBM) to solve the segmentation problem of multi-structural. Unlike most high breakdown robust estimators (e.g. \cite{102, 109, 111}) that use random sampling for objective function optimisation which is computationally cumbersome, HBM uses an iterative reweighed ‘Least Squares’ regression. In this method, the objective function is a differentiable equation for which a closed form updating formula is derived to search the parameter space. This estimator, instead of minimising the single $K$-th order statistics of the smallest squared residuals, minimises a smoothed window around the $K$-th order statistics of the smallest squared residuals. The experimental results on HBM show that this estimator is able to accurately extract the structures embedded in the data of a complex scene. In addition, the comparison of computational time of this method to the other estimators such as ASSC \cite{120}, MSSE \cite{111} and PbM \cite{116-118} shows the outperforming of this method as its search strategy converge fast.

The main difficulty of range segmentation using robust estimators is that the performance of these techniques relies highly on the correct selection of underlying surface model. This problem leads to a model selection criteria choosing from a model library \cite{121} \cite{122}. The effectiveness of these methods also largely depends on the accuracy of their scale estimates \cite{123}. In return, there are several advantages for parametric range segmentation methods. First, they overcome the occlusion problem, meaning that these methods (unlike
regional growth and edge detecting methods) are able to directly identify and extract occluded surfaces that belong to one specific geometric primitive. These methods would reduce a significant amount of the post-processing work involved in merging occluded surfaces. Moreover, unlike level set approaches in region growing methods that constrain segmentation boundaries explicitly, the parametric segmentation techniques are performed at the pixel level and therefore the embedded regions can be correctly segmented.

3.3.3.1 Modified Selective Statistical Estimator (MSSE)

This section describes the implementation of MSSE for segmentation applications. This method uses random sampling to determine a number of candidate fits, ranks these candidate fits by least $K$-th order residuals and estimates the scale from the best preliminary fit. The algorithm classifies inliers and outliers by using scale estimation. Figure 3.4 shows the flowchart of range segmentation using MSSE.

As shown in the flowchart, a value of $K$ is set (by the user) as the lower limit of the size of populations one is interested in. A localised data group inside the data space in which all the pixels appear on a flat plane is found using random sampling. A planar model with the least $K$-th order squared residuals is selected from the planar models and fitted to those samples. Residuals are then ranked for the selected fit.

For the accepted model, starting from $n = K$, the unbiased estimate of scale of noise is calculated using the smallest $n$ residuals:

$$\sigma_n^2 = \sum_{j=1}^{n} \frac{r_j^2}{n-p}$$  \hspace{1cm} (3.6)

where $r_j$ is the $j$th smallest residual and $p$ is the number of parameters in the model. The $(n+1)^{th}$ residual squared is weighed against a threshold set by the $\sigma_n^2$. If $r_{n+1}^2 > T^2 \sigma_n^2$, then the residual $r_{n+1}^2$ is considered as the first outlier residual. Then, those points whose squared residuals is greater than the threshold multiple of the scale of noise are rejected.
Chapter 3 — Range Image Segmentation

Figure 3.4: Robust segmentation using MSSE. The process has two parts of estimation and segmentation.
The value of the threshold $T$ is often considered to be 2.5 based on the desired level of significance in the normal distribution\(^2\) [104]. The equivalent characterisation of the point of transition from inliers to outliers occurs when:

\[
\frac{\sigma_n^2 + 1}{\sigma_n^2} > 1 + \frac{T^2 - 1}{n - p + 1}
\]  

(3.7)

The above procedure, from random sampling to outlier rejection, is repeated as long as the remaining data is large enough to hold the remaining segments and each time the segmentation process leads to a new segment containing all the inliers to the obtained fit (regardless of their geometrical location) to be generated. As a result, the algorithm has the advantage of detecting and resolving occlusion while segmenting the data. The above tasks are iteratively performed until the number of remaining data points become less than the size of the smallest possible region in the considered application. Details of implementation of MSSE are shown in Algorithm 1:

\(^2\) In Normal (Gaussian) distribution, the majority (\(\%50 + 1\)) of the samples lie within 2.5 times the standard deviation from 0.
Chapter 3 — Range Image Segmentation

### Algorithm 1 MSSE Algorithm

1. **Perform** random sampling:
   a) *Calculate* number of required samples (*nsample*), considering the size of smallest structure to be segmented (*K*) and the probability of success (*P*).
   b) *Generate* random samples
   c) *Choose* a random sample (3 data points for *p*=3)

2. **Calculate** residuals to the plane fitted to the sample.

3. **Sort** square of residuals (*r^2*) in ascending order

4. **Start** from *n=K* to calculate the unbiased scale estimate as:

   \[
   \sigma_n^2 = \frac{\sum_{j=1}^{n} r_j^2}{n - p}
   \]

   If *r^2_{(K+1)} < (T^2 \sigma_n^2)*
   
   **Calculate** \( \sigma_{new}^2 \) and **Continue**
   
   **elseif** *r^2_{(K+1)} > (T^2 \sigma_n^2)*
   
   Residual \( r^2_{(K+m)} \) is classed as the first outlier residual for the chosen sample
   
   **Record** inliers, outliers and estimated \( \sigma_{new}^2 \)

   **end**

5. **Repeat** steps 1-c to 5 for all random samples (*nsample* times)

6. **Find** the smallest estimated scale (\( \sigma_{new}^2 \)). Inliers associated with this scale belong to the desired segment

7. **Repeat** steps 1 to 7 for outliers associated with the smallest estimated scale

8. **Continue** until all desired segments are extracted (it depends on the value of *K*)
3.3.4 Combined methods

Many segmentation algorithms reported in the literature combine some aspects of region-based, edge-based and model-based approaches to overcome their individual drawbacks and to create an appropriate method for a desired application [124, 125]. An example of a combined segmentation algorithm is the WSU method. This algorithm first proposed by Hoffman and Jain [126], was then improved by Flynn and Jain [127] for 3D object recognition. The WSU algorithm has seven steps: labelling, estimating the surface normal, sampling, clustering, merging, separating planar surfaces and again merging. In the first step, the absolute difference between the depths of a range data and its eight neighbouring pixels are measured. If this measurement is greater than a predetermined threshold ($T_1$) for all points, the pixel is labelled as a jump edge point. In the second step, the surface normal of each non-jump edge point is calculated using an $n$-by-$n$ window. The algorithm then performs sampling on a regular grid to generate a data set less than 1,000 points in size. The 3D normal vector and 3D data of all points belonging to this data set are joined to generate a six-dimensional data space. In this case, each data point is assigned to the nearest corresponding cluster centre. The algorithm is designed to avoid assigning the same labels to regions that are not connected. In the fifth step, an edge-based merging algorithm aggregates the regions. The merging happens when the average angle between the surface normal of the data points on both sides of the edge is less than a predetermined threshold ($T_2$). Next, the planar surfaces are distinguished and separated from non-planar regions using a principal component procedure and the neighbouring co-planar surfaces are merged. The last two steps are iteratively repeated until all segments are separated.

The first weakness of this method is that the predetermined thresholds ($T_1$ and $T_2$) need to be precisely tuned for different scenes and different levels of noise. In addition, size and shape of the boundaries for each segment require an accurate assumption. Lastly, this method tends to under-segment the range image [89].

Another example of a combined segmentation algorithm is UB method presented by Jiang and Bunke [128]. This algorithm divides each scan-line into a number of 3D line segments (straight or curved) using a splitting method. Then, specified values called jump edge strength and crease edge strength are computed separately for each point (in four directions). A point is considered as an edge point if one of these measurements is greater than a pre-defined threshold. At the next step, a component labelling algorithm groups the
3D line-segments, based on a hypothesis-testing approach. In order to fill the gaps and detect under-segmentation, a verification test is also performed. This test starts with a planarity test. If the region is not planar it then tests the validity of a second order model. If this is successful, the region is accepted. Otherwise the edge points inside the region are expanded to fill the breakage. These steps are recursively performed until all the regions have been verified or they are too small to be regarded as separate regions. The evaluation tests of Powell et al. [129] and Hoover et al. [89] have shown that this algorithm performs well for the segmentation of planar surfaces, however it does not perform reasonably well for segmenting curved surfaces. To solve this problem, Boulanger et al. [130, 131] have present a region-based hierarchical range segmentation method with contour constrains. In this technique, the first level of hierarchy starts with an initial small partition of first order regions using Least Median of Squares (LMS) fitting algorithm. The algorithm then group the first order regions to the larger regions using a parametric compatibility function until an approximation limit is reached. In the next step, the method generalises the first order regions to the second order regions to generate corresponding regions with the highest order using Bayesian decision theory. Unlike parametric robust segmentation methods such as MSSE, this method needs to re-adjust the boundaries of each region by performing a geometric intersection (as most of region growing-based algorithms do).

### 3.4 Range segmentation of large man-made objects

Although the image segmentation of urban structures has been extensively studied for many years (e.g. [132-135]), their 3D range segmentation (especially data obtained by laser rangescanners) is a relatively new subject. Due to the lack of large-scale 3D data until recently, the majority of range segmentation techniques are devoted to the segmentation problem of 3D simulated data or range data of laboratory-size objects [89, 136]. Significant progress in 3D measurement technology and advancements in computer systems in recent years has led to the availability of accurate dense range data of large outdoor scenes. As an example, Figure 3.5 shows an image and a set of point cloud data obtained from the exterior of a large historical building in Melbourne, Australia. The availability of such data means that the production of automated urban models of whole buildings and streetscapes is emerging as one of the viable applications of three-dimensional data. To create such a model, a range segmentation algorithm is required to overcome the problems associated with
measurement uncertainties and structural complexities of range data of outdoor scenes. These problems will be elaborated upon in Chapter 5.

Figure 3.5: The front view of the Royal Melbourne Exhibition Building, Melbourne, Australia: (top) intensity image, (bottom) colour-coded range image.

Depending on the application and the type of objects of interest, there are different approaches to the segmentation of dense range images of outdoor scenes. The existing range segmentation methods of a large-scale outdoor scene can be classified into two main categories. The first class of approaches can be referred to as group segmentation. In these methods, the 3D dataset is considered as a collection of pre-defined classes of segments [137-140]. In other words, a learning method is employed to find the various instances of different classes of objects such as ground, trees, buildings, shrubs and crowds. Zhao and Shibasaki [137] initially classified range points of urban scenes into vertical lines, horizontal lines, non-vertical lines, vegetation and outliers. They then extracted each segment based on a hierarchical procedure. Anguelov et al. [138] has also developed a learning-based approach for the segmentation of complex scenes based on Markov Random Field (MRF) theory. In
Chapter 3 — Range Image Segmentation

this method, the scan points are labelled and weighted according to an appropriate object class (e.g. ground, building, tree and shrub) applicable to both indoor and outdoor range data. To reduce the computational cost associated with using MRF in the learning algorithm, Triebel et al. [139] have presented a new method named Associative Markov Networks (AMNs). Their technique is an extension to the work of Anguelov et al. [138], which uses a max-margin optimisation technique in the learning phase of the segmentation to adaptively reduce the training data. In this case, the training process is performed efficiently which, in turn, reduces the computational cost. Besides the cost, another advantage of this algorithm is that it does not require any tuning parameter. In the method of Haala et al. [77], a curvature based segmentation algorithm is employed to recognise tree species for forestry applications. Früh and Zakhor [5] have also presented a learning based segmentation technique which exploits the information embedded in both laser and intensity images. Wolf et al. [141] have employed a Markov Random Fields (MRF) technique to segment 3D terrain maps for autonomous navigation purposes. They classified an outdoor scene into navigable and non-navigable regions.

A general drawback of these methods is that they rely highly on the existence of distinguishing features embedded in the scene and therefore the learning phase (pre-segmentation phase) of the algorithm becomes very application dependant. In this case, the segmentation algorithm requires a lot of pre-processing work which in most cases are performed manually.

The second class of approaches which has been proposed particularly for the range segmentation of building exteriors is based on using architectural features. This class of approaches can be referred to as feature segmentation. Attributes such as vanishing points [63], parallelism of walls and orthogonality of edges [142] are employed to extract linear features of buildings. The method of Stamos and Allen [63] is an instance of this group of segmentation. They present a two level systematic approach to recover the geometry and texture of large-scale buildings by extracting features embedded in the dense range images of urban scenes. At the first level, 3D curves of a finite extent are extracted using a region-based method, then 3D linear features are segmented using an edge-detection based segmentation algorithm. Another example of this class of segmentation is proposed by Cantzler et al. [142]. In this method, the input is a 3D model consisting of vertices, edges and triangulated meshes. The architectural features such as planes and edges are first extracted
from the 3D model using a RANSAC approach. The RANSAC algorithm starts with random sampling of triple vertices to fit to a set of random planes. The distance of the triangle centroid to the hypothesis plane is then calculated. Those triangles where their distance to the hypothesis plane is less than threshold are considered to belong to the hypothesis plane. The edges also are extracted by filtering the 3D model. A tree search strategy is then used to automatically find the relationship between the planes and parallel edges are grouped using a clustering method. To this purpose, planes matched against a semantic net of a generic house and edges with a similar orientation are clustered together as illustrated in Figure 3.6.

![Figure 3.6: Semantic net model for the architectural scene (adapted from [142]). Nodes represent the model entities (e.g. walls, roofs, floors) and are linked by architectural relationships (e.g. parallelism and orthogonality).](image)

In the last stage of segmentation, parameters of the model are obtained and optimised through the Downhill Simplex optimisation algorithm. When the parameter vector for the optimised model is found, the vertices are projected onto the features. The result is a model
with aligned features (e.g. parallel walls). This method of segmentation is independent of the sensor data and is applicable to the range image as well as the intensity image.

Han et al. [143] introduced a stochastic algorithm based on the Markov Chain Monte Carlo (MCMC) theory for model-based segmentation of range data obtained by 3D laser rangescanner and its associated reflectance data. To reduce the speed of Markov Chain searching, this method employed a coarse-to-fine COF approach together with a data-driven strategy such as edge detection and clustering. Later, they extended this method to a stochastic jump-diffusion algorithm using a Bayesian framework [144].

A different class of outdoor range segmentation method is presented by Yuan et al. [145]. This method employs a sketch based user interface in order to segment 3D data of outdoor environments. It partitions a user-defined part of the 2D image in the projection plane of the camera and uses the result of this stage for range data segmentation of the same scene.

In general, these techniques rely on the existence of distinguishable features embedded in the scene, which their availability is often application dependent. In addition, they mostly employ edge detection or region growing techniques which are not robust and require significant post-processing to overcome the occlusion problem and the complexity of 3D images of large-scale objects.

### 3.5 Chapter review

This chapter has introduced the range segmentation task in general and outlined a well known evaluation method of a range segmentation algorithm. It has also presented the classification of existing 3D segmentation techniques, particularly those that have been used for range data of large-scale objects. This provides the necessary background for presenting the new range segmentation scheme for large-scale applications.

The review of existing range segmentation algorithms for large-scale applications shows that although these techniques are generally able to extract large segments of various buildings, they are not designed to detect fine details (e.g. decorative features, stairs and staircases) from a range image. Moreover, these techniques rely on the existence of distinguishable features embedded in the scene, where their availability is often application
dependent. In addition, these techniques mostly employ edge detection or region growing techniques which are not robust and require significant post-processing to overcome the occlusion and complexity problem in 3D images of large-scale objects. It was shown that, among range segmentation algorithms, those that rely on parametric robust estimators have the advantage of being able to resolve the occlusion problem and produce more accurate surface parameters while preserving boundaries.
Chapter 4

Laser Rangescanner System

4.1 Introduction

This chapter describes the design, calibration and validation of a low cost long-range laser rangescanner system. The system is later used for conducting part of the required experiments supported this research. Section 4.2 explains the stages of design and implementation of the rangescanner system including hardware and software design, sensor fusion and system calibration. Section 4.3 discusses the sources of error in the range measurement obtained by the experimental rangescanner system followed by the system characterisation. To validate the performance of the laser rangescanner system, two methods of evaluation were implemented and their results are detailed in Section 4.4.

4.2 Laser rangescanner system design

Automatic three-dimensional data acquisition devices (often called 3D scanners) allow the building of highly accurate models of real world objects in a relatively time effective manner. This chapter is an attempt to experiment with this technology in a particular
application context: the data acquisition of large building exterior. The specific requirements of this application domain are: medium to high accuracy, ease of use, affordable cost of the scanning device and safe operation for both the operator and the scanned building artefacts. Considering these requirements, a low cost 3D scanner is designed, based on the active laser rangefinding method. Active laser rangefinding is generally the preferred method for large-scale/outdoor applications for two main reasons: firstly, it provides a direct range computation while reducing the computational cost. Secondly, because this method is not as sensitive to the ambient illumination as the passive methods, the image processing is significantly simpler. This issue is more important in outdoor scenes where the ambient light is not always controllable or predictable.

\subsection{Hardware architecture}

The proposed 3D laser rangescanner device is composed of a laser rangefinder (LRF) (model ILM300HR\textsuperscript{1}, manufactured by MDL Measurement Devices Ltd) (see Appendix \textit{A} for laser safety consideration and Appendix \textit{B} for the mechanical detail of the LRF), a controllable pan-tilt unit (PTU) (model PTU-46-70, manufactured by Directed Perception) (see Appendix \textit{C} for the mechanical detail of the PTU) and a digital video camera (model 400 DVi, manufactured by Canon). As shown in Figure 4.1, two vision sensors (LRF and camera) are rigidly attached on the same platform. To simplify the data fusion procedure, the camera is modelled as a pinhole camera and its optical axis is aligned with the rangefinder optical axis so that the coordinate system of the camera and LRF are parallel, where Y-axes are coaxial with a fixed distance of \(d=25\) centimetres (Figure 4.1 (middle)). Both the laser rangefinder and pan-tilt unit are controlled through serial communication (RS232). The centre of gravity of the rangefinder is placed on the nodal point\textsuperscript{2} of the PTU. The LRF is able to measure distance up to 300 meters at the typical rate of 1 KHz and accuracy of 30cm (0.1\% of the maximum measured distance) (see Table 4.1 for LRF factory specifications). The digital video camera captures intensity images of size \(720 \times 576\) (pixels). The PTU provides maximum horizontal and vertical field of view (FOV) (\(\Phi, \theta\)) of \([-159^\circ, 159^\circ]\) and \([-30^\circ, 30^\circ]\) respectively (see Table 4.2 for PTU factory specifications). It has a maximum resolution of 0.0128\(^\circ\) where resolution of the scanner device is directly related to this angle.

\textsuperscript{1} The ILM300HR is now modified by \textit{MDL} and presented under new name, LaserAce\textsuperscript{®} IM-HR.

\textsuperscript{2} Nodal point of the PTU is the intersection of pan and tilt axis.
Figure 4.1: Laser rangescanner system: (top) Overall system architecture (middle) System alignment ($r \gg d$, $O_c$ is the centre of camera coordinate system and $O_l$ is the centre of LRF coordinate system) (bottom) Experimental system set-up
Chapter 4 — Laser Rangescanner System

In the laser rangescanner system, raw range data is measured in the Spherical coordinate system \((r_{ij}, \Phi_{ij}, \theta_{ij})\)\(^3\) and is then converted to the Cartesian coordinate system using the following equation:

\[
\begin{bmatrix}
X_{ij} \\
Y_{ij} \\
Z_{ij}
\end{bmatrix} = r_{ij}
\begin{bmatrix}
\sin (90 - \theta_{ij}) \cos (\Phi_{ij}) \\
\sin (90 - \theta_{ij}) \sin (\Phi_{ij}) \\
\cos (90 - \theta_{ij})
\end{bmatrix}.
\]

(4.1)

\(^3\) Hereafter, \((i,j)\) refers to indices of the horizontal and vertical coordinates of each point in the range image and \(r_{ij}\) is the corresponding range measured by the LRF.

### Table 4.1: Technical specification of LRF (http://www.mdl.co.uk)

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>5m to 300m</td>
</tr>
<tr>
<td>Accuracy</td>
<td>30cm (typically)</td>
</tr>
<tr>
<td>Resolution</td>
<td>10cm</td>
</tr>
<tr>
<td>Measuring Rate</td>
<td>1000Hz</td>
</tr>
<tr>
<td>Wavelength</td>
<td>905mm</td>
</tr>
<tr>
<td>Laser Eye Safety Class</td>
<td>Class 1</td>
</tr>
<tr>
<td>Weight</td>
<td>950g</td>
</tr>
<tr>
<td>Power</td>
<td>12 to 48 volts DC, &lt;5watts</td>
</tr>
<tr>
<td>Operating Temp</td>
<td>-10°C to +60°C</td>
</tr>
</tbody>
</table>

### Table 4.2: Technical specification of PTU (http://www.dperception.com/)

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Rated Payload</td>
<td>4.08 Kg</td>
</tr>
<tr>
<td>Maximum Speed</td>
<td>Over 60 degrees /sec</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.012857° (0.771 arc minutes)</td>
</tr>
<tr>
<td>Tilt Range (approximately)</td>
<td>Min. 31° up and 47° down</td>
</tr>
<tr>
<td>Pan Range (approximately)</td>
<td>± 150°</td>
</tr>
<tr>
<td>Input Voltage</td>
<td>8 to 30 volts DC (unregulated)</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>13W @ full-power load</td>
</tr>
<tr>
<td></td>
<td>6W @ low-power load</td>
</tr>
<tr>
<td></td>
<td>1W @ holding power-off mode</td>
</tr>
</tbody>
</table>

---

\(^3\) Hereafter, \((i,j)\) refers to indices of the horizontal and vertical coordinates of each point in the range image and \(r_{ij}\) is the corresponding range measured by the LRF.
4.2.2 Software architecture

The laser rangescanner software consists of three main parts including system calibration (pre-processing), data acquisition, and data fusion and visualisation (post-processing). Figure 4.2 shows the structural design of the software. The required code for the system calibration and data acquisition is implemented in C language (Microsoft Visual Studio 6.0) and data processing, including data fusion and visualisation are implemented in MATLAB language (MATLAB 7.1).

Figure 4.2: Laser rangescanner system: Software architecture
4.2.3 Automatic system calibration

Calibration is a common term for any kind of alignment procedure or compensation routine to improve the accuracy of a measurement system. For the 3D scanner constructed in this project a few system parameters are initialised by a general automatic calibration procedure. The resulting parameters confine the scene to be scanned (start and end position for pan and tilt motion) and define the resolution of the data acquisition (pan and tilt step size in degree), based on the relationship between the pixels in the range and intensity image (pixel ratio). This section describes the steps of the self-calibration procedure for the experimental rangescanner system.

A class I infrared (IR) laser with wavelength of 905 nm is used in the LRF device. The IR lasers are not visible to the human eye but they are more sensitive and accurate than the visible lasers. To be able to trace the IR laser spot with the digital camcorder, an optical IR band-pass filter (B+W 092) is mounted on the camera lens (see Appendix D for filter characterisation). The band-pass filter blocks visible light up to 650 nm, and passes only 50 per cent of the radiation below the 700 nm. From 730 nm to 2000 nm, transmission is greater than 90 per cent.

Figure 4.3 shows the rangescanner set-up for general calibration. To be able to automatically calibrate the system, a client-server program using the DirectShow platform is implemented. The server program creates a Transmission Control Protocol (TCP) stream socket to be able to listen to the client. The incoming data received from the client are displayed and stored and a received confirmation message is sent back to the client. Here, the data are the coordinates of the projected laser beam into the image plane. These coordinates are obtained from a $5 \times 5$ median filter and implemented in the client side.
The rangescanner, in calibration mode, starts to capture a live video of the scene while it is tracking the laser beam. The calibration tool is a 110cm × 110cm non-glossy whiteboard which is located perpendicular to the rangescanner system at a certain distance from the camera. The laser beam is tracked by an appropriate infrared filter that is mounted on the camera lens. The laser beam first initialised to be at its home (original) position \( P_{\text{home}} (X_{\text{home}}, Y_{\text{home}}, Z_{\text{home}}) \). In this stage, the projection of the laser beam at the home position into the image plane is estimated by the calibration routine. Once the coordinate of the image plane is recorded as \( p_{\text{home}} (x_{p,\text{home}}, y_{p,\text{home}}) \), the laser beam is shifted horizontally (or vertically) to the second known position \( P_k (X_k, Y_k, Z_k) \). The projection of the laser beam at this position, into the image plane, is then calculated and recorded as \( p_k (x_{p,k}, y_{p,k}) \). As the laser footprint is larger than a pixel in the image plane, a 3 × 3 median filter is used to estimate the exact position of the centre of the laser beam in the image plane. Based on the desired field of view and the angular resolution of the range image, a set of parameters including a start and end position of the pan and tilt movement and Pixel Ratio are obtained from the system.

Figure 4.3: System set-up for automatic system calibration procedure
calibration. Figure 4.4 shows the relationship between the range and intensity images in the automatic calibration mode.

The relationship between the resolution of range image and intensity image is obtained from the following equation:
Chapter 4 — Laser Rangescanner System

\[
\alpha = \frac{(P_k - P_{\text{hom}e})}{(x_{pk} - x_{\text{phom}e})} \text{ or } \frac{(y_{pk} - y_{\text{phom}e})}{(y_{pk} - y_{\text{phom}e})}. \quad (4.2)
\]

Based on the value of \( \alpha \) and the desired size of range image, the start and end positions of LRF pan and tilt movement are then calculated. In all experiments, the home position is considered to be at \( X_{\text{home}} = Y_{\text{home}} = 0 \) and \( Z_{\text{home}} \) which is equal to the distance measured at that point.

Table 4.3 presents the parameters determined by the system calibration procedure for the experimental laser rangescanner at the distance of 7 meters. The default size of the range image is set to be \( 715 \times 400 \) (points) and the default angular resolution of the range image is set to be \( 0.1^\circ \) (the best angular resolution of the system is around \( 0.0128^\circ \)). The Pixel Ratio shows the relationship between the number of pixels in the range image and intensity image. For instance, a Pixel Ratio of five means that there are five points in the range image associated with every one pixel in the intensity image. This value needs to be accurate and is critical for the range and intensity data fusion procedure.

<table>
<thead>
<tr>
<th>Table 4.3: System calibration parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel Ratio</td>
</tr>
<tr>
<td>Start Pan Position</td>
</tr>
<tr>
<td>End Pan Position</td>
</tr>
<tr>
<td>Start Tilt Position</td>
</tr>
<tr>
<td>End Tilt Position</td>
</tr>
</tbody>
</table>

It is necessary to note that the calibration outcome, reported in Table 4.3, was obtained at the distance of 7 meters from the rangescanner system and was used for the entire experiments with good degree of accuracy. The accuracy of calibration will be explained in Section 4.2.4.3 in more detail. Based on the experimental result, there is a close agreement between the measured and estimated value of the calibration in a range between 16 to 32
meters. The experiments also show that in larger distance the laser footprint became larger and so the size of the soft filter to process calibration should be considered larger which reduces the accuracy of calibration by itself. To avoid this issue and for the ease of data acquisition, it was decided to locate the calibration tool at 7 meter from the system.

4.2.4 Range and intensity data fusion

In order to obtain a rich six dimensional data set of large building exteriors, volumetric and intensity data are fused (integrated/registered). In fact, every point in the range image \((X_{ij}, Y_{ij}, Z_{ij})\) is assigned to its corresponding point in the intensity image \((x_{ij}, y_{ij})\) (inverse mapping). In the experimental rangescanner system designed for this project (Figure 4.1), the laser rangefinder and camera are rigidly attached to the same platform. This ensures that the position and orientation of the LRF and camera are fixed with respect to each other. Therefore, the range and intensity data fusion can be simplified to an inverse mapping (camera calibration) problem. In this method, a pinhole camera model is utilised in which radial and tangential distortions are compensated. To find the corresponding points in the range and intensity image, the image formation process is reversed to project the intensity image back to the range image, based on the camera parameters and position estimation.

The next subsections describe the most common camera models and present the camera calibration method used for range and colour fusion in this project.

4.2.4.1 Camera models

In order to map the intensity image into the range image (inverse mapping), it is required to compute geometric information from an intensity image. Therefore, a camera model is required to be used and its parameters tuned. A camera model is described by its intrinsic and extrinsic parameters. The intrinsic or internal camera parameters, including the optical and electronic properties of the camera, are listed as \( (f, r, c_x, c_y, s, k) \) where \( f \) represents the camera focal length, \( r \) is the aspect ratio of the pixels, \( c_x \) and \( c_y \) are coordinates of the principal (or focal) point, \( s \) is skew coefficient (the angle between \( x \) and \( y \) axes in the image plane) and \( k \) describes the lens distortion. Those parameters that describe the geometric relationship between the camera and the scene (rotation and translation matrix) are named
extrinsic or external camera parameters. A simple camera model is shown in Figure 4.5.

Let the coordinates of a 3D world point $P_{ij}$ be denoted as $(X_{ij}, Y_{ij}, Z_{ij})$ in the world coordinate system and as $(X_{cij}, Y_{cij}, Z_{cij})$ in the camera coordinate system. The projection of point $P_{ij}$ onto the image plane is denoted by vector $u_{ij} = [x_{ij}, y_{ij}]^T$ with the corresponding homogeneous coordinate $u_{ij,h} = [x_{ij}, y_{ij}, 1]^T$ (see Appendix E for more detail on Homogeneous Coordinates).

In a simple form, the relationship between points $P_{ij}$ and $u_{ij}$ in the camera coordinate system can be defined as a normalised image projection matrix given by:

$$M_{n,ij} = \frac{1}{Z_{cij}} \begin{bmatrix} X_{cij} \\ Y_{cij} \end{bmatrix} = \begin{bmatrix} x_{ij} \\ y_{ij} \end{bmatrix}$$ (4.3)

Figure 4.5: Intrinsic camera parameters in a pinhole camera model. $\Phi$ and $R$ are focal and image planes, respectively. $c$ is the camera principle point and $f$ is the camera focal length. $(O_c, X_c, Y_c, Z_c)$ is the coordinate system of camera and $(c, u, v)$ is the coordinate system in the image plane.

There are several algebraic and matrix representations of a camera model (e.g. [27, 71, 146]). In order to simplify mathematical formulation of the relationship between 3D coordinates of objects and their corresponding values in the image coordinate, most of the
existing camera models approximate the real camera projection to the ideal pinhole camera. In such a camera model, the principal point of the image plane is assumed to be the centre of the image and the value of the aspect ratio is assumed one. In addition, the skew coefficient is assumed 90° and the lens distortion is ignored. The most common methods of mapping 3D to 2D using an ideal pinhole camera model are: projective, perspective, affine, weak-perspective and orthographic [147, 148].

The **projective** camera model can be described by Equation 4.4 [149] below:

\[
\begin{bmatrix}
    x_1 \\
    x_2 \\
    x_3
\end{bmatrix} =
\begin{bmatrix}
    T_{11} & T_{12} & T_{13} & T_{14} \\
    T_{21} & T_{22} & T_{23} & T_{24} \\
    T_{31} & T_{32} & T_{33} & T_{34}
\end{bmatrix}
\begin{bmatrix}
    X_1 \\
    X_2 \\
    X_3 \\
    X_4
\end{bmatrix}
\]

(4.4)

where \([x_1, x_2, x_3]^T\) and \([X_1, X_2, X_3, X_4]^T\) are homogeneous coordinates related to \(x\) and \(X\) by \((x, y) = (x_1/x_3, x_2/x_3)\) and \((X, Y, Z) = (X_1/X_4, X_2/X_4, X_3/X_4)\). The transformation matrix of projective camera, \(T\), has 11 degrees of freedom since only the ratios of elements \(T_{ij}\) are important.

The **perspective** camera model is a simplified form of the projective model, where the leftmost \(3\times3\) sub-matrix of \(T\) is a rotation matrix with its third row scaled by the inverse focal length, \(1/f\). The projection matrix is then in the following form:

\[
T_p = \begin{bmatrix}
    1 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 \\
    0 & 0 & 1/f & 0
\end{bmatrix}
\]

(4.5)

which gives the equation:

\[
\begin{bmatrix}
    x \\
    y
\end{bmatrix} = f \begin{bmatrix}
    X \\
    Z
\end{bmatrix} Y.
\]

(4.6)
Equation 4.6 shows that each point is scaled by its depth and all projection rays converge to the optical centre.

The affine camera model is also a special form of the projective camera model by constraining the matrix $T$ such that $T_{31} = T_{32} = T_{33} = 0$ to preserve the parallelism. The transformation matrix $T$ has 8 degrees of freedom. The 3D–2D affine mapping takes from:

$$x = MX + t$$  \hspace{1cm} (4.7)

where $M = [M_{ij}]$ is a $2 \times 3$ matrix with elements $M_{ij} = \frac{T_{ij}}{T_{34}}$ and $t = (T_{14}/T_{34}, T_{24}/T_{34})$ is a 2-vector.

The weak-perspective camera model is a combination of the perspective and affine camera model, where $M$ forms a uniformly scaled rotation matrix. The $T$ is shown as:

$$T_{WP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{Z_{ave}/f}{Z} \end{bmatrix}$$  \hspace{1cm} (4.8)

thereby:

$$M_{WP} = \frac{f}{Z_{ave}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z_{ave}} \begin{bmatrix} X \\ Y \end{bmatrix}.$$  \hspace{1cm} (4.9)

The Equation 4.9 is a simplified form of the perspective equation with point depths $Z_i$, replaced by an average constant depth $Z_{ave}$. The weak perspective model is used when the field of view is small and is valid when the average variation of the depth of the object along the line of the sight ($\Delta Z$) is much smaller than the $Z_{ave}$.  

74
The orthographic camera model is a simplified affine camera model, where $M$ represents the first two rows of a rotation matrix. The transformation matrix of orthographic camera is:

$$T_{\text{orth}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.10)$$

thereby:

$$M_{\text{orth}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix}. \quad (4.11)$$

When high accuracy is required, an ideal pinhole camera model needs to be extended to correct lens distortion. There are two types of lens distortion in real camera: tangential and radial [150]. Radial distortion is because of the shape of the real lens whereas the tangential distortion is mainly due to the process of lens manufacturing. In practice, the lens distortion is generally dominated by radial distortion [67, 151] and so the complex camera models can be designed to correct radial lens distortion. Radial lens distortion is symmetric [150] and may cause the image to minimise or bulge toward the centre. Also, because the surface of the real lens is slightly curved, each point in the image plane corresponds to a set of rays from the front side of the lens, all converging to the focal point. Therefore, a distinctive principle point cannot be defined for such a camera model (Figure 4.6). The lens distortion is more noticeable in wide camera lenses and can be ignored when using a small camera aperture.
4.2.4.2 Camera calibration

The aim of camera calibration is to estimate the intrinsic and extrinsic parameters of the camera to find the geometric relationship between the location of the points in the world coordinate system and the camera coordinate system. A number of camera calibration methods have been presented in both photogrammetry (e.g. [153, 154]) and computer vision literature (e.g. [67, 150, 155-159]). As the camera used in this project is a camcorder with an extremely wide angle, the lens distortion is more noticeable and the camera calibration methods have to be able to compensate for the lens distortion. Among the camera calibration methods, Bouguet’s camera calibration toolbox [159] is used here, as this model considers radial components of the distortion model up to the fifth order and can accurately estimate the skew factor.

The internal camera model that has been used in this method is very similar to the method presented by Heikkilä [156, 160]. To obtain camera parameters, 15 photographs of the calibration board (a flat checkerboard pattern) are taken from different views. The camera calibration toolbox is implemented in MATLAB and the camera parameters for a (720 × 576) Canon camcorder used in this project, computed as shown in Table 4-4. For complete report on the camera calibration result generated by MATLAB code, see Appendix F.
Table 4.4: Camera calibration parameters: estimated vs. expected

<table>
<thead>
<tr>
<th>Camera Parameters</th>
<th>Computed (Real Camera)</th>
<th>Expected (Ideal Pinhole Camera)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_x )</td>
<td>352.900704875125830</td>
<td>360.00</td>
</tr>
<tr>
<td>( c_y )</td>
<td>270.870434347690720</td>
<td>288.00</td>
</tr>
<tr>
<td>( f_x )</td>
<td>968.985461566317550</td>
<td>1000.00</td>
</tr>
<tr>
<td>( f_y )</td>
<td>1062.876259169658400</td>
<td>1000.00</td>
</tr>
<tr>
<td>( k_1 )</td>
<td>-0.179995453454130</td>
<td>0.00</td>
</tr>
<tr>
<td>( k_2 )</td>
<td>-0.141294595257734</td>
<td>0.00</td>
</tr>
<tr>
<td>( k_3 )</td>
<td>-0.003449635043187</td>
<td>0.00</td>
</tr>
<tr>
<td>( k_4 )</td>
<td>-0.002026472282628</td>
<td>0.00</td>
</tr>
<tr>
<td>( k_5 )</td>
<td>0.00000000000000000000</td>
<td>0.00</td>
</tr>
<tr>
<td>( s )</td>
<td>90.00000000000000000000</td>
<td>90.00</td>
</tr>
</tbody>
</table>

The distortion coefficients are indicated as \( k_1 \) to \( k_5 \). The skew coefficient, \( s \), determines the angle between the \( x \) and \( y \) axes. \( c_x \) and \( c_y \) are the coordinates of the principal points and is consistent with the image centre of projection affected by radial distortion. Next, the computed parameters are used to map range data into intensity data.

As shown in Figure 4.1 and 4.5, every point in the camera coordinate system is denoted by \((X_{cij}, Y_{cij}, Z_{cij})\). Similarly, the coordinate system of the laser rangefinder is assumed to be the world coordinate system and every point on that coordinate system is denoted by \((X_{ij}, Y_{ij}, Z_{ij})\). Equation 4.12 shows the relationship between the LRF and camera coordinate system:

\[
\begin{bmatrix}
X_{c_{ij}} \\
Y_{c_{ij}} \\
Z_{c_{ij}}
\end{bmatrix} = \begin{bmatrix}
X_{ij} \\
Y_{ij} + d \\
Z_{ij}
\end{bmatrix} .
\]  

(4.12)

Considering lens distortion, the real normalised image coordinates are defined as a distorted normalised image projection matrix \((M_{ij})\) and given by:
\[
M_{d,ij} = (1 + k_1 r_{ij}^2 + k_2 r_{ij}^4 + k_3 r_{ij}^6)M_{n,ij} + T_{d,ij}
\] (4.13)

where \( k_1 \) to \( k_5 \) are distortion coefficients and \( r_0 \) and \( T_{d,ij} \) (tangential distortion) are given as:

\[
r_{ij}^2 = x_{ij}^2 + y_{ij}^2
\] (4.14)

and

\[
T_{d,ij} = \begin{bmatrix}
2k_3 x_{ij} y_{ij} + k_4 (r_{ij}^2 + 2x_{ij}^2) \\
k_3 (r_{ij}^2 + 2y_{ij}^2) + 2k_4 x_{ij} y_{ij}
\end{bmatrix}
\] (4.15)

Once distortion is applied, the pixel coordinates of the projection point \( P_{ij}(X_{ij}, Y_{ij}, Z_{ij}) \) on the image plane \( u_{ij}(x_{ij}, y_{ij}) \) are calculated as:

\[
\begin{align*}
x_{ij} &= f c_x (M_{dx,ij} + s \times M_{dy,ij}) + c_x \\
y_{ij} &= f c_y \times M_{dy,ij} + c_y
\end{align*}
\] (4.16)

where

\[
M_{d,ij} = \begin{bmatrix}
M_{dx,ij} \\
M_{dy,ij}
\end{bmatrix}
\] (4.17)

The skew coefficient \( s \) is zero according to the camera calibration. As a result, Equation 4.16 is simplified to:
To obtain a rich set of data including geometry (XYZ) and intensity (RGB), the above equation together with a linear interpolation function is applied to the range and intensity data, which are captured by the developed rangescanner system. The experimental result for this system is presented in Section 4.3.

\[
\begin{bmatrix}
  x_{ij} \\
  y_{ij}
\end{bmatrix} =
\begin{bmatrix}
  f_x \\
  f_y
\end{bmatrix} \times
\begin{bmatrix}
  M_{dx,ij} \\
  M_{dy,ij}
\end{bmatrix} +
\begin{bmatrix}
  c_x \\
  c_y
\end{bmatrix} \quad (4.18)
\]

4.2.4.3 **System calibration verification for fusion**

The data fusion algorithm is heavily influenced by the camera calibration parameters. Therefore, finding a solution to evaluate the accuracy of the camera calibration is important for determining the accuracy of the proposed data fusion. For this purpose, an experimental set-up is used to measure the quality of camera calibration parameters. For various points, at different distances from the camera, the coordinates of the laser beam in the image plane, \( P(x_p, y_p) \), are calculated and recorded using the infrared filter and the client-server software as described in Section 4.2.3. At the same time, the above coordinates are directly calculated based on the geometry of the points and obtained by the rangescanner system (Equation 4.18). Table 4.5 shows the calculated and measured value of \( x_p \) and \( y_p \) (in pixel) at eight different distances from the camera. In addition, maximum and minimum calculated value for \( x_p \) and \( y_p \) is calculated based on the errors obtained from camera calibration result (Appendix F) and presented in the table. A plot of the result for the values of the laser beam coordinate in the direction of y axis is illustrated in Figure 4.7. The plot shows that there is a close agreement between the laser beam coordinates obtained by observation and direct calculation and the results are within the uncertainty bounds (Appendix F and [159]) computed from camera calibration parameters.
Table 4.5: Calibration validation result for eight distances

<table>
<thead>
<tr>
<th>Range (m)</th>
<th>$x_p$ calculated</th>
<th>$x_p$ measured</th>
<th>$y_p$ calculated</th>
<th>$y_p$ measured</th>
<th>$y_{p,\text{max}}$</th>
<th>$y_{p,\text{min}}$</th>
<th>$y_{p,\text{max}}$</th>
<th>$y_{p,\text{min}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>352.8997</td>
<td>324</td>
<td>111.2866</td>
<td>132</td>
<td>368.6364</td>
<td>317.1630</td>
<td>131.043</td>
<td>91.258</td>
</tr>
<tr>
<td>18</td>
<td>352.9001</td>
<td>322.5</td>
<td>129.0301</td>
<td>135</td>
<td>368.6374</td>
<td>317.1627</td>
<td>149.070</td>
<td>108.775</td>
</tr>
<tr>
<td>21</td>
<td>352.9004</td>
<td>324</td>
<td>149.3061</td>
<td>153</td>
<td>368.6382</td>
<td>317.1625</td>
<td>169.667</td>
<td>128.787</td>
</tr>
<tr>
<td>23</td>
<td>352.9005</td>
<td>325</td>
<td>159.8837</td>
<td>163</td>
<td>368.6385</td>
<td>317.1625</td>
<td>180.411</td>
<td>139.225</td>
</tr>
<tr>
<td>26</td>
<td>352.9006</td>
<td>325</td>
<td>172.6973</td>
<td>174</td>
<td>368.6387</td>
<td>317.1624</td>
<td>193.424</td>
<td>151.868</td>
</tr>
<tr>
<td>28</td>
<td>352.9006</td>
<td>325.5</td>
<td>179.7135</td>
<td>184</td>
<td>368.6388</td>
<td>317.1624</td>
<td>200.549</td>
<td>158.790</td>
</tr>
<tr>
<td>30</td>
<td>352.9006</td>
<td>327</td>
<td>185.7938</td>
<td>189</td>
<td>368.6389</td>
<td>317.1623</td>
<td>206.722</td>
<td>164.788</td>
</tr>
<tr>
<td>330</td>
<td>352.9006</td>
<td>327.5</td>
<td>193.5318</td>
<td>196</td>
<td>368.6390</td>
<td>317.1623</td>
<td>214.579</td>
<td>172.421</td>
</tr>
</tbody>
</table>

Figure 4.7: Plot of calibration validation result in $y$ direction.

4.3 Laser rangescanner system experimental results

This section presents the results of data acquisition and fusion using the proposed laser range scanner system. The early experiments, shown in Figure 4.8 and Figure 4.9, were mainly conducted in order to assess the design of the proposed system in terms of timing, power consumption and the correctness of the generated data for general evaluation of the software and hardware of the system. Figure 4.8 shows the intensity image and the colour-coded
range image of the first data set captured by the experimental rangescanner system. In this experiment, data is captured from objects located up to 20 metres from the laser rangescanner system. Figure 4.9 shows the intensity image and the colour-coded range image of a building exterior located at Swinburne University of Technology, Melbourne, Australia. As observed in this figure, the laser rangescanner system is unable to capture correct data of very fine details such as tree branches. In addition, the system fails to obtain data of window glasses, as these data are read by the laser rangefinder as out-of-limit data.
Figures 4.10(a) and 4.11(a) show the intensity images, captured from the Melbourne Royal Exhibition Building and the building of the Shrine of Remembrance, respectively. The photos are taken by the camcorder installed on the rangescanner system. Figure 4.10 (b) shows the corresponding colour-coded range image obtained from the front side of the Royal Melbourne Exhibition Building (a world heritage listed site) and Figure 4.11 (b) presents the corresponding colour-coded range image taken from the front side of the building of the Shrine of Remembrance (a significant building in Melbourne).
Figure 4.10: Royal Melbourne Exhibition Building: (a) Intensity image (720×576 pixels) (b) Colour-coded range image (715×400 pixels) (c) Range and intensity data fusion (720×400 pixels). The areas coloured in white by the fusion algorithm shows that algorithm could not find any valid range or intensity data.
Figure 4.11: The Shrine of Remembrance, Melbourne, Australia: (a) Intensity image (720×576 pixels) 
(b) Colour-coded range image (715×400 pixels) (c) Range and intensity data fusion (720×400 pixels). 
The areas coloured in white by the fusion algorithm shows that algorithm could not find any valid 
range or intensity data.
The parts of range images that appear in navy indicate the areas for which the rangefinder has not been able to measure the depth. These areas include: very close objects (less than 3 metres), the sky or objects located at distance of more than 300 metres, transparent objects such as window glasses and highly specular surfaces. Figures 4.10(c) and Figure 4.11(c) present the result of range and intensity data fusion. The visual inspection of the experimental results shows that the range and intensity images are highly consistent. In the figures of range and intensity fusion, the areas with no range or intensity data (or neither) are shown in white.

4.4 Laser rangescanner system characterisation

Data quality is a key factor in any measurement system that aims to effectively and accurately acquire data. The validity of data processing is, in turn, directly related to the validity of the acquired data. In general, the quality of data obtained by a laser rangescanner system is affected by the accuracy of the system components which are laser rangefinder and the scanning head. The next subsections discuss the sources of error in both laser rangefinder and the pan-tilt unit as the main parts of the proposed 3D scanning system. The operational characterisation of the experimental laser rangescanner system is also obtained and presented.

4.4.1 Rangefinder errors

Different types of interference may corrupt range measurements in a time-of-flight laser rangefinder in both temporal and spatial fields, including warm up time, divergence of the detection beam, target surface and colour properties and mixed pixels effect. The next subsections first present the temporal effects that are mainly related to the hardware of the laser time-of-flight measurement. These include laser temperature effect and ambient temperature effect on the measurement. The errors pertaining to mixed pixels and object reflective properties appear in the spatial domain and they are presented in the following subsections.
4.4.1.1 Temperature effect on the laser hardware system

Fluctuations of range measurements over time are common to all laser rangefinder devices, mostly due to the system hardware. To investigate the time it takes for the laser measurement to stabilise, an experiment is conducted. In this test, \(27 \times 10^4\) measurements are taken in \(10^{-3}\) second intervals over an hour period at a distance of nine metres.

![Figure 4.12: Measurement fluctuations over time due to temperature effect on laser rangefinder electronics.](image)

Figure 4.12 shows the average of the measurements within an hour period. It shows that a warm up time of about twelve minutes is required for the laser measurement to become steady. Due to the fact that this measurement error is unavoidable, all experiments performed from here onwards, will disregard measurements taken in the first fifteen minutes. The figure also shows that the repeatability of the measurement is poor in the first few minutes of operating the laser.

4.4.1.2 Surrounding temperature effect

To avoid uncertainty in range measurements, a temperature range in which the laser can work properly is given by the manufacturers. Due to the fact that in this project the laser rangefinder operates in an uncontrolled outdoor environment, the measurement fluctuations
caused by varying environmental temperatures cannot be ignored. To study this effect, a number of range measurements at different ambient temperatures are conducted. Figure 4.13 shows range measurements for a variety of temperatures at two different distances. In this experiment, each measurement is the average of $20 \times 10^3$ range readings at 50 Hz over fifteen minutes.

**Figure 4.13: Range measurements vs. temperature at distances of nine and nineteen meters**
The above graphs show that the range measurement of the same distance becomes smaller at higher temperatures. A twenty degree change in temperature results in around a one metre change in the average range measurements at the distance of approximately 20 metres.

### 4.4.1.3 Range accuracy at different distances

The range accuracy given by the manufacturer for the laser rangefinder used in this project is typically three decimetres, where in practice it is variable at different distances. To determine the range accuracy, a number of measurements were taken at different distances. Table 4.6 shows a statistical analysis of the measurements obtained at different distances. The mean, variance and RMS are calculated from over $10^5$ range readings for every distance.

**Table 4.6: Range accuracy at different distances and statistical values of range measurements**

<table>
<thead>
<tr>
<th>Ground Truth Value (dm)</th>
<th>Measurement Mean (dm)</th>
<th>Measurement Variance (precision) (dm)</th>
<th>Measurement Error (RMS) (dm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>71.71015</td>
<td>0.438596</td>
<td>0.722888</td>
</tr>
<tr>
<td>91</td>
<td>90.65403</td>
<td>0.346785</td>
<td>0.682969</td>
</tr>
<tr>
<td>198</td>
<td>196.6556</td>
<td>0.348135</td>
<td>1.688177</td>
</tr>
<tr>
<td>515</td>
<td>514.666</td>
<td>0.339101</td>
<td>0.671274</td>
</tr>
<tr>
<td>1193</td>
<td>1192.54</td>
<td>0.335321</td>
<td>0.358136</td>
</tr>
</tbody>
</table>

The ground truth value is the distance measured from the front window of the laser rangefinder to the target surface. The *measurement mean* is calculated by the following equation:

$$M = \frac{\sum_{i=1}^{n} R_i}{n} \quad (4.19)$$
where $R_i$ is the experimental range measurement and $n$ is the number of measurements. These results show that the mean value of the measurement is always less than the ground truth value. It is mainly due to a measurement bias inside the laser rangefinder.

The measurement variance is a measure of the dispersion shown by a set of observations and is calculated by the following formula:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (M - R_i)^2}{n} \quad (4.20)$$

where $M$ is measurement mean and can be calculated from Equation 4.19. The Root Mean Square (RMS) error of the measurements is defined by following equation:

$$e_{rms} = \sqrt{\frac{\sum_{i=1}^{n} (T - R_i)^2}{n}} \quad (4.21)$$

where $T$ is the ground truth value. Figure 4.14 shows that the precision of the measurement is better than 0.35 decimetres except for the closest range reading. Also the error of the measurements significantly decreases as the distance is increased. These facts prove that this laser rangefinder is optimised to work at large distances of up to 300 metres.
Figure 4.14: Comparison of the variance and RMS error for five different distances.

### 4.4.1.4 Mixed Pixel effects

As mentioned in Chapter 2 (Section 2.2.3), one of the limitations of the laser TOF measurement is the mixed pixel effect problem. This problem arises when a laser beam points to the edge of an object and so the data measured for a single measurement comes from two different surfaces (see Figure 2.10). This problem is mainly due to the size of the laser footprint and the range measurement is a weighted average of the distances to the two surfaces where laser footprint lies. To discover the effect of mixed pixels on range measurement, an analysis of the range data on a number of edges in the range data of an outdoor scene is conducted. To this purpose, several cross sections of the range image were selected and depth discontinuities were observed. Figure 4.15 shows one of the observations.
Figure 4.15: Range image of the front side of the Shrine of Remembrance building, Melbourne, Australia: (a) edges appear as sawtooth due to the mixed pixel effect, (b) mixed pixel effect in the part of the three cross sections of the range image, between columns 250 and 500 (row 240 appears in red, row 250 appears in cyan and row 260 appears in green)
4.4.1.5 Target reflectance properties

Due to the fact that the laser measurements are based on the energy absorption of the target, the reflectance properties of the objects in the scene affect the range measurements. To study this effect, several experiments have been conducted. Three surfaces of white, black and shiny metal were selected. The three targets were each placed at three different distances. To avoid the effect of angle of incident, the laser rangefinder is located perpendicular to each target.

Figure 4.16 shows the effect of the reflectance property of the target on the range measurement. The target of low reflectance (e.g. black) appears further than the target of high reflectance (e.g. matt white), although both are measured from the same distance. A target of high reflectance saturates the photo detector and causes the target to appear closer, while a target of low reflectance absorbs more energy and as a result the target appears further away. A target of very high reflectance diffuses the signal so that the energy of the received signal is greatly reduced. This effect causes the target to appear much farther away than the ground truth.

An observation from the results of the final experiment of this project shows that the laser rangefinder fails to accurately measure the distance to highly reflective materials such as windows, shiny metals and mirrors (e.g. Figure 4.10(b) and 4.11(b)).
4.4.2 PTU Errors

The accuracy of the measurement made by the laser range scanner system is also limited by the pan-tilt unit. The manufacturer claims that the PTU (at its best resolution) can be positioned every 0.0129° in vertical or horizontal directions. In other words, the PTU can move 13,954 positions in 180°, however when a pan of 180° is performed, the position is off by tenths of a degree.

To estimate the error of both the pan and tilt movement of the PTU, an experiment was performed using the original set-up of the rangescanner system with an extra visible infrared laser, mounted between the transmitter and receiver of the laser rangefinder. Figure 4.17 shows the PTU error measurement set-up. In this set-up the angular resolution of the PTU is set to 0.0129° (the best resolution of the PTU) and a wooden rectangular target board of 1.5m × 2m is placed at the distance of about fifteen meters from the laser rangescanner system. The steps of finding angular error of the PTU are as follows:

1) aim the visible laser horizontally at the target board, mark its position on the target board and record the distance of the target measured by the LRF;
2) move the visible laser clockwise by panning the PTU 100 steps from its home position and mark its position on the target board;
3) move the visible laser anticlockwise by panning the PTU 100 steps from its home position and mark its position on the target board;
4) repeat the last two procedures when panning the PTU 500 steps and
5) repeat the stages 2, 3 and 4 when tilting the PTU.

Table 4.7 shows the result of the above experiment to find the PTU angular errors. Based on the measurements presented in the table, vertical and horizontal angular resolutions (VAR and HAR) of the system are not equal and both are greater than the expected value given by the manufacturer. Based on the results, the angular resolution of the system is 0.014° for panning while this value is 0.015° for tilting.
Chapter 4 — Laser Rangescanner System

![Diagram of laser rangescanner system](image)

Figure 4.17: System set-up to determine errors in PTU measurement: (left) horizontal angular resolution (HAR) and (right) vertical angular resolution (VAR). In this experiment, the distance from target (D) measures 1520 centimetres by LRF.

<table>
<thead>
<tr>
<th>PTU Angular Position</th>
<th>Estimated Resolution at the Position</th>
<th>Estimated Resolution</th>
<th>Angular Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>±100 steps panning and tilting</td>
<td>100 × 0.0129° = 1.29°</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>+100 steps <strong>panning</strong></td>
<td>(\tan^{-1}(37.1/1520) = 1.4°)</td>
<td>0.0140°</td>
<td>0.0011°</td>
</tr>
<tr>
<td>+100 steps <strong>tilting</strong></td>
<td>(\tan^{-1}(39.8/1520) = 1.5°)</td>
<td>0.0150°</td>
<td>0.0021°</td>
</tr>
<tr>
<td>-100 steps <strong>panning</strong></td>
<td>(\tan^{-1}(37.8/1520) = 1.42°)</td>
<td>0.0142°</td>
<td>0.0013°</td>
</tr>
<tr>
<td>-100 steps <strong>tilting</strong></td>
<td>(\tan^{-1}(38.8/1520) = 1.46°)</td>
<td>0.0146°</td>
<td>0.0017°</td>
</tr>
<tr>
<td>±300 steps panning and tilting</td>
<td>300 × 0.0129° = 3.87°</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>+300 steps <strong>panning</strong></td>
<td>(\tan^{-1}(111.5/1520) = 4.19°)</td>
<td>0.0140°</td>
<td>0.00106°</td>
</tr>
<tr>
<td>+300 steps <strong>tilting</strong></td>
<td>(\tan^{-1}(118.6/1520) = 4.46°)</td>
<td>0.0149°</td>
<td>0.00196°</td>
</tr>
<tr>
<td>-300 steps <strong>panning</strong></td>
<td>(\tan^{-1}(111/1520) = 4.17°)</td>
<td>0.0139°</td>
<td>0.00100°</td>
</tr>
<tr>
<td>-300 steps <strong>tilting</strong></td>
<td>(\tan^{-1}(117.7/1520) = 4.42°)</td>
<td>0.0147°</td>
<td>0.00183°</td>
</tr>
</tbody>
</table>
4.5 Verification of the functionality of the 3D scanner system

A design process is a systematic procedure to transform a set of specifications into an implementation. The specifications state the functionality that the design is expected to perform. An implementation explains the details of how the design is executed. Design verification confirms that the implementation meets its specifications. In this project, to ensure that the design of the rangescanner system meets the expected accuracy and usability, the output of the 3D scanner system (range data) was verified using two methods of verification: application checking and equivalence checking. Both validation techniques are adapted from the electronic hardware design modification methods, presented by W. Lam [161], and modified for the rangescanner system.

The application checking verification method is used to examine the performance of the laser rangescanner system in terms of the usability of its output. The equivalence checking verification method is validation by redundancy. This method is employed to examine the accuracy of the resulting range data. Figure 4.18 shows how the design of the laser rangescanner system is validated through the above mentioned methods.

![Figure 4.18: Laser rangescanner system verification](image-url)
4.5.1 Application checking verification of 3D measurement system

The application checking verification is used to examine the functionality of the laser rangescanner system in terms of the usability of its output. To this purpose, as shown in Figure 4.19, a parametric segmentation technique has been adopted to segment the point cloud data, obtained by the prototype system, into planar surfaces. Among the range segmentation methods, Modified Selective Statistical Estimator (MSSE) \cite{111} has been chosen as it is robust to the effect of outliers and pseudo-outliers \cite{162} and is quite straightforward and easy to implement. If the output of the rangescanner system is correct and accurate, it is expected that the MSSE could extracts those segments that contain at least $K$ percent of the whole data. As mentioned in Chapter 3, $K$ is a tuning input for the MSSE segmentation and is the size of the smallest structure to be segmented.

In MSSE algorithm, a group of data, which initially was supposed to be laid on a planar surface, is found (even if it is not on a planar surface). To this purpose, a window of size $15 \times 15$ pixels is considered randomly as a sampler and an over-determined linear equation system is generated, based on the obtained samples. If more than half of the samples are inliers, the data group is accepted as a good sample. A planar model is then fitted to all the accepted samples and the residual for every point in the data set is estimated.

\footnote{The implementation of MSSE was described in Chapter 3.}
Also, the scale of noise is computed using the MSSE. The above steps are repeated a number of times to recognise the samples with the least $K$-th order residual. The choice of $K$ depends on the application [111]. Here, it is set to be 2 per cent (for large-scale application). Next, the points with the residuals of greater than a threshold $T$ multiple of 2.5 times of the scale of noise are rejected as outliers. A new partition is then created containing all the inliers to this fit. The inliers are excluded from the whole data and the above steps repetitively executed on the remaining data until the number of data becomes less than the threshold $K$.

It is necessary to note that a hole-filling algorithm (here, a median filter of 10 by 10 pixels) is applied to the inliers to remove the holes produced by moving objects and different sources of noise. As the fitting has already been performed, this step only improves the look of the result and has no effect on the segmented parameters of the surface.

Figure 4.20 shows the range segmentation result on the range data captured from the front view of the Royal Melbourne Exhibition building and Figure 4.21 presents the range segmentation result on the range data acquired from the front side of the Shrine of Remembrance building. As shown in Figure 4.20 (c) and 4.21 (c), segmentation results verify that the range data captured by the experimental laser rangescanner is appropriate enough to satisfy a challenging computer vision application such as range segmentation as range segmentation algorithm has successfully detected the planar and coplanar areas.
Figure 4.20: The Royal Melbourne Exhibition Building: a) colour-coded range image (715×400) pixels, b) intensity image (715×400) pixels, and c) range segmentation result. Different colour shades in grey-scale shows each segment.
Figure 4.21: The Shrine of Remembrance: a) colour-coded range image (715×400) pixels; b) intensity image (715×400) pixels; and c) range segmentation result. Different colour shades in grey-scale show each segment.
4.5.2 Equivalence checking verification of 3D measurement system

The equivalence checking verification method is used to examine the accuracy of range data obtained by the experimental 3D scanner system. As shown in Figure 4.22, in this method the output of the prototype system is compared with the same range data set obtained by an accurate commercial laser rangescanner.

![Figure 4.22: Equivalence checking verification of laser rangescanner system](image)

The output of the experimental rangescanner system and the reference commercial rangescanner system are presented in the form of colour-coded range images, shown in Figure 4.23 and 4.24. The visual comparison of the range images shows that although there is a very close agreement between two sets of data, the experimental rangescanner system is not able to correctly measure the distance of windows and darker areas (e.g. doorway of the Shrine of Remembrance in Figure 4.23 (b)). This comparison also shows that the edges of the reference range image appear smoother than the edges in the experimental range image. This difference might be because the laser beam of the commercial rangescanner has less divergence than the laser beam of the experimental rangescanner.
Figure 4.23: Range images of the front side of the Shrine of Remembrance, Melbourne, Australia: (a) obtained by the Riegl laser rangescanner (b) obtained by the experimental laser rangescanner

Figure 4.24: Range images of the front side of the Royal Melbourne Exhibition Building, Melbourne, Australia: (a) obtained by the Riegl laser rangescanner (b) obtained by the experimental laser rangescanner

In order to precisely evaluate the accuracy of the measurement in the experimental rangescanner, different geometric features of the spatial data are extracted and measured in both range images. Figure 4.25 and 4.26 show where the measurements are taken and Table
4.8 and 4.9 summarise the measurement results for the range images of the Shrine of Remembrance and the Royal Melbourne Exhibition Building, respectively.

Figure 4.25: Corresponding features of the range images of the Shrine of Remembrance, Melbourne, Australia obtained by: (a) Riegl laser rangescanner (b) Experimental laser rangescanner
Figure 4.26: Corresponding features of the range images of the Royal Melbourne Exhibition Building, Australia obtained by: (a) Riegl laser rangescanner (b) Experimental laser rangescanner.
Chapter 4 — Laser Rangescanner System

Table 4.8: Comparison of estimated corresponding line segments of range image obtained by Riegl and experimental laser rangescanner (image of the Shrine of Remembrance)

<table>
<thead>
<tr>
<th>Estimated Distances Obtained from Riegl (m)</th>
<th>Estimated Distances Obtained from Experimental Laser Rangescanner (m)</th>
<th>Measurement Error (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 = 19.6 (horizontal)</td>
<td>D'1 = 19.3</td>
<td>0.3</td>
</tr>
<tr>
<td>D2 = 3.0 (vertical)</td>
<td>D'2 = 3.4</td>
<td>0.4</td>
</tr>
<tr>
<td>D3 = 11.6 (slope)</td>
<td>D'3 = 10.7</td>
<td>0.9</td>
</tr>
<tr>
<td>D4 = 20.2 (horizontal)</td>
<td>D'4 = 20.7</td>
<td>0.5</td>
</tr>
<tr>
<td>D5 = 27.3 (horizontal)</td>
<td>D'5 = 27.7</td>
<td>0.4</td>
</tr>
<tr>
<td>D6 = 2.9 (vertical)</td>
<td>D'6 = 3.2</td>
<td>0.3</td>
</tr>
<tr>
<td>D7 = 1.8 (vertical)</td>
<td>D'7 = 2.0</td>
<td>0.2</td>
</tr>
<tr>
<td>D8 = 18.1 (horizontal)</td>
<td>D'8 = 18.8</td>
<td>0.7</td>
</tr>
<tr>
<td>D9 = 1.7 (vertical)</td>
<td>D'9 = 1.9</td>
<td>0.2</td>
</tr>
<tr>
<td>D10 = 1.7 (vertical)</td>
<td>D'10 = 1.9</td>
<td>0.2</td>
</tr>
<tr>
<td>S1 = 0.96 (diameter)</td>
<td>S'1 = 0.94</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 4.9: Comparison of estimated corresponding line segments of range image obtained by Riegl and experimental laser rangescanner (image of the Royal Melbourne Exhibition Building)

<table>
<thead>
<tr>
<th>Estimated Distances Obtained from Riegl (m)</th>
<th>Estimated Distances Obtained from Experimental Laser Rangescanner (m)</th>
<th>Measurement Error (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 = 12.2</td>
<td>D'1 = 12.3</td>
<td>0.01</td>
</tr>
<tr>
<td>D2 = 29.1</td>
<td>D'2 = 29.3</td>
<td>0.2</td>
</tr>
<tr>
<td>D3 = 10.5</td>
<td>D'3 = 10.9</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Comparison of the measurements shows that the peak to peak horizontal variation in the samples taken from the range images of the Shrine of Remembrance is 2.71 – 42.99
centimetres and this variation is 1.35-38.46 centimetres for the range images of the Royal Melbourne Exhibition Building. The vertical variations are 32.33 and 28.13 centimetres, respectively. The verification shows that range images obtained by the experimental rangescanner is equivalent to the range images obtained by the commercial rangescanner with regard to the accuracy of their measurement components. The observed differences are mainly due to the accuracy and characteristics of the laser rangefinder and the mechanical parts of the rangescanner systems, which are different in each system.

4.6 Chapter Review

This chapter has covered the main aspects of the design and implementation of a prototype laser rangescanner system. It is comprised of four main sections:

In the first section, the structure design of the hardware and software of the laser rangescanner system has been described. The design criteria for the experimental device were to: a) facilitate scientific research at an affordable cost, b) provide adequate accuracy and resolution for large-scale civil applications, and c) provide simple operation, portability and flexibility. The rangescanner system has two vision sensors (3D and 2D) and has been designed to capture the data of large buildings in outdoor environments. In addition, to obtain a rich set of data containing geometry and colour information for each point of the scene, a 3D-2D data fusion has been applied to the acquired range and intensity data. To this purpose, a camera calibration routine was used to precisely extract the intrinsic and extrinsic parameters of the intensity sensor. The camera calibration algorithm was selected to satisfy the perspective projection camera model. Moreover, the algorithm takes into account the radial and tangential distortion of the lens. After the camera was calibrated and all equipment was installed, an automatic calibration procedure was applied to the system to establish a set of parameters for proper system operation. To measure the quality of data fusion, a validation technique was performed, based on the system calibration technique.

In the second section, the range data obtained by the prototype system have been presented in the form of colour-coded range images. The results of the range and intensity data fusion have also been presented. The results of the data fusion show that in spite of the
large number of outliers, produced by external factors, the relationship between range and intensity data of the same scene are highly consistent.

In the third section, the sources of the range measurement error (systematic errors) in the prototype system have been accurately specified. It has been shown that a number of environmental parameters such as the temperature, light and property of the objects could affect the accuracy of range measurement. Also, it has been investigated that some of the specifications of the mechanical and electronic components of the system, given by its manufacturer, are only typical values and the rangescanner system needs to be calibrated based on the practical characteristics of its components.

In the last section of this chapter, the performance of the prototype rangescanner system has been verified by two verification methods: equivalence checking and application checking. In the equivalence checking method, the output of the experimental rangescanner system (range data) is compared with the output of a reference rangescanner system. The reference system is considered to be an accurate commercial laser rangescanner system. This verification has shown that the most prominent limitation of the experimental system (against the reference system) is its limitation to measure the distance from the windows and dark areas. This limitation is due to the fact that the laser beam is very low power and is absorbed by the dark environment. Also the edges of the objects are not as smooth as expected. In application checking method, the output of the prototype rangescanner system is used as an input for a segmentation algorithm. Range segmentation is a potential application for range data in the field of computer vision. The result of range segmentation verified that the 3D data produced by prototype system is reliable for computer vision applications.

There are a few concerns that are worth mentioning here that have a direct affect on the way in which the experiments were carried out. During the practical experiments on real world environments, it has been realised that range segmentation of large-scale/outdoor objects is a more challenging problem than the range segmentation of laboratory-sized or simulated objects. The first issue is the fact that there is always a trade-off between the resolution of the range image and the time consumed for data acquisition and data processing (e.g. range segmentation). A low resolution range image contains less information about the
objects but its segmentation is straightforward. Where fine details are not important or the scene does not contain small details, the availability of such a dataset could be satisfactory. On the other hand, to capture all existing details in a scene of interest (e.g. decorative details embedded in the façade of historical buildings), a set of dense range data is required. In this case, because of the large size of the data and the existence of the large disparity in the size of the objects, the processing of such data is very challenging and most of the available algorithms have failed to accurately process such data. For instance, Figure 4.21 shows that the segmentation algorithm could not detect and extract the stairs and decorative details embedded in the pediment and frieze of the Shrine of Remembrance building. The reason is that the algorithm ignores structures that contain less than 2 per cent of data. Examples of such details are shown in Figure 4.27 below.

Moreover, the disparity of depth in outdoor scenes is often substantial. This disparity means that the difference in range can easily be much larger than the object features in an outdoor scene. Such a disparity in depth has serious implications in terms of the scale of the noise, as in many practical cases the level of the noise is proportional to the magnitude of data and most heuristic methods to estimate the scale of noise would be likely to fail in such circumstances.

Another major problem in capturing the 3D data of large-scale/outdoor objects is the fact that the range data is always marred by the influence of environment conditions. During
data acquisition in the places of interest, it is very difficult to control certain factors
including the flow of tourists, emergency vehicles and wind.

In addition, there were some difficulties related to finding appropriate scenes for data
acquisition. As mentioned previously, the application of the author's research project is
mainly large-scale structured building exteriors. Such buildings are usually historical or
governmental and capturing any data from their internal and external sides requires the
necessary permissions. Also, seclusion from the public to satisfy Occupational Health and
Safety (OHS) guidelines is almost impossible. Furthermore, the duration/number of
repetition of each experiment is limited to the power source (i.e. car batteries for PTU and
LRF, back-up power of laptop). As a result, each experimental set-up should be carefully
planned.

In the next chapter, the main problems associated with data acquisition and processing
of large building exteriors that are mentioned above will be mathematically formulated and
analysed.
Chapter 5

Range Data of Large Building Exteriors

5.1 Introduction

In the previous chapter, a low-cost laser rangescanner system was devised and a robust parametric range segmentation algorithm [111] was applied to the output of the acquisition system. The experimental results on the data acquisition and the range segmentation of large building exteriors led to identifying a number of unavoidable and challenging problems in this domain.

This chapter aims to introduce and analyse the characteristics of the geometric data of large buildings (complex multi-structure data) that make the processing (e.g. segmenting) of such data a challenging problem. To this purpose, several sets of simulated data are produced to examine the problems.

5.2 Characteristics of range data of building exteriors

The datasets of outdoor man-made scenes obtained by modern 3D measurement devices,
compared to conventional range data produced in laboratories, pose serious challenges to the existing range data processing techniques [13, 163, 164]. The characteristics of such data sets are described in the following sections.

### 5.2.1 Existence of moving objects

The production of range data of outdoor scenes is unavoidably affected by moving objects such as birds, pedestrians and vehicles. As shown in Figure 5.1, the existence of those unwanted objects commonly results in a large number of outliers in the range data. The random effect of moving objects leads to issues broadly referred to as **occlusion problem**. Therefore, any processing techniques that are applicable to this data have to be robust to the effect of outliers.

![Intensity and range images of the Royal Melbourne Exhibition Building](image)

*Figure 5.1: Intensity and range images of the Royal Melbourne Exhibition Building. Cars, passengers and vegetation are obstacles that are unavoidable at the time of data collection. Moving objects appear as straight lines in the range image.*
5.2.2 Disparity in size

Man-made objects in outdoor scenes appear with different sizes and geometric complexities. This difference is in contrast with the range data of indoor objects such as those depicted in the ABW range image database [165]. For example, the building of the Shrine of Remembrance in Figure 5.2 contains large surfaces such as walls and ceilings; medium size objects such as staircases and statues; and small size objects such as detailed decorative carving in the façade of the building. In this example, when walls of a building may contain as many as 30 per cent of all data points, the surfaces associated with roof decoration may only contain one per cent of all data. Moreover, an object of interest that is located close to the laser rangescanner includes more data points compared to the similar object located further away. Since distant and small structures have a small number of data samples, a range segmentation algorithm that relies on a minimum size for structures - as most techniques do explicitly (e.g. LKOS and MSSE) or implicitly (most random sampling-based methods) may not be able to extract all possible structures of interest. The significant difference in the size of the objects of interest highly complicates the extraction of small structures as it often leads to over-segmentation of larger sections.

Figure 5.2: The front view of the Shrine of Remembrance building. Large disparities in the size of features are highlighted in different colours. This building is obviously composed of a collection of large, medium and very small structures
To show the effect of disparity in size on the segmentation process, an experiment using synthetically generated data has been designed and conducted. Figure 5.3(a) shows a sample of various sets of synthetic data generated in the experiment. The scene represented by 3D synthetic data contains two parallel planar structures and outliers as following:

1) 
\[ P_1 : 0 < X < D_1; 0 < Y < D_1; Z = 20 \pm \varepsilon; \Delta X = \Delta Y = 0.1; \varepsilon \in N(0,0.1); n_1 = 1600 \]

2) 
\[ P_2 : D_2 < X < D_2 + (D_1 S); D_2 < Y < D_2 + (D_1 S); Z = 40 + \varepsilon; \Delta X = \Delta Y = 0.1; \varepsilon \in N(0,0.1); S = 0.1: 0.1: 1; D_2 = (1-S)(D_1 / 2); n_2 = 16: 1600 \]

3) 
\[ \text{Outliers : } D_1 / 2 < X_o < -D_1 / 2 + 2D_1; -D_1 / 2 < Y_o < -D_1 / 2 + 2D_1; 0 < Z_o < 60; n_o = 80 \]

The large plane \((P_1)\) is a 4×4 metres planar surface, located in the depth of 20 metres and contains 1,600 data points while the size of the small plane \((P_2)\) varies from 0.4×0.4 metres to 4×4 metres (defined by \(S\)) and number of data points of the small plane \((P_2)\) varies from 16 to 1,600 (defined by \(n_2\)) respectively in 10 steps. Data of both structures are generated using square shape regular grids (the same grid size for both, defined by \(\Delta X\) and \(\Delta Y\)) corrupted by additive Gaussian noise \(N(0,0.1)\). Around 80 uniformly distributed wrong measurements (representing 2.44%-4.72% of the whole population) which emulate the effect of gross outliers have also been added to the mix. A parametric range segmentation algorithm based on robust estimation (MSSE [111]) is then applied to segment this data.

The above experiment was repeated 100 times for every value of the parameter \(K\) in MSSE (the proportion of the size of the smallest data group that would be considered a structure by the estimator – here, varying from 2 per cent to 12 per cent of the whole population) and the success rates of the robust estimator in separating both planes were recorded as shown in Figure 5.3(b). This figure indicates that successful segmentation of all possible structures greatly depends on the size of the embedded structure. Structures
containing less than 20 per cent of all the data population are less likely to be segmented as a separate structure.

Figure 5.3: a) A sample of the synthetic data used to demonstrate the effect of disparity in size for large-scale range data segmentation. b) A plot representing the percentage of success in segmenting
both small and large structures with a robust estimator versus the ratio of the size of small structure to the size of whole population for different values of $K$.

5.2.3 Existence of very fine details

Modern 3D laser rangescanners are able to capture high resolution and dense geometric data points of building exteriors in a short period of time. As a result, outdoor range datasets are rich in detail. Moreover, most of the buildings of interest (e.g. historical buildings) contain architectural details such as columns, statues and staircases. Figure 5.4 shows examples of fine details in the Shrine of Remembrance building.

![Figure 5.4: Fine details in the façade of the Shrine of Remembrance building](image)

Generation of a simplified model of such buildings is now feasible using available techniques — for instance, see [166]. However, extraction of fine details of the ornamental buildings has remained a challenging task that can only be performed either manually [9] or by a huge amount of computation. The computation cost associated with either of those methods increases rapidly with modest increases in the desired level of interest.
The main reason for the expensive computations involved in segmentation of fine details is the very large number of random samples that are required by robust segmentation techniques based on RANSAC. The RANSAC-based estimators use random sampling as a search method to optimise their objective (cost) function. As shown by Fischler and Bolles [102], if in a random sampling search scheme, the number of \( p \)-tuples is \( m \), the probability of having at least one good sample (a sample belonging to the segment of interest) in \( p \)-dimensional parameter space is:

\[
P = 1 - (1 - (1 - \varepsilon)^p)^m
\]

(5.1)

where \( \varepsilon \) is the fraction of data points not belonging to the segment one tries to find by random sampling. Therefore, the minimum number of random samples required for having at least one good sample (which by itself is far from satisfying the sufficiency condition [116]) is calculated by:

\[
m = \frac{\log(1 - P)}{\log[1 - (1 - \varepsilon)^p]}.
\]

(5.2)

The above formula shows that the required number of random samples rapidly increases with small inlier ratios (\( \varepsilon \) values very close to 1) encountered when fine details are to be segmented. For instance, if the size of the smallest structure of interest in a multi-structure three dimensional scenario (\( p=3 \)) is 1 per cent (\( \varepsilon=0.99 \)) of all the data population, then more than two million random samples are required to find the structure of interest 90 per cent of times. Figure 5.5 shows the rapid change in the number of required random samples for a range of inlier ratios for different probabilities of success.

It is important to note that in practice, the outlier ratio (\( \varepsilon \)) is not known \textit{a priori} and has to be assumed to be fairly large to guarantee that small structures are not overlooked. Therefore, for the successful segmentation of small patches, \( \varepsilon \) values very close to one must be chosen which would result in prohibitive computation as exemplified above.
5.2.4 High uncertainty due to the construction errors

As mentioned in the previous chapter, in the experiments with the range data of building exteriors, it has been found that segmentation of building exteriors is highly affected by the construction accuracy of the modelled building. Construction errors in large buildings are generally unavoidable and their scales are significant when compared to the accuracy of the 3D measurement systems (a few centimetres and a few millimetres, respectively). In particular, the effect of construction error becomes a significant issue when different parts of one structure are located apart. In such circumstances, and depending on the level of construction error, model-based range segmentation algorithms may no longer be able to detect coplanar surfaces as single structures.

To investigate the effect of construction error on the segmentation process, a simulation experiment has been carried out. The detail of this experiment is shown in Figure 5.6(a). The scene represents 3D synthetic data containing coplanar surfaces as follows:

1)
Chapter 5 — Range Data of Large Building Exteriors

\[ P_1 : 0 < X < W; 0 < Y < L; Z = 10 \pm \varepsilon; \Delta X = \Delta Y = 0.25; \varepsilon \in N(0,0.1); W = 5; L = 10; n_1 = 500 \]

2)

\[ P_2 : W + mW < X < 2W + mW; 0 < Y < L; Z = 10 \pm \mu \sigma \pm \varepsilon; \Delta X = \Delta Y = 0.25; \varepsilon \in N(0,0.1); m = 0 : 1 : 12; \mu = 0.5 : 0.5 : 3; \sigma = 0.1; W = 5; L = 10; n_2 = 500 \]

3)

Outliers : \( 0 < X_0 < W(2 + m); 0 < Y_0 < L; Z_0 = 10 \pm \varepsilon; \varepsilon \in U(0,1); n_{out} = 80 \) .

Each plane \( (P_1 \text{ and } P_2) \) contains a total of 500 data points and is similar in size (5 meters wide \( W \) by 10 metres in length \( L \)). The distance between planes varies from 0 to 60 metres \( mW \). Data of both planar surfaces are generated using rectangular (and regular) grids corrupted by additive Gaussian noise \( N(0, 0.1) \). The construction errors are then modelled by moving one surface parallel to the other in depth \( \mu \) times the scale of noise. A number of randomly distributed gross outliers (around 30 per cent of population, representing wrong measurements or miscellaneous building parts) have also been added to the set and a robust estimator (MSSE) is applied to segment this data.

The above experiment was repeated 100 times for different values of \( \mu \) (the construction error in depth) ranging from 0.5 to 3 times of measurement error (here, 10 mm). The number of times that the robust estimator has successfully labeled both patches as a single plane was recorded and is shown in Figure 5.6(b). The plot shows that successful segmentation of coplanar surfaces directly depends on the distance separating those coplanar structures and the amount of the construction error. Coplanar structures separated by more than two times their dimension are not likely to be segmented as coplanar and the situation worsens as construction error increases.
Figure 5.6: a) Sample of the synthetic data used for segmentation analysis of distant co-planar surfaces. b) Likelihood of detecting co-planar surfaces as one segment vs. distance of structures for cases where $K=0.1$.

### 5.3 Chapter review

In this chapter, the key issues relating to segmentation of 3D surfaces using range data obtained from large building exteriors were identified and analysed. The observations made during the data acquisition of large outdoor scene and experimental results on simulated data
imply that these issues are mainly related to disparities in the size of the objects, construction errors and unwieldiness of data of large scenes.

In order to overcome those issues, in the following chapter, a hierarchical framework will be introduced for high breakdown robust estimation of surface parameters in large building range segmentation applications. The proposed algorithm accomplishes segmentation at three levels — coarse, medium and fine. Hence the algorithm recovers small structures without the interference of the larger ones, while avoiding the over-segmentation issue and reducing the total computation time and memory requirements.
Chapter 6

Hierarchical Robust Range Segmentation

6.1 Introduction

In the previous chapter the problems associated with the range segmentation of large building exteriors were examined. The approach that has been chosen to address these issues is to employ a high breakdown robust estimator. This technique is selected for two main reasons (see Chapter 4). Firstly, it uses high-level information such as the geometry of the objects which is appealing for segmentation of man-made objects that are mainly built of low-order surfaces. Secondly, the spatially separated parts of an object can be detected simultaneously, regardless of their geometrical locations. However, using such a method to extract all possible details from range data has a drawback. The minimum presumed size of a segment for such a segmentor has to be set to very small values, requiring a huge number of samples (randomly taken) to solve the optimisation problem. Therefore, such a scheme would be practically infeasible in terms of both memory and computation time. Furthermore, the existence of construction errors and fine details lead to over-segmentation when a robust estimator is used.
In this chapter, a novel segmentation scheme to overcome these issues, whilst maintaining the robustness of the process, is presented. The scheme uses high breakdown robust estimators in a multi-step global-to-local approach to extract both coarse and fine details of large building exteriors.

This chapter consists of four sections. Section 6.2 explains the usage of the robust estimators in the context of range segmentation. Section 6.3 explains the steps of the novel Hierarchical Robust Segmentation (HRS) scheme that is designed to extract fine structures embedded in the data of building exteriors. The experimental results on synthetic and real datasets and the evaluation of the proposed segmentation scheme are presented in Section 6.4.

6.2 From robust estimation to robust segmentation

Robust estimation techniques have been broadly used in many computer vision tasks as they have been successfully demonstrated to tolerate outliers of various types (such as impulsive noise generated by sensors, neighbouring structures — called pseudo outliers [167] — and environmental noise (e.g. [111, 116]). As mentioned in Chapter 3, these methods are either adopted from the statistics community (e.g. Least Median of Square (LMedS) [100]), or innate to the computer vision field (e.g. Hough Transform (HT) [101], RANdom SAmple Consensus (RANSAC) [102]). Most of these techniques, especially those introduced by the computer vision community, have a breakdown point of more than 50 per cent. Examples are, Minimize the Probability of Randomness (MINPRAN) [168], Minimum Unbiased Scale Estimator (MUSE) [108], Adaptive Least $K$-th Order Squares (ALKS) [109], Modified Selective Statistical Estimator (MSSE) [111], Maximum Density Power Estimator (MDPE) [115], projection-based M-estimator (pbM) [117] and most recently High Breakdown M-estimator (HBM) [119].

In general, robust segmentation algorithms are a class of techniques that are designed to extract geometric primitives from raw data with multiple structures by sequentially using a robust estimation. Segmentation of the structures is usually performed in a sequential manner. In each iteration, first, the inliers (the data samples belonging to one structure) are determined by fitting a surface model to the range data and simultaneously estimating the
model parameters such as surface normal and curvature. The resulting inliers are then masked out not to be processed in the next iterations. The algorithms also eliminate the outliers produced by false measurements due to range sensor errors or malfunctioning, or environment changes. The extraction of geometric primitives, using robust estimators, often entails an optimisation problem. To minimise the cost function of the RANSAC based robust estimators, a random search is employed involving a minimal number of subsets of randomly selected points. Each minimal subset (also called a random sample or sample for short) is a $p$-tuple of measurement data points where $p$ is the dimension of parameter space. For instance for line models $p = 2$, for planar surface models $p = 3$, and for quadratic surface models $p = 9$. The number of samples has to be adequate to ensure, with a probability close to one, that at least one of the samples is a good sample. As described in Chapter 5, according to Equations 5.1 and 5.2, for higher dimension models (large $p$) or in the existence of small structures to be extracted, the number of required samples significantly increases and therefore, optimisation becomes very time-consuming.

### 6.3 Hierarchical Robust Segmentation scheme

As mentioned previously, robust estimation is the tool that has been chosen to address the complexity and uncertainty issues associated with the segmentation of range images. To deal with the problems explained in Chapter 5, a single global approach is unlikely to be satisfactory. The Hierarchical Robust Segmentation (HRS) technique presented here is designed to overcome the issues associated with range segmentation of building exteriors. In order to process range data, it has been assumed that most of the structural and decorative parts of large building exteriors are either planar or can be approximated by small planar patches [169, 170]. This assumption leads to the usage of a parametric approach. However, for applications where nonlinear forms are of importance, the proposed hierarchical framework can be extended to include model selection strategies similar to those introduced in [121]. In this approach, a robust range segmentation method is sequentially applied to the range data at different levels. As a result, the overall computation requirement is significantly reduced to a stage that is achievable by ordinary computers (see Tables 6.1 and 6.2 for a comparison) while issues with scale and size are also addressed.
The proposed algorithm starts by specifying a user-defined input to the robust estimator - a threshold $K$ which is the ratio of the size of the smallest region that can be regarded as a separate region (smallest ratio of inliers) to the whole data. Without this constraint, the segmentation task becomes a philosophical question as every three points in the dataset can in theory represent a planar patch. It is also important to note that the scale of typical measurement error of the rangefinder is usually available and can also be measured from data. Details of the implementation of each stage of the proposed range data segmentation algorithm are shown in Figure 6.2 and described as follows:

- **Range data pre-processing** — In this stage, those data points whose associated depths are not valid (due to the limitation of the laser rangefinder used for measuring the depth) are eliminated. These points are usually marked by the rangescanner software with an out-of-range number. Data of outdoor man-made objects, captured by laser technology are contaminated by noise due to the ‘mixed-pixel’ effect and moving objects. To reduce these effects, a 2D median filter of $5 \times 5$ pixels is applied to the entire valid range data.

- **Robust range segmentation** — A robust segmentation algorithm is applied to the entire data. This algorithm is initially tuned to extract a preliminary collection of coarse/large segments with appropriately setting $K$ (the value of $K$ is application dependant). The remaining data (which contain pseudo-outliers and outliers) are marked as outliers and stored for further processing. In this work, the Modified Selective Statistical Estimator (MSSE) [111] is used for robust segmentation, because it is straightforward and has the least finite sample and asymptotic bias in comparison with other popular robust estimators [162] [123]. In addition, this estimator simultaneously calculates the scale of the noise of each separated structure which is used as hierarchy criterion in the HRS algorithm. Figure 6.1 shows the inputs and outputs of the MSSE technique as a segmentor. The detail of this estimator was elaborated in Chapter 3.
It is important to note that other highly robust estimators such as pbM and HBM could also be used in this step and would be expected to produce similar results.

- **Surface fit** — A planar surface is then fitted to the data of each coarse/large segment (of the previous stage) using a least-squares fitting and calculate the scale of noise.

- **Hierarchy criterion** — If the calculated value of scale is more than the nominal scale of noise of the measurement unit, the extracted segment is considered as a coarse segment and once again, the robust segmentation algorithm is applied. Otherwise, it will be labelled as a large segment. Where applicable, this step is repeated to extract all possible details embedded in the data.

- **Sequential segmentation** — Data marked as outliers in the first segmentation stage is not discarded since the majority of such points may belong to some small structures. The robust segmentation algorithm is again applied to these data points. Smaller structures are normally detected at this stage.
Chapter 6 ─ Hierarchical Robust Range Segmentation

**Data Pre-Processing**

(e.g. deleting invalid data, median filtering, etc.)

**Coarse Segmentation:**
- **Inputs:** All range data (from previous stage). \( K \) and \( \sigma \) (here, \( K = 0.3 \) and \( \sigma = 10 \) times of accuracy of data acquisition system (\( \sigma \))
- **Method:** Robust estimation (here, MSSSE)
- **Condition:** No condition
- **Output(s):** Geometric data of each large/coarse structure and remaining data (known as outliers)

**Estimation of Scale of Noise:**
- **Input:** Range data of each coarse segment/outliers (resulted from previous stage)
- **Method:** Least-square regression
- **Condition:** No Condition
- **Output:** Estimated scale of noise (\( \hat{\sigma} \)) and parameters of each segment

**Intermediate Segmentation:**
- **Inputs:** Range data of each coarse segment/outliers, \( K \) (here, 0.25)
- **Method:** Robust estimation (here, MSSSE)
- **Condition:** \( \hat{\sigma} \) is greater than 5 times of accuracy of data acquisition system (\( \sigma \))
- **Output(s):** Geometric data of medium size structures (if there is any)

**Estimation of Scale of Noise:**
- **Input:** Range data of each intermediate segment (resulted from previous stage)
- **Method:** Least-square regression
- **Condition:** No Condition
- **Output:** Estimated scale of noise (\( \hat{\sigma} \)) and parameters of each segment

**Fine Segmentation:**
- **Inputs:** Range data of each intermediate segment, \( K \) (here, 0.15)
- **Method:** Robust estimation (here, MSSSE)
- **Condition:** \( \hat{\sigma} \) is greater than the accuracy of data acquisition system (\( \sigma \))
- **Output(s):** Small size structure(s) (if there is any)

**Estimation of Scale of Noise:**
- **Input:** Range data of each fine segment (resulted from previous stage)
- **Method:** Least-square regression
- **Condition:** No Condition
- **Output:** Estimated scale of noise (\( \hat{\sigma} \)) and parameters of each segment

**End Segmentation**

---

*Figure 6.2: Cascade of Hierarchical Robust Segmentation (HRS). \( K \) is the relative size of smallest structure to be segmented. \( \sigma \) is range accuracy of measurement system and is varied by instrument.*
6.3.1 How does the hierarchical scheme reduce computation costs?

A general drawback of using robust estimation is that no explicit formula exists to solve the objective function optimisation problem for most of the estimators. An accurate solution can only be determined by searching in the space of all possible estimates. Consider all estimates determined by all possible $p$-tuples (i.e. three-tuples for 3D segmentation) of data points. In an exhaustive search scheme to minimise the median of square residuals (using the well-known LMS estimator), there are $(n_p = n!/(p!(n-p)!))$ $p$-tuples and it takes $O(n_p)$ time to find the median of the residuals of the whole data for each $p$-tuple and its cost will therefore increase very fast with $n$ (number of data points) and $p$.

In order to reduce the cost of computation, instead of searching all space, one can apply random sampling as described in Chapter 5. However random sampling by itself is again costly. As mentioned previously, because of the large disparity in the size of the objects and the existence of fine details in the structures, a very small ratio of inliers have to be assumed when using the random sampling. The number of required random samples is given by Equation 5.2, where $(1-\epsilon)$ is the smallest possible ratio of inliers in the application (denoted by $K$). Since $K << 1$ the denominator of Equation 5.2 can be closely approximated with $K^3 (log (1-x) \approx x$ for small $x$) and therefore:

\[
N \approx \log(1-P)K^{-3}. \tag{6.1}
\]

Let assume that the minimum ratios of inliers in the three levels of the proposed hierarchical scheme are: $K_1$, $K_2$ and $K_3$, then the total number of random samples required by the hierarchical technique is:

\[
N_{HRS} = \log(1-P)[K_1^{-3} + K_2^{-3} + K_3^{-3}] \tag{6.2}
\]
On the other hand, the segmentation process would be able to extract segments as small as \( K = K_1 K_2 K_3 \) times the total number of data. Therefore, the number of samples required to segment the same size of structure by direct usage of a robust estimator is:

\[
N_{\text{dir}} = \log(1 - P) K_1^{-3} K_2^{-3} K_3^{-3}.
\]

From Equations 6.2 and 6.3, it is evident that for small \( K_1 \) and \( K_2 \) and \( K_3 \) values, the number of random samples required by the proposed hierarchical scheme is far less than a straightforward random sampling approach. For instance, the experimental results (see Table 6.2 in Section 6.4.2 for more details) show that for segmenting fine structures as small as 1.2 per cent of the whole data (where \( K_1 = 0.2, K_2 = 0.15 \) and \( K_3 = 0.4 \)), with the HRS method, the number of required samples (and consequently the computation time) is reduced by a factor of 1325 compared to MSSE.

6.4 Experimental results

The HRS technique has been applied to a number of both synthetic and real range datasets to evaluate the performance of the proposed algorithm. The synthetic data includes complex multiple structures. Real data include range data of real-world scenes (here, historical building exteriors) captured by the experimental rangescanner device, developed by the author (see Chapter 4), as well as two commercially available rangescanner systems. The experimental results are presented in Section 6.4.1 and 6.4.2. All experiments are performed by a desktop PC with a dual core 2.2 GHz CPU and 3Mb of RAM in 64 bits Matlab environment.

6.4.1 Synthetic data

Three-dimensional synthetic data containing four structures is generated here. As illustrated in Figure 6.3, the inliers belong to the planar surfaces - each representing a structure - , containing 1000, 800, 400 and 40 data points corrupted by an additive Gaussian noise of
The inliers are mixed with 30 gross outliers. The synthetic dataset is designed to represent data of a complex multi-structure scene where outliers and disparities in the population (i.e. size) exist. In this dataset, the number of outliers is close to the number of data of the smallest structure which is about 1.7 per cent of total data. The objective of the experiment on this dataset is to extract all structures and eliminate outliers.

The first experiment was performed with $P = 0.99$ and $K = 0.017 (40/2270)$. From Equation 5.2, the number of random samples required for segmentation using MSSE is $10^6$. The output of this experiment was ‘out of memory’ error returned by the segmentation software, mainly due to the large size of random samples and the limitation of the computation.

In the second experiment, the size of the smallest structure to be segmented is again considered to be $K = 0.017 (40/2270)$. Therefore, the number of samples required to extract the smallest structure using the MSSE, is 237,138 samples with a modest probability of
selecting good sample \((P = 0.85)\). The process takes approximately six hours for an ordinary computer to complete the segmentation task. Besides the high computational cost, the resulting segmentation detects seven segments instead of four, which means that under-segmentation has been occurred.

In the third experiment with \(P=0.85\) and \(K=0.05\), the number of required samples for segmentation using MSSE, significantly decreased to about \(15 \times 10^4\) and the process takes about two hours. Figure 6.4 shows the result of robust segmentation. It is observed that although the computational cost (in terms of time and memory) is significantly decreased in this experiment, the smallest structure could not be extracted. In addition, a false surface component was extracted (the red and blue structures in Figure 6.3 are detected as one structure). This problem is referred to as a ‘bridging fit’ error [171] and the output of range segmentation evaluated as over-segmentation (see Chapter 3).

A bridging fit occurs when two distinct neighbouring surfaces are segmented as one region and it occurs here due to the lack of enough samples with \(P\) set to 0.85 and size of the
smallest structure of interest set to 5 per cent of whole data. As mentioned in Section 6.3, to overcome this problem and delineate all possible features that exist in such a dataset; a single global use of robust estimation is not adequate. A step-wise robust approach needs to be considered for the segmentation problem of complex and large datasets.

To evaluate the performance of the HRS, the proposed algorithm (as detailed in Figure 6.2) has been applied to the synthetic dataset. Figures 6.5-6.7 show the steps of the HRS on the synthetic data of multiple structures presented in Figure 6.3. Figure 6.5 shows the result of segmentation of coarse/large structure using the HRS algorithm. In this stage, $K_1 = 0.15$ and $P = 0.95$. As shown in this figure, only one structure is extracted and other structures are separated as a very coarse segment that needs to be further segmented.

In the second stage, the segmentation algorithm is applied to the remaining data resulting from the first level of segmentation (those data is shown in green points in Figure 6.5). In this stage, $K_2 = 0.1$ and $P = 0.95$, where $K_2$ is the size of smallest structure in the remaining data and $P$ is the probability of selecting good sample. Figure 6.6 illustrates the result of the second level of segmentation.

In the last stage of segmentation, the smallest (or finest) segment is detected and the remaining gross outliers are eliminated from the dataset. This segmentation result is shown in Figure 6.7. In this stage, $K_3 = 0.1$ (size of the smallest structure one is interested in, in the remaining data) and $P = 0.95$. 
Chapter 6 — Hierarchical Robust Range Segmentation

Figure 6.5: First level of robust segmentation.

Figure 6.6: Second level of robust segmentation
6.4.2 Real data

To show the performance of the proposed algorithm on real data, a number of real world experiments are conducted and detailed here (Figures 6.8, 6.9 and 6.10). The exteriors of the chosen buildings are highly structured with many large planar objects such as walls, doors and roofs as well as smaller planar objects such as stairs, doorways and decorative parts. Range data of the first experiment is captured from the front view of a large building called the Shrine of Remembrance (Melbourne, Australia) by a Riegl (LMS-Z210) laser rangescanner. The building is pictured in Figure 6.8 (a – left). The scanned range image of the building is sampled on a 250 × 382 grid and contains almost $10^5$ data points. The angular resolution of the scanner was set to be 0.1 degree and its measurement error is typically 10mm. The proposed algorithm has all been implemented in MATLAB. The original range image and the results of the first and last stages of the segmentation strategy are shown in Figure 6.8 (b) and (c). To show the high accuracy of the proposed segmentation algorithm,
the decorative part of the front exterior and its segmentation outcome are magnified in Figure 6.8 (d).

Figure 6.8: Hierarchical range segmentation of the front side of the Shrine of Remembrance, Melbourne – Australia. Range data of the building is captured by Riegl (LMS-Z210) laser rangescanner: a) left) Intensity image, a) right) Range image, b) First level of segmentation, c) Final result of hierarchical segmentation. d) All possible detail (planar and decorative) of the pediment is successfully segmented. This figure is best viewed in colour.
To highlight the advantages of the hierarchical strategy, the results of the direct and the step-wise implementations of the robust range segmentation technique have been compared by using the range image of a typical building. Table 6.1 and 6.2 summarise the various outcomes of each approach for different values of $K$ (the proportional size of the smallest data group that would be considered a structure).

As shown in Table 6.1, when the value of $K$ is decreased to extract more details (finer structures), the required computational cost is significantly increased to levels that would be considered impractical for most vision applications. At the same time, the value of $\sigma$ (estimated noise scale) for each segment also decreases, pointing to the fact that the segmentation has become more accurate with the smaller values of $K$. The under-segmentation problem however worsens when $K$ is fairly small.

Table 6.1: Outcome of direct implementation of the robust range segmentation algorithm for the range data of the Shrine of Remembrance with different values of $K$ (size of the smallest structure)

<table>
<thead>
<tr>
<th>$K$</th>
<th>No. of Samples</th>
<th>Segmentation Time (in seconds)</th>
<th>No. of Segments</th>
<th>Segmentation Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>84</td>
<td>28</td>
<td>2</td>
<td>No fine details is detected</td>
</tr>
<tr>
<td>0.2</td>
<td>286</td>
<td>36</td>
<td>3</td>
<td>No fine detail is detected</td>
</tr>
<tr>
<td>0.1</td>
<td>2,301</td>
<td>150</td>
<td>8</td>
<td>No fine detail is detected</td>
</tr>
<tr>
<td>0.08</td>
<td>4,496</td>
<td>398</td>
<td>16</td>
<td>Moderate number of fine details are detected</td>
</tr>
<tr>
<td>0.05</td>
<td>13,419</td>
<td>1,661</td>
<td>15</td>
<td>Moderate number of fine details are detected</td>
</tr>
<tr>
<td>0.02</td>
<td>287,821</td>
<td>44822 (~13 hours)</td>
<td>29</td>
<td>Most details are detected</td>
</tr>
<tr>
<td>0.01</td>
<td>2,302,583</td>
<td>Stopped due to the computational limitations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2 shows the results of the hierarchical approach to the robust range segmentation of the Shrine of Remembrance (see Figure 6.8) requiring a three level pyramid. At the first level, the algorithm has focused on the large/coarse segments (e.g. structures that contain at least 20 per cent of the whole population) and has separated data into four parts.
Stages two and three have further refined those parts into smaller/finer segments where a very accurate segmentation is achieved at its final stage. This table also shows that the hierarchical approach to the segmentation drastically decreases the required time of computation (three minutes versus 13 hours) and has taken full advantage of the high accuracy that the MSSE can produce.

**Table 6.2: Outcome of hierarchical implementation of robust segmentation algorithm for the range data of the Shrine of Remembrance**

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>$K$</th>
<th>Number and Quality of Generated Segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>0.2</td>
<td>Segment I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma &gt;&gt; 0.01$</td>
</tr>
<tr>
<td>Second</td>
<td>0.15</td>
<td>Segments I-1 and I-2 to Segment II-7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma &gt; 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma &gt; 0.01$</td>
</tr>
<tr>
<td>Third</td>
<td>0.4</td>
<td>Segment I-1.1 to Segment I-1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma &lt; 0.01$</td>
</tr>
</tbody>
</table>

Segmentation time = 191 (s)

In the second experiment, the proposed segmentation algorithm was applied to the geometric data of the exterior of the Melbourne Exhibition Building shown in Figure 6.9(a). The dataset was captured by the experimental laser rangescanner system that has been described in Chapter 4. Angular resolution and measurement error of the system are 0.1 degree and 300mm respectively and the range image is sampled on a 320×615 lattice. Figure 6.9 (b) and (c) illustrate the first and final level of segmentation for this set of data.
The results show that although the measurement noise in this dataset is much higher than the data obtained by commercial systems (mostly due to the limitations of the system components), the range segmentation has been successful in extracting fine details embedded in the front side of the building.

Range data of the Notre-Dame Church, France, is used for the third test (Figure 6.10). The range data is captured by Leica (HD2500) laser rangescanner and contains $5 \times 10^5$ data points. The measurement error of the system is typically 4mm in depth and its angular resolution generates about 25mm space between the samples at 20 metres. Figure 6.10 (b) and (c) show the first and last levels of range segmentation respectively. Figure 6.10 (d), (c) and (f) are close views of the different parts of the building that have been segmented to demonstrate the level of detail that can be extracted by the algorithm. It is important to note that although parts of data, shown in Figure 6.10 (e), are missing due to the limitations of measurement for reflective surfaces using laser technology, the segmentation algorithm has been able to correctly partition the existing parts.
Figure 6.9: Hierarchical segmentation of the back side of the Melbourne Exhibition Centre – Australia. Range data of the building is captured by the experimental laser rangescanner explained in Chapter 3: a) Intensity image of the building. b) First level of segmentation (coarse segmentation). c) Final result of hierarchical segmentation. This figure is best viewed in colour.
Figure 6.10: Hierarchical segmentation of the south view of Notre-Dam Church – France. Range data of the church captured by Leica HDS2500 laser rangescanner. a) Intensity image of the church [172]. b) First level of segmentation (coarse segmentation). c) Final result of hierarchical segmentation d, e, f) All possible details (planar and decorative) of different parts of building are successfully segmented. This figure is best viewed in colour.
Chapter 6 — Hierarchical Robust Range Segmentation

6.5 Chapter review

This chapter has introduced a novel method of partitioning the range data of significant buildings into planar patches using a hierarchical COF (coarse-to-fine) segmentation approach. The segmentation algorithm is based on a high breakdown robust estimator and extracts detailed planar patches of large and complex datasets with very modest computational costs. In one of the case studies presented here, it has taken around 190 seconds to segment 28 planar patches embedded in the range dataset of a decorative monument. The same task takes approximately 13 hours when a similar robust segmentation technique is applied directly to this dataset. The experimental results also show that the segmentation outcomes of the proposed method are more accurate and less prone to the usual over or under segmentation issues.
Chapter 7

Conclusions and Future Directions

This chapter concludes the thesis by outlining its key contributions to the range data acquisition and segmentation problem and identifies promising avenues for future research.

7.1 Conclusions

This thesis has explored a sample computer vision system and its associated problems from data acquisition to segmentation. The domain of this research was restricted to applications relating to large building exteriors. A cost-effective, robust technology and technique for the accurate measurement and segmentation of geometric details embedded in the exterior surfaces of large buildings were devised. The relevant literature was also reviewed (Chapter 2 and 3).

The thesis consists of three main parts. In the first part (Chapter 4), the need for accurate photorealistic representation of the surrounding environment was addressed. A versatile large-scale rangescanner system was implemented using state-of-the-art rangefinding and scanning technology. The main purpose of this part was to develop a
reliable and accurate 3D scanner system for large-scale objects with minimum complexity and cost. The system design had four stages. The first stage was the design of the range data acquisition system which consists of a time-of-flight laser rangefinder (LRF) and a programmable pan-tilt unit (PTU). The sensors were aligned, and the hardware and software were designed and developed to acquire range and intensity images of building exteriors. In the second stage of the design, the range and intensity data were fused (integrated/registered) using a novel yet simple calibration algorithm. The results of range and intensity data fusion show that the relationship between range and intensity data of the same scene is highly consistent. In the third stage, the sources of 3D measurement errors were studied and the key operational parameters of the experimental laser rangescanner system were estimated. In the last stage, the laser rangescanner design was verified using two methods of verification — application and equivalence checking verification methods. The first method of verification was used to examine the performance of the laser rangescanner system in terms of the usability of its output. To this purpose, a parametric segmentation technique has been adopted to segment the point cloud data into planar surfaces. The experimental results show that the resulting range data is acceptable, as the range segmentation algorithm has successfully detected the planar and coplanar areas. The second method of verification was employed to examine the accuracy of the resulting range data. To determine this, the output of the experimental system was compared with the same range data set obtained by an accurate commercial laser rangescanner. The result of the comparison shows a very close agreement between the two sets of data.

In the second part of this thesis (Chapter 5), problems associated with range data acquisition and the processing of large building exteriors was studied. The key challenges include large size of data sets, significant disparities in the size and depth of the objects, the existence of moving objects (e.g. passengers, vehicles and birds) and construction error. Window glass and specular surfaces are also found as another source of uncertainty in the range data captured from outdoor scene, obtained by a laser time-of-flight technique.

In the last part (Chapter 6), a computationally cost-effective and robust segmentation technique capable of extracting geometric details embedded in the 3D data of large building exteriors was developed. The segmentation algorithm, called *Hierarchical Robust*
Chapter 7 — Conclusions and Future Directions

Segmentation (HRS), uses a high breakdown robust estimator in a hierarchical coarse-to-fine approach. The proposed algorithm was tested on several range data sets acquired by different laser rangescanners (i.e. the experimental and two commercial 3D laser scanners) The experimental results show that the proposed algorithm is capable of extracting coarse and fine details from range data in a relatively short period of time. In addition, the segmentation outcomes of the proposed method are less prone to the usual over or under segmentation issues.

It is necessary to note that the technology and technique presented in this thesis can be used in any application where the accurate measurement and fine segmentation of large-scale man-made objects (e.g. buildings interior and large statues) is required.

7.2 Future directions

The inherent complexity of the computer vision system, especially for large-scale and heterogeneous real world applications; and the availability and richness of the acquired range and intensity data sets open a number of future venues along the lines of the technology and technique presented in this thesis.

In the context of range data acquisition system:

• An error model can be developed based on the key practical parameters of the rangescanner system, obtained from the characterisation, to correct systematic errors. This model can be general and applicable to the output of any laser rangescanner system, regardless of the scanning and rangefinding method used. This model can be applied to the raw range measurements to improve the accuracy of the 3D measurement.

• A single range image of large building exteriors is not sufficient to generate a complete 3D model of the building. Several range and intensity images need to be obtained and combined (or registered) to produce a photorealistic model of the scene. To this purpose, the range image registration can use the proposed range segmentation scheme presented in this thesis. Due to the large size and complexity of the data of large buildings, a robust COF
approach may be used for range data registration. In addition, sensor planning can be used for enhancing incomplete scene models by carefully choosing the position of the sensor.

In the context of range data segmentation:

- The range segmentation algorithm presented in this thesis is able to be successfully segment planar or close to planar surfaces. However, for applications where non-planar forms are also of importance, the proposed hierarchical segmentation can be extended to include a model selection strategy similar to the one introduced in [173].

- The range data of large building exteriors also includes elements such as the ground, shrubs and crowds. For those applications that only the building exteriors is of importance, a robust segmentation method can first be used to classify data to its components (e.g. building, vegetation, ground and sky), then the extracted building can be used as the input for the hierarchical segmentation, proposed in this thesis.

- The robust hierarchical segmentation scheme presented in this thesis used the MSSE estimator. A comparative analysis of hierarchical range segmentation based on other high breakdown estimators (such as PbM and HBM) can be conducted.

- To solve the objective function of the estimator used for range segmentation (i.e. MSSE), random sampling approach has been used. Random sampling is computationally expensive (it may even be impossible for very large sets of data) and it can be replaced by other methods such as guided sampling [174], direct search [175] and genetic algorithms[176, 177].
Bibliography


349-356.


Acquisition from Range Images," Department of Artificial Intelligence, University of Edinburgh 1995.


[123] R. Hoseinnezhad and A. Bab-Hadiashar, "Consistency of Robust Estimators in Multi-structural Visual Data Segmentation," Pattern Recognition, vol. 40, pp. 3677-


Appendices
Appendix A

Laser Safety Considerations

Adapted from Acuity Laser Measurement: www.acuiltylaser.com/resources/safety.shtml

Laser devices are divided into “classes” based on the power of the laser, whether it is a visible or IR laser, and the potential exposure duration. Class I devices are eye-safe under any circumstances. The maximum permissible output varies with the laser light frequency and other factors. Class II devices are visible lasers with output of less than 1mW. Classifications apply to both pulsed and continuous wave lasers, with various formulae for determining class. For continuous wave lasers, Class IIIa lasers are visible lasers with output power of more than 1mW but less than 5mW, as measured through a 7 millimetre aperture. Class IIIb lasers are those with output above 5mW, or any laser outside the visible frequency band that is not unconditionally eye safe. Class IIIb extends up to 500mW output power.

Visible and Infrared Laser Beams

Both visible and infrared (IR) lasers are used in distance measurement. For some applications, the advantage of being able to see the spot is an advantage, while others do not want the spot to be visible. The IR lasers are slightly more sensitive and accurate than the visible version, and IR models benefit from a wider range of available laser powers. At close ranges, it is possible to view an infrared laser spot on a target surface. The easiest way to view an infrared spot is to view it through a CCD digital camera or to put an IR detector card in front of the spot. The human eye cannot see IR, but the digital camera can.
Appendix B

Multipurpose Laser Distance Meter

Model LaserAce® IM Mechanical Dimension
Appendix C

Computer Controlled Pan-Tilt Unit

Model PTU-46-70-*N Mechanical Dimension
Appendix D

B+W Infrared Filter 092 -Specifications


The nearly opaque Infrared Filter 092, which looks dark purplish red when held in front of a light source, blocks visible light up to 650 nm, and passes only 50% of the radiation just below 700 nm (thus the dark red colour). From 730 nm to 2000 nm, transmission is greater than 90%. This makes photographs of pure red and infrared images possible making the best use of the relatively low sensitivity of infrared films. As the sensitisation of infrared black-and-white films barely extends beyond 1000 nm, the red portion that is transmitted still makes a relevant contribution to the exposure. This makes the 092 the preferred filter for pictorial photography on IR black-and-white film. Its filter factor is 20 to 40. The transmission curve of the filter is shown in the following figure.
Invisible infrared radiation

The spectral range that is visible to the human eye ends at a wavelength of approximately 50 nm. This is where infrared radiation begins (only certain birds can see portion of infrared radiation that better penetrate haze). Most films parallel the spectral response of our eyes. But there are special infrared-sensitive films for colon-and black-and-white photographs which depending on their sensitization, react to 850 nm, 900 nm or nearly 1000 nm radiation. Like other films, these are also very sensitive to visible light. If we wish to image only in the infrared, filters must be used to suppress the visible, or to attenuate it strongly, so that the (weak) image produced by the infrared radiation will be sufficiently prominent.
Appendix E

Homogeneous Coordinates

Obtained from CVonline - by Bill Triggs

Suppose that the camera is at the origin (0, 0, 0). The ray represented by homogeneous coordinates \((X, Y, T)\) is that passing through the 3D point \((X, Y, T)\). The 3D point \(\lambda (X, Y, T) = (\lambda X, \lambda Y, \lambda T)\) also lies on (represents) the same ray, so we have the rule that rescaling homogeneous coordinates makes no difference:

\[
(X, Y, T) \sim \lambda (X, Y, T) = (\lambda X, \lambda Y, \lambda T).
\]

If we suppose that the image plane of the camera is \(T=1\), the ray through pixel \((x, y)\) can be represented homogeneously by the vector \((x, y, 1) \sim (xT, yT, T)\) for any depth \(T \neq 0\).

Hence, the homogeneous point vector \((X, Y, T)\) with \(T \neq 0\) corresponds to the inhomogeneous image point \((X/T, Y/T)\) on the plane \(T=1\).

But what happens when \(T=0\) \((X,Y,0)\) is a valid 3D point that defines a perfectly normal optical ray, but this ray does not correspond to any finite pixel: it is parallel to the plane \(T=1\) and so has no finite intersection with it. Such rays or homogeneous vectors can no longer be interpreted as finite points of the standard 2D plane. However, they can be viewed as additional "ideal points" or a limit as \((x, y)\) recedes to infinity in a certain direction:

\[
\lim_{T \to 0} \left( \frac{X}{T}, \frac{Y}{T}, 1 \right) \sim \lim_{T \to 0} (X,Y,T) = (X,Y,0)
\]

We can add such ideal points to any 3D plane. In 2D images of the plane, the added points at infinity form the plane's "horizon". We can also play the same trick on the whole 3D space, representing 3D points by four homogeneous coordinates:

\[(X, Y, Z, T) = (\lambda X, \lambda Y, \lambda Z, \lambda T) \sim (X/T, Y/T, Z/T, 1).\]
Appendix F

Camera Calibration Result
% Intrinsic and Extrinsic Camera Parameters
% This script file can be directly executed under Matlab to recover the camera intrinsic and extrinsic parameters.
% IMPORTANT: This file contains neither the structure of the calibration objects nor the image coordinates of the calibration points.
% All those complementary variables are saved in the complete matlab data file Calib_Results.mat.
% For more information regarding the calibration model visit http://www.vision.caltech.edu/bouguetj/calib_doc/

%-- Focal length:
fc = [ 968.985461566317550 ; 1062.876259169658400 ];

%-- Principal point:
cc = [ 352.900704875125830 ; 270.870434347690720 ];

%-- Skew coefficient:
alpha_c = 0.000000000000000;

%-- Distortion coefficients:
kc = [ -0.179995453454130 ; -0.141294595257734 ; -0.003449635034187 ; -0.002026472282628 ; 0.000000000000000 ];

%-- Focal length uncertainty:
fc_error = [ 14.180392480670905 ; 15.372585838410146 ];

%-- Principal point uncertainty:
cc_error = [ 15.738411602057747 ; 22.198120299951711 ];

%-- Skew coefficient uncertainty:
alpha_c_error = 0.000000000000000;

%-- Distortion coefficients uncertainty:
kc_error = [ 0.055089980449442 ; 0.523431596667884 ; 0.002822378303759 ; 0.003459410487333 ; 0.000000000000000 ];

%-- Image size:
nx = 720;
ny = 576;

%-- Various other variables (may be ignored if you do not use the Matlab Calibration Toolbox):
%-- Those variables are used to control which intrinsic parameters should be optimized
n_ima = 15;  % Number of calibration images
est_fc = [ 1 ; 1 ];  % Estimation indicator of the two focal variables
est_aspect_ratio = 1;  % Estimation indicator of the aspect ratio fc(2)/fc(1)
center_optim = 1; % Estimation indicator of the principal point
est_alpha = 0; % Estimation indicator of the skew coefficient
est_dist = [1; 1; 1; 1; 0]; % Estimation indicator of the distortion coefficients

%-- Extrinsic parameters:
%-- The rotation (omc_kk) and the translation (Tc_kk) vectors for every calibration image and their uncertainties

%-- Image #1:
omc_1 = [-2.203504e+000; -2.166118e+000; 3.177037e-001 ];
Tc_1  = [-5.972853e+001; -7.800750e+001; 1.052056e+003 ];
omc_error_1 = [1.711544e-002; 1.470947e-002; 3.459550e-002 ];
Tc_error_1  = [1.707125e+001; 2.194213e+001; 1.595231e+001 ];

%-- Image #2:
omc_2 = [-1.920762e+000; -1.787161e+000; 7.527158e-001 ];
Tc_2  = [3.305909e+001; -6.096367e+001; 9.218776e+002 ];
omc_error_2 = [1.582614e-002; 1.026243e-002; 2.567660e-002 ];
Tc_error_2  = [1.497175e+001; 1.922447e+001; 1.339780e+001 ];

%-- Image #3:
omc_3 = [1.953013e+000; 1.961894e+000; -7.280352e-001 ];
Tc_3  = [-5.744603e+001; -4.236694e+001; 8.497827e+002 ];
omc_error_3 = [1.191505e-002; 1.676087e-002; 2.537502e-002 ];
Tc_error_3  = [1.377155e+001; 1.771846e+001; 1.234154e+001 ];

%-- Image #4:
omc_4 = [2.932295e+000; -1.851088e-001; 7.981097e-001 ];
Tc_4  = [-8.002802e+001; 7.383227e+001; 8.160785e+002 ];
omc_error_4 = [2.041346e-002; 9.398826e-003; 2.494133e-002 ];
Tc_error_4  = [1.329513e+001; 1.707607e+001; 1.221119e+001 ];

%-- Image #5:
omc_5 = [2.996951e+000; -1.874005e-001; 1.895532e-001 ];
Tc_5  = [-7.175384e+001; 1.110328e+002; 7.137153e+002 ];
omc_error_5 = [2.078387e-002; 3.647222e-003; 2.592983e-002 ];
Tc_error_5  = [1.168544e+001; 1.499297e+001; 1.068850e+001 ];

%-- Image #6:
omc_6 = [2.979132e+000; -6.712011e-002; -4.347889e-001 ];
Tc_6  = [8.769393e+001; 1.380759e+002; 8.411585e+002 ];
omc_error_6 = [1.971568e-002; 3.702493e-003; 2.828190e-002 ];
Tc_error_6  = [1.379727e+001; 1.774556e+001; 1.233682e+001 ];

%-- Image #7:
omc_7 = [1.881850e+000; 1.875743e+000; 3.463279e-001 ];
Tc_7  = [3.775256e+001; -7.805471e+001; 7.755419e+002 ];
omc_error_7 = [1.745472e-002; 1.077277e-002; 2.642377e-002 ];
Tc_error_7  = [1.264361e+001; 1.619809e+001; 1.241886e+001 ];

%-- Image #8:
omc_8 = [ -1.747472e+000 ; -1.750983e+000 ; 7.780566e-001 ];
Tc_8  = [ 7.161829e+001 ; -1.135984e+002 ; 8.664500e+002 ];
omc_error_8 = [ 1.584374e-002 ; 1.044173e-002 ; 2.438511e-002 ];
Tc_error_8  = [ 1.414648e+001 ; 1.811453e+001 ; 1.297830e+001 ];

%-- Image #9:
omc_9 = [ 3.113671e+000 ; 3.486038e-003 ; -1.666291e-001 ];
Tc_9  = [ -2.102021e+001 ; 1.017762e+002 ; 7.271846e+002 ];
omc_error_9 = [ 2.036294e-002 ; 1.966777e-003 ; 3.132511e-002 ];
Tc_error_9  = [ 1.184225e+001 ; 1.520317e+001 ; 1.055261e+001 ];

%-- Image #10:
omc_10 = [ 3.111207e+000 ; 7.939210e-002 ; -2.253115e-001 ];
Tc_10  = [ 1.698430e+001 ; 9.377105e+001 ; 1.164627e+003 ];
omc_error_10 = [ 3.102293e-002 ; 3.227327e-003 ; 5.459200e-002 ];
Tc_error_10  = [ 1.893780e+001 ; 2.433633e+001 ; 1.703524e+001 ];

%-- Image #11:
omc_11 = [ -3.077535e+000 ; -6.184010e-003 ; 1.419933e-001 ];
Tc_11  = [ -3.683375e+001 ; 8.867541e+001 ; 1.196103e+003 ];
omc_error_11 = [ 3.144121e-002 ; 2.979158e-003 ; 5.331546e-002 ];
Tc_error_11  = [ 1.943513e+001 ; 2.494728e+001 ; 1.701249e+001 ];

%-- Image #12:
omc_12 = [ 2.733257e+000 ; -5.029761e-002 ; -2.565081e-001 ];
Tc_12  = [ 2.507139e+001 ; 7.855275e+001 ; 1.221545e+003 ];
omc_error_12 = [ 2.201667e-002 ; 5.398612e-003 ; 2.450952e-002 ];
Tc_error_12  = [ 1.986162e+001 ; 2.553628e+001 ; 1.833697e+001 ];

%-- Image #13:
omc_13 = [ 2.619713e+000 ; -5.306421e-002 ; -8.197556e-001 ];
Tc_13  = [ -2.657130e+001 ; 1.284471e+002 ; 8.243016e+002 ];
omc_error_13 = [ 2.06195e-002 ; 8.527801e-003 ; 2.151806e-002 ];
Tc_error_13  = [ 1.346235e+001 ; 1.724284e+001 ; 1.168708e+001 ];

%-- Image #14:
omc_14 = [ 2.474195e+000 ; -1.479150e-001 ; -4.296844e-001 ];
Tc_14  = [ -4.752950e+001 ; 1.008828e+002 ; 7.122567e+002 ];
omc_error_14 = [ 2.073610e-002 ; 7.242594e-003 ; 2.009777e-002 ];
Tc_error_14  = [ 1.161303e+001 ; 1.492414e+001 ; 1.067109e+001 ];

%-- Image #15:
omc_15 = [ 1.809502e+000 ; 1.840367e+000 ; -3.649776e-001 ];
Tc_15  = [ -4.897574e+001 ; -4.396542e+001 ; 8.364753e+002 ];
omc_error_15 = [ 1.438423e-002 ; 1.386410e-002 ; 2.395815e-002 ];
Tc_error_15  = [ 1.356563e+001 ; 1.743977e+001 ; 1.241381e+001 ];