3D Shape Reconstruction By Dynamic Sensing With A Range Sensor

レーザ・レンジ・センサを用いた形状復元のための
動的計測に関する研究

BY

Xiangqi Huang

A Doctoral Dissertation

Submitted to the Graduate School of
the University of Tokyo

東京大学
THE UNIVERSITY OF TOKYO

JANUARY 2013

Thesis Supervisor: Takafumi KOSEKI
Thesis Co-supervisor: Katsushi IKEUCHI

6

Committee:

Takafumi KOSEKI (Chair)
Katsushi IKEUCHI
Yoichi SATO
Yoichi HORI
Takashi KUBOTA
Tatsuaki HASHIMOTO

Supervisor: Takafumi KOSEKI

Co-supervisor: Katsushi IKEUCHI

ABSTRACT

Building 3D shape models of interesting targets and getting their localization are very important fundamental tasks in the fields like digital culture heritage, robotics, mixed reality, medical application and so on.

Most commonly used sensors for 3D modeling are camera and laser range sensor. Compared with camera, range sensor has the following advantages: 1. It can easily get direct and highly accurate 3D data. 2. Range data is stable, whereas the 3D modeling techniques based on 2D cameras are much more sensitive to illumination, shadows, scale and pose. However, range sensor costs a relatively long time of data acquisition. This causes data distortion and cannot be ignored especially in the case of continuously gathering data on a moving platform.

Existing common solutions to this problem are either taking ”stop-scan-go” strategy to avoid distortion or correcting the sensor motion using secondary sensors like GPS, inertial sensor, camera or even another range sensor. Some other researches apply linearization, discretization and other specified constraints to the problem to achieve compromised results.

This thesis focuses on efficient and accurate 3D shape reconstruction under a moving sensing system only with a single range sensor. Different from previous solutions, the proposed system works in an efficiently continuous manner. We don’t have to stop the platform to obtain an stationary scan. Sensor can move under a reasonable motion mode and simultaneously scan the target or environment. Data distortion caused by continuous movement will be rectified. This continuous manner is much more efficient and attractive in practical applications.

According to the different prior conditions of targets, we divide the problem to two categories. The first one is based on the assumption that the target shape model is known. The second one is focusing on the unknown targets.

For the known targets, since there is no need to build 3D model for the object, our interesting mainly focus on how to accurate localize them. We propose a method which utilizes the prior shape model to estimate sensor motion. Based on the estimated motion model, distorted measurements can be rectified and accurate localization of targets can be achieved. As an application, we build a real-time pile driver position system using laser range finder. The detail is described is Chapter 2.

For the unknown targets, we develop a polynomial fitting method based on corresponding points to estimate 6 DOF motion parameters of a moving range sensor. The 3D model can be reconstructed according to highly accurate sensor motion estimation. When using a 2D scanning sensor instead of a 1D scanning sensor, the same region is measured in multiple times when the sensor moves. We will show that we can reconstruct the sensor motion and the scene from only the measured coordinates and times of the same set of points. This reconstruction is intrinsic, which relies on only the intrinsic properties of the distortion, and not relying on the extrinsic information from other sensors. Firstly, for robust estimating sensor movement, we model the sensor motion using polynomial with respect to time. Secondly, to build simple representation of common motion like uniform circular motion or rectilinear motion, we introduce twist coordinates for the representation of rigid body transformation. This method doesn't need the secondary sensor and is not limited with specific environment features. Without linearization of constraint and discretization of trajectory, distorted data is accurately rectified. Details are in Chapter 3.

To obtain the corresponding constraints described in chapter 3, we propose a 3D affine invariant shape feature which is designed for the deformed 3D data collected by moving range sensor. Firstly two Morse function which measure the object shape under different conditions are designed. A disconnected graph based method is proposed to extract stable affine invariant region feature in multiple scale. Applied a moment-based affine transformation, the deformed region feature can be normalized. A multiple scale Spin Image is designed to describe the normalized region feature. The synthetic data gives robust feature detection and

matching results. Details are presented in Chapter 4.

Finally, we present our conclusions and summarize our possible future works in Chapter 5.

Acknowledgements

First I would like to express my gratitude to my supervisors, Prof. Hideki Hashimoto, Professor Katsushi Ikeuchi, and Prof. Takafumi Koseki. They gave me a precious chance to study in the University of Tokyo, and give me a lot valuable advices and help on my research.

I would like to record my gratitude to Prof. Katsushi Ikeuchi who is not just a great researcher but also wise tutor. He gave me a wonderful opportunity and take me into an interesting field, computer vision. Not only has he direct me to pursue my Ph.D., he has offered me an open door and an open mind, encouraging me and providing me with countless ideas through discussions. What most worth appreciating is the time when he spent hours weekly with me discussing my research face to face. I would never forget all his selfless help.

I would like to record my gratitude to Dr. Bo Zheng for his advice and guidance as well as giving me extraordinary experiences through out the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways. His truly scientist intuition has made him as a constant oasis of ideas and passions in research, which exceptionally inspire and enrich my growth as a student, a researcher and a scientist want to be. I am indebted to him more than he knows.

I would like to express my gratitude to Dr. Takeshi Masuda for his advice and guidance. His rigorous scholarship and patience to details in research give me great example and inspiration.

I would like to express my gratitude to Dr. Atsuhiko Banno for his and guidance and help in research and programming.

I would like to thank Professor Takeshi Oishi and Dr. Shin Ono who give me a lot useful comments and suggestions and have always been so patient in weekly research meeting.

Many thanks go to the people that I have been working with, the current members of the Computer Vision Laboratory, the former members of the Hashimoto Laboratory at the University of Tokyo, Dr.Fumihiro Inoue and other members from Obayashi-gumi Inc, and members from IHI Corporation.

Special thanks go to Bjoern Rennhak and Shaodi You who helped me a lot with my paper writing. They kindly improved my writing and gave me many appropriate suggestions.

I would also like to thank my committee members, Prof. Yoichi Sato, Prof. Yoichi Hori, Prof. Takashi Kubota, and Prof. Tatsuaki Hashimoto, for giving valuable advice on this thesis.

I would like to thank many people in my life. In particular, I am very grateful to all the members of the Chinese Tennis Club in the University of Tokyo, to which I have belonged for 2 years, for their contributions to the enjoyment of my student life.

Last but not least, I would like to thank my mother for always being patient and supporting me during my student life. It has been a long effort but worth doing, and I could not finish it without your support. I dedicate this dissertation to my mother.

January 2013

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Building accurate 3D models of interesting targets is a very important fundamental task in the fields like using range finder in digital culture heritage, robotics, mixed reality, medical application and so on. [4] [6] [64] [44]

When investigating 3D reconstruction, it is necessary to describe the kind of data being considered. Most commonly used sensors for 3D modeling are camera and laser range sensor. Compared with camera, range sensor has the following advantages: 1. It can easily get direct and highly accurate 3D data. 2. Range data is stable, whereas the 3D modeling techniques based on 2D cameras are much more sensitive to illumination, shadows, scale and pose. However, it ususally takes range sensors a relatively long time of data acquisition, e.g., Vivid 9i takes 30 seconds for one 3D scan. This causes data distortion and cannot be ignored especially in the case of continuously gathering data on a moving platform with relatively fast speed.

The way of continuously obtaining range data is obviously more efficient and practical in actual applications. This thesis challenges the problem of 3D shape reconstruction by dynamic sensing using a range sensor.

1.2 Related Work

1.2.1 3D Reconstruction from Range data

For example, to digitize large-scale culture heritage, [5] has used laser range sensor mounted on flying balloon to cover the whole site. The uncontrollable movement of balloon causes distortions on the data. Utilizing the overlapped areas with another range sensor fixed on the ground, Banno and Ikeuchi adopt polynomial fitting to approximate the sensor motion parameters, then align and rectify distorted data.

Another typical issue coming from robotics field is to build a 3D map of complex non-flat terrain. Although range sensors can afford accurate 3D point clouds of the environment, relatively slow data acquisition rate compared with fast vehicle speed leads to distorted data and difficulties for alignment. Existing common solutions to this problem are either taking ″stop-scan-go″ strategy [29] [50] to avoid distortion or correcting the sensor motion using secondary sensors [26] [2] like GPS, inertial sensor, camera or even another range sensor. To achieve more practical and efficient system, Bosse and Zlot [7] proposed a 3D scan-matching method varied from the ICP algorithm, in which sensor trajectory is discretized and constraint are linearized to deal with continuously collected data from a moving vehicle.

Harrison and Newman [23] have made effort to utilize the vertical plane feature in man-made environment to refer low amplitude roll, pitch and yaw movements of vehicle moving on uneven ground.

Structured light range sensors with rolling shutter, like Microsoft Kinect, have the similar data distortion problem when it is used on a moving platform. In [48], Ringaby and Forssen proposed a scan rectification method by estimating 3D camera trajectory. Taking advantages from the NIR images and the limited range of the sensor, their problem is easier than the one which uses range data as the single source of data.

Another related line of work is rectification of rolling shutter video. This problem has been studied, and solved to some extent [3, 17]. What is different here is that in range sensor systems we have access to depth values in all pixels, and these allow us to robustly solve for the full 3D sensor trajectory, instead of resorting to affine motion [3], or rotation only models [17].

1.2.2 3D Shape Features

Finding correspondence between 3D shapes is very important and fundamental task in the field of shape analysis. A lot works have been done on this problem. Classical rigid alignment includes methods based on sampling and verifying candidate transformation, or by applying the iterated closest point algorithm or its variations [69, 49]. More recent works focus on shapes with large variations. There are some successful works about non-rigid registration of surfaces [8, 40, 67, 18], which usually consider the whole shape as a single feature and can not handle with large missing of data. Several recent works also deal with matching approximately isometric shapes [32, 62].

In the category of using shape features to find correspondence, representative points are selected from the shapes and shape descriptors are calculated to construct correspondence by comparing similarity between the descriptors. 3D shape features can be categorized into global feature and local feature based methods [58]. Global features characterize the whole 3D shape into a single feature. Local features work from the neighborhood of points on the surface and characterize each small local patch into a feature.

Among the global shape features: Ohbuchi et al. [41] investigate shape histograms that are discretely parameterized along the principal axes of inertia of the model. Osada et al. [42] introduce and compare shape distributions, which measure properties based on distance, angle, area and volume measurements between random surface points. Taubin [59] proposed Implicit Polynomial feature that first fits the shape with implicit polynomial fitting, and then extract rotation invariant feature from polynomial coefficients.

While global feature is both efficient in the aspects of time and storage; and is generally used in shape retrieving, it has obviously drawbacks that the feature cannot stay invariant on partial shape. Local feature techniques work from the neighborhood of points on the surface and characterize each small local shape patch into a feature. There are many tasks and applications based on the partial shape found the range finder, and thus, there are more local shape feature techniques than global feature techniques.

Among the local shape feature: Gelfand et al. [19] use a small 3D ball which centers at a given point on the shape in order to find the intersection space that insides both the 3D ball and the 3D shape; and then calculate the integration of the volume of the space. Johnson and Hebert studies the Spin Image [27] that project the local shape patch on to a spinning small image window and then characterize the small image window. Zaharescu et al. [65] extend the popular 2D image feature 2D DOG (Difference of Gaussian) on to 3D mesh shape. Sun et al [56] start from thermal physics that considers the 3D shape as an heat conductor and propose the Heat Kernel Signature which is a local shape feature invariant to non-rigid deformations.

All the methods mentioned above are considering 3D object as rigid. There are also many works on non-rigid invariant shape feature [45, 54, 66, 63, 67, 62, 51, 68, 56],

In the case of deformation coming from relative motion between range sensor and objects, it has its own properties. It usually includes large amount of missing data, due to occlusion or change of scene or view points. The shape might be deformed significantly between scans and the deformation doesn't keep the geodesic distance invariant. Thus we consider local shape feature is better choice for this problem.

Although we focus on 3D data in this research, there are plenty of excellent works in 2D image feature which can also give us inspirations [34, 37, 37, 60, 28].

- MSER detector
  Maximally Stable Extremal Regions (MSER) have been proposed by Matas [36]. A Maximally Stable Extremal Region is a connected component of an appropriately thresholded image. The word ″extremal″ refers to the property that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary. The ″maximally stable″ in MSER describes the property optimized in the threshold selection process.

- SIFT descriptor
  Scale Invariant Feature Transform (SIFT), proposed by Lowe [35], is a gold standard for evaluation of feature descriptor. It combines a scale-invariant region detector (DoG detector) and a descriptor (SIFT descriptor) based on the gradient distribution in the detected regions. SIFT descriptor can be applied alone on other kinds of feature regions for image matching. The descriptor is represented by a 3D histogram of gradient locations and orientations.

  Before computing the feature vector, the orientation of the feature region, which is covariant to the image rotation, should be assigned. An orientation histogram is formed from the gradient orientations of sample points within a region around the key point. The orientation histogram has 36 bins covering the 360 degree range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window. The magnitude, $m(x, y)$, and orientation, $\theta(x, y)$, are precomputed using pixel differences:

$$m(x, y) = ((L(x+1, y)-L(x-1, y))^2+(L(x, y+1)-L(x, y-1))^2)^{1/2} \quad (1.1)$$

$$\theta(x,y) = tan^{-1}((L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1,y))) \quad (1.2)$$

The several highest bins (higher than 80% of the highest bin) in the histogram are selected as orientation, to guarantee the stability of matching. Then the normalized feature region is rotated by the degree of orientation.

SIFT descriptor is a 3D histogram of gradient location and orientation, where location is quantized into a 4×4 location grid and the gradient angle is quantized into eight orientations. The resulting feature vector is of dimension 128. To obtain illumination invariance, the descriptor is normalized by the square root of the sum of squared components.

1.3 Summary of Contributions

This thesis challenges the problem of 3D reconstruction under a moving sensing system only with a single range sensor. Different from ″stop-scan-go″ strategy, our system works in an efficiently continuous manner. We don't have to stop the platform to obtain an stationary scan. Sensor can move under a reasonable motion mode and simultaneously scan the target or environment. Data distortion caused by continuous movement will be rectified. This continuous manner is much more efficient and attractive in practical applications.

- We propose a method which utilizes the prior shape model to estimate sensor motion. Based on the estimated motion model, distorted measurements can be rectified and accurate localization of targets can be achieved. As an application, we build a real-time pile driver positioning system using laser range finder.

- We propose a polynomial fitting method based on corresponding points to estimate 6 DOF motion parameters of a moving range sensor. The 3D model can be reconstructed according to a highly accurate sensor motion estimation. When using a 2D scanning sensor instead of a 1D scanning sensor, the same region is measured in multiple times when the sensor moves. We will show that we can reconstruct the sensor motion and the scene from only the measured coordinates and times of the same set of points. This reconstruction is intrinsic, which relies on only the intrinsic properties of the distortion, and not relying on the extrinsic information from other sensors. Firstly, for robust estimating sensor movement, we model the sensor motion using polynomial with respect to time. Secondly, to estimate the parameters of sensor motion model, we utilize the 3D corresponding points extracted from the overlapped parts between consecutive frames. This method doesn't need the secondary sensor and is not limited with specific environment features. Without linearization of constraint and discretization of trajectory, distorted data is accurately rectified.

- To obtain corresponding points in distorted data , we propose a novel 3D affine invariant feature detection and matching method which is designed for the deformed 3D data collected by moving range sensor. Firstly a Morse function which measures the object shape is designed. A level-set based method is proposed to extract stable affine invariant region feature in multiple scale. Applied a moment-based affine transformation, the deformed region feature can be normalized. A multiple scale Spin Image is designed to describe the normalized region feature.

1.4 Overview of the Thesis

Chapter 2 presents the first contribution of this thesis, a real-time pile driver positioning system.

Chapter 3 presents the second contribution of this thesis, 3D shape reconstruction using a moving range sensor based on estimation of sensor motion.

Chapter 4 presents the third contribution of this thesis, a 3D affine invariant shape feature designed for distorted data from moving range sensor.

Chapter 5 summarizes the thesis, clarifies the contributions and discusses a few directions for future research.

Chapter 2

Dynamic Sensing with Known Targets

– A Real-time Pile Driver Positioning System using Laser Range Finder

2.1 Introduction

For the known targets, since there is no need to build 3D model for the object, our interesting mainly focus on how to accurate localize them. We propose a method which utilizes the prior shape model to estimate sensor motion. Based on the estimated motion model, distorted measurements can be rectified and accurate localization of targets can be achieved.

Assume targets are rigid moving objects. Assume that range sensor is static to the absolute reference frame. Let $\mathcal{S}$ be a implicit function in object coordinate system, describing the surface of the target. It means that for any point p on the object surface, there is:

$$\mathcal{S}(p) = 0 \tag{2.1}$$

Let $G(t)$ be a general smooth function describing the motion of the target relative to the sensor. Given a set of range sensor measurement points $\{x\}$ of the target under the senor reference frame, with measuring time t of each point, its corresponding object coordinates p satisfy:

$$p(t) = G(t) \cdot x \tag{2.2}$$

Since $p(t)$ are points on the target surface, they should satisfy Eq. 2.1. We thus can build a general cost function as the sum of square of errors of points satisfying Eq. 2.1:

$$\sum_{i=1}^{K} ||\mathcal{S}(G(t_i) \cdot x_i)||^2 \tag{2.3}$$

where K is the number of points.

We can deduce the target motion $G(t)$ by minimizing the cost function Eq. 2.3. How to choose the parameterization of $G(t)$ depends on the task at hand. It could differ from the simple linear modeling to more complex polynomial approximation or spline curves.

As an application, we build a real-time pile driver position system using laser range finder. Details are in following sections.

2.2 Application Background

Pile driver is a mechanical device used in the construction field. It drives piles into soil to provide foundation support for buildings or other structures. In the process of pile driving, a marker is first placed at the designed pile point. A hole is then dug at this marked point and a pile is driven in to it.

The problem during this procedure is that the marker couldn't be seen after starting drilling the hole. The pile needs to be kept at the desired position manually. This manual way requires well trained operators and long operation time and may degrade the accuracy. The cost is high and the procedure is not safe for workers. This is why a well designed automatic positioning system is needed.

The problem is caused by the non-real-time property of current surveying equipments. In this paper, inspired by positioning techniques in the robotic field, we propose a novel pile driver positioning system, as shown in Fig.2.1. Instead of measuring preset marker on ground in pre-process work, the proposed system realizes real-time accurate tracking of pile position, which can be used for navigation of pile driver. Compared to the manual way, the pre-process work and additional workers during the process can be saved and it becomes easier to evaluate a construction error.

On one hand, in modern construction field, there are two main surveying equipments: total station and global satellite positioning system. Total station is the most widely used one. It can provide extremely precise measurement of a single point as long as a direct line of sight can be established between the point and the equipment. Its surveying procedure usually requires another worker to hold the prism reflector. The other surveying tool, global satellite positioning system, like GPS, is not limited to the line of sight observation, but not stable at the areas where high buildings or mountains block the signal. It needs a relatively long time to initialize in order to get an reliable and stable measurement.

On the other hand, in the research field of robotics, position measurement of specified objects such as people or mobile robots is also an important task. Many real-time positioning systems using distributed devices including cameras [21], [55], [38], Laser Range Finders (LRF) [1], [14] and ultrasound sensors [33], [9] have been proposed. For applications in the construction field, LRFs can scan a plane and get multiple measurements in a very short time interval, while total staion or GPS can only get one measurement with a long operation time. LRFs have already been used for real-time position measurement in large outdoor areas [70]. It is considered that a position measurement system using LRFs can also be applied to survey tasks at construction sites [31], [13], [57], [53]. Unlike ultrasound sensor, LRF is easy to deploy in the construction site. Compared to camera, LRF is more stable and accurate even under poor illumination

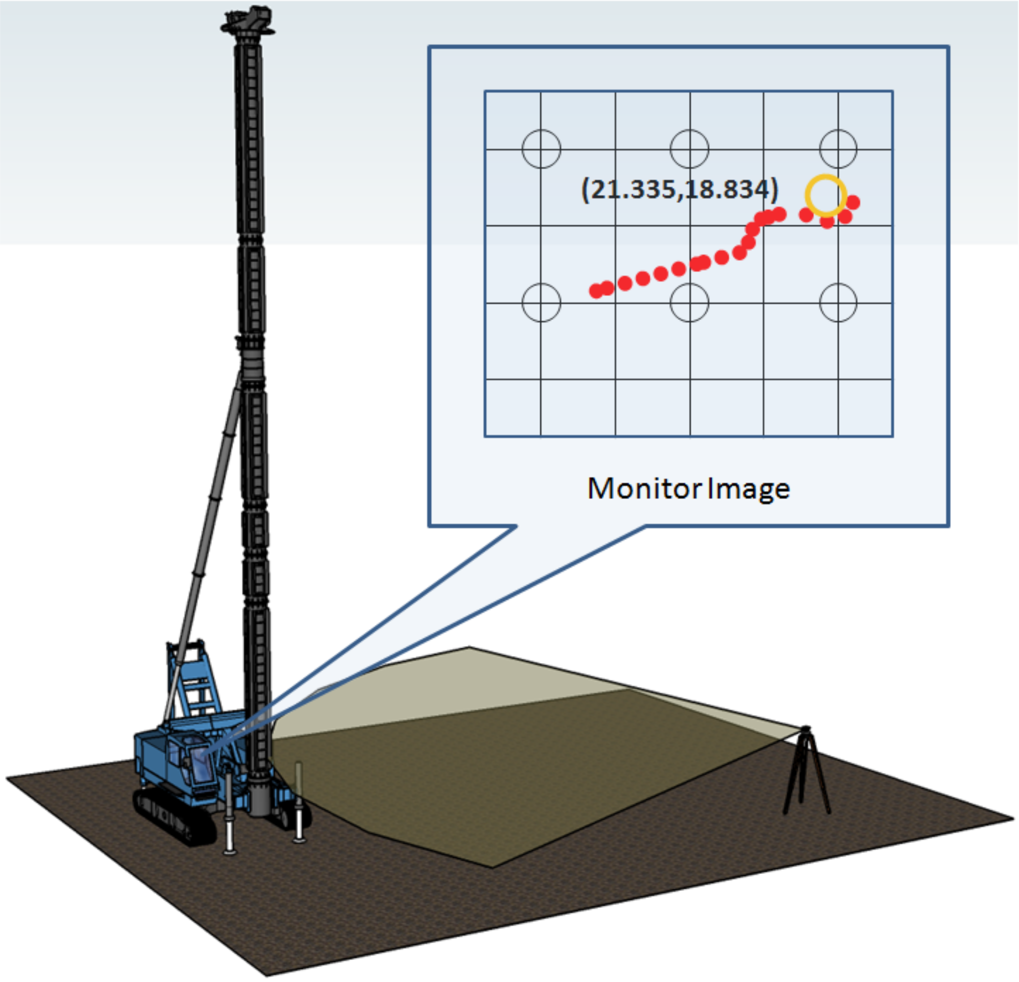


Figure 2.1: System Overview: LRF scans the construction workspace, estimates the position of the pile and shows it on the display terminal to help the worker operate the pile driver. The red points are measured data points of the pile driver and orange circle stands for the estimated pile position. The dark circles stand for the designed pile positions.

condition. Camera-based positioning system achieves high accuracy only under a small scale and well controlled industry environment, since it suffers from the complexity of scene and change of illumination. This makes camera-based positioning systems less reliable in an unstructured construction site. We consider LRF as the most suitable sensor, for its advantages such as real-time, relatively high accuracy, large covering area, low noise and simple installation [24], [25].

As shown in Fig.2.2, our proposed pile driver positioning system consists of three parts: data acquisition from LRF, position estimation process and visualization of results. It first gathers real-time 2D LRF data of scene. While waiting for the next frame of data from LRF, pile position is estimated based on the obtained data. The estimated pile position is shown in the map together with the designed one to assist workers with the pile driving operation. The whole procedure is completed in very short time period, which is mainly determined by data acquisition rate of LRF and CPU speed. For this reason, the proposed system can be used in real-time.

In the position estimation process, target detection is a very important step. As shown in Fig.2.1, in the typical application scene, besides cylindrical pile, there may also be other shaped objects near to the target in the observation range of the LRF. An effective target detection algorithm is required to identify the pile from other objects. LRF data of pile should be part of a circle since the data scanned from the LRF is the contour of object. Inspired by RANSAC and Hough Transform, we propose an algorithm of circle detection for positioning system.

After extracting the contour of cylindrical targets from raw data, we use the Maximum Likelihood Estimation (MLE) to fit the arc shaped contour to a circle with a given radius to estimate the circle center. The orientation-invariant property of the circle guarantees that from any view direction, the arc-shaped contour obtained from LRF always stands for the same center point of the reference target.

We give a description of the system configuration in Section **??**. The

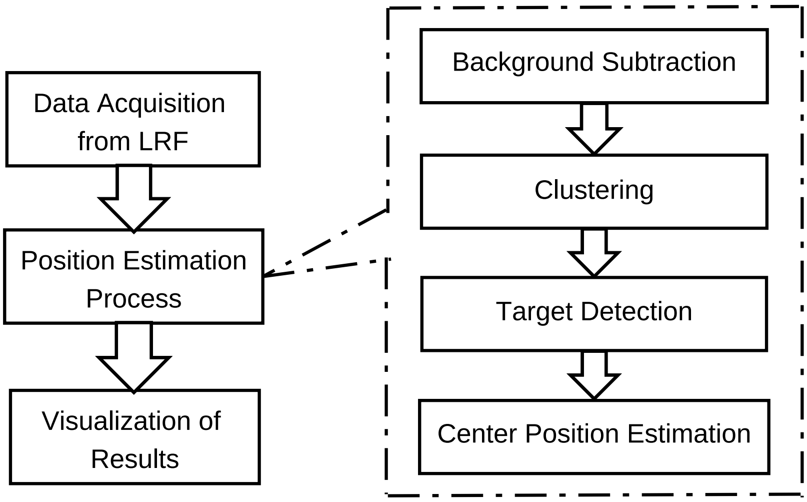

Data Acquisition from LRF
Position Estimation Process
Visualization of Results
Background Subtraction
Clustering
Target Detection
Center Position Estimation

Figure 2.2: Process flow graph

detailed algorithms of position estimation are in Section **??**. Section **??** are experiments on both simulation data and real data, followed by the conclusion in Section **??**.

2.3 Configuration of A Real-time Pile Positioning System

The configuration of the proposed system is shown in Fig.2.3. In the system, an LRF is set at a position higher than human height by using a tripod stand and kept horizontal by using a leveling system. The LRF is mounted on a pan unit (as shown in Fig.2.4), which is used to improve the angular resolution of the LRF. The data acquired from the LRF are sent to a processing computer in which the center positions of cylindrical piles are estimated. Then, the estimated center positions are sent to mobile display devices through a wireless network. Workers can easily get access to visualized information about the current measured positions and the designed positions.

In the data acquisition module, for the purpose of getting more data

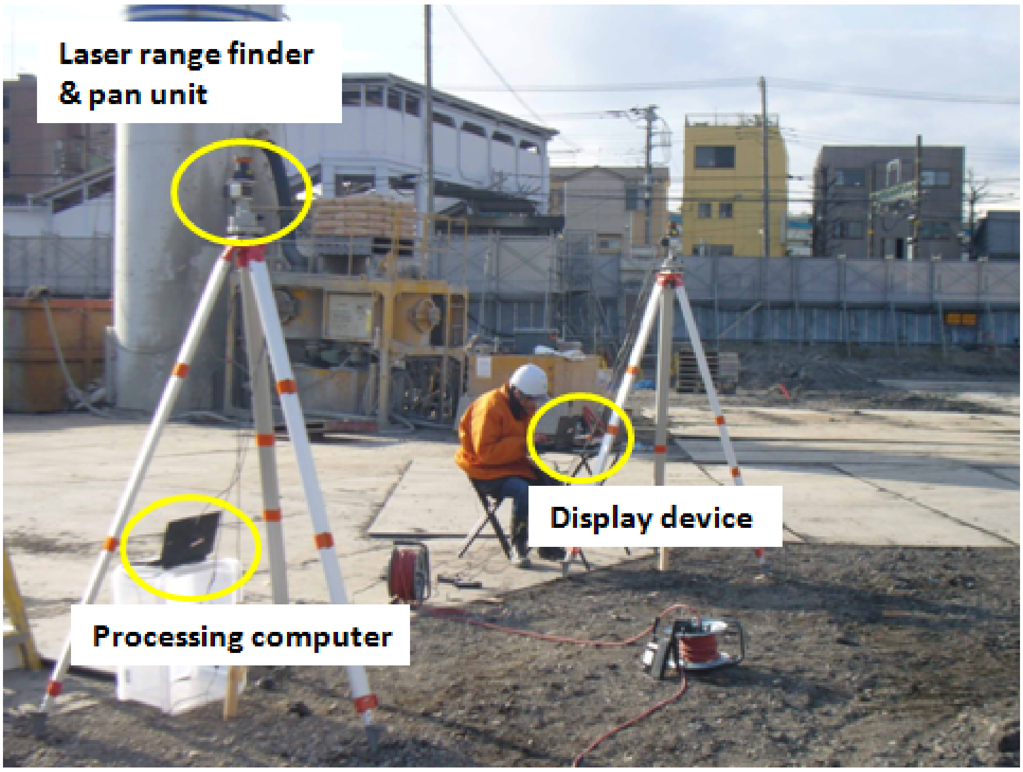


Figure 2.3: System configuration

Table 2.1: Specification of UTM-30LX

Model Number	UTM-30LX
Light source	Laser diode $\lambda = 870[nm]$
Measurable area	$0.1[m] \sim 30[m], 270[deg]$
Measurement accuracy	$0.1[m] \sim 10[m]$: $\pm 30[mm]$ $10[m] \sim 30[m]$: $\pm 50[mm]$
Angular resolution	$0.25[deg]$
Scan time	$25[msec]/scan$

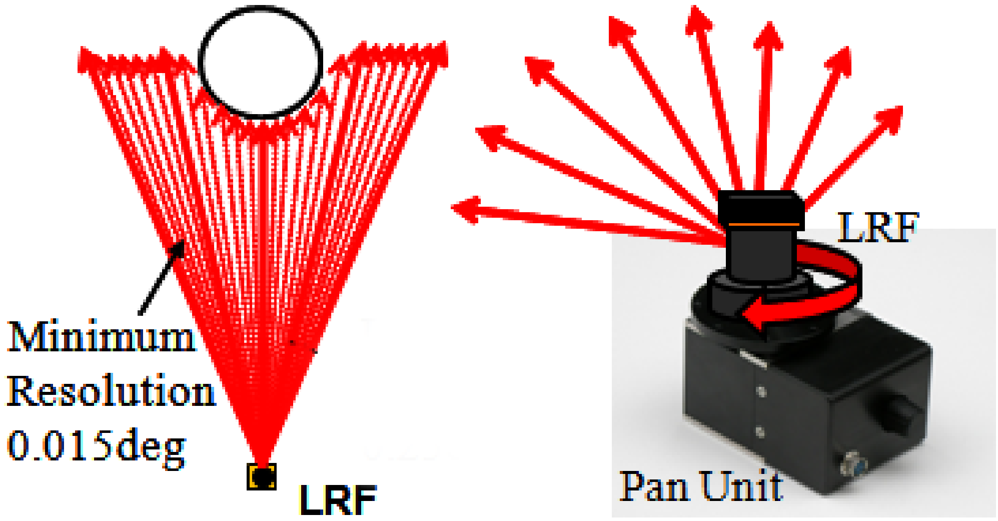


Figure 2.4: A combination of LRF and pan unit

points of the object to decrease the affect of noise, a pan unit is combined with the LRF to improve the angular resolution. As shown in Fig.2.4, LRF is rotated by a small angle after each scan. The relatively low angular resolution of LRF is improved to the same as pan unit with higher resolution. In our system, we use the type of UTM-30LX LRF with its specification described in Table 2.1. Its original angular resolution is 0.25 degrees. Here we mount the LRF on a SPU-01 pan unit which has a higher angular resolution of 0.015 degrees. We can get much denser data since the resolution can be improved 17 fold at most. The system works in a data acquisition frequency between $2Hz - 40Hz$. The working speed of pile driver is between $0 - 2 \times 10^3 mm/s$. The speed in the final stage of the piling procedure, which especially needs high positioning precision, is around $0 - 50mm/s$. For this slow movement, the system data acquisition frequency is satisfied for the real-time requirement of the application of assistance of pile driver.

2.4 Position Estimation Based on Target Shape Model

There are four main steps in the process of position estimation after obtaining raw data from LRF. The background of the scene is assumed to be static. Firstly moving parts of data are extracted as foreground objects using a background subtraction algorithm. Then all the foreground data points are clustered using a neighborhood verification method. The resulted data set is denoted as S. Section 2.4.1 describes how to detect circle target from S using the proposed algorithm, which is inspired by RANSAC and Hough transform. Section 2.4.2 is about how to estimated the circle center with a MLE algorithm using detected arc data.

2.4.1 Pile Detection

RANSAC [16] is the abbreviation of "random sample consensus". Its basic idea is that there should be a good subset which is composed of only inliers. The procedure of RANSAC is that a subset is randomly sampled from the data, a hypothesis of the model is established from the subset, the hypothesis is tested on all the data, and the number of data acceptable to the model within the predetermined threshold is counted. After iterations of this procedure, the model that fits to the maximum number of data is determined as the answer. The disadvantage is that it requires thresholds predetermined depending on the problem. In our application, because of the presence of large noise and objects with similar shape to target, RANSAC breaks down.

Hough transform [12] is an algorithm which makes a vote on all possible set of parameters for each data point and count the parameter set with the most votes as "winner". According to the requirement of estimation accuracy, it usually needs a huge number of accumulator units and is computational expensive. Its voting procedure helps to find the most possible result.

Inspired by RANSAC and the voting procedure of Hough transform, we propose the following circle detection algorithm, utilizing the prior knowledge of the circle radius. Fig.2.5 shows the flow chart of the algorithm.

We assume that the pile is perpendicular, and the cross section is modeled by a circle:

$$(x-a)^2 + (y-b)^2 = r^2. \tag{2.4}$$

The proposed algorithm works using the following steps:

Step 1. In a data cluster S, a random subset of data is chosen to calculate position a, b and radius r in Eq.2.4. Because there are three unknowns in Eq.2.4, three non-colinear points, can determine one set of parameters of a circle. So the number of points contained in the randomly chosen subset of data is determined to be three.

Step 2. As the radius of the pile R is given in advance, a threshold T_r is used to eliminate any invalid model. The current hypothesis will be rejected, if this inequality is not satisfied:

$$|R-r| \leq T_r. \tag{2.5}$$

The reference target used in this positioning system is cylinder, so from the viewpoint of LRF, the shape of the contour of target is convex, which means the center of object can not be observed by sensor. Utilizing this assumption, another constraint to get rid of false model can be described as Eq.2.6:

$$(k_i \cdot x_{LRF} + t_i - y_{LRF})(k_i \cdot x_{new} + t_i - y_{new}) < 0, \tag{2.6}$$

where l_i:$y = k_i \cdot x + t_i, i = 1, 2, 3$ are lines determined by two points from the current subset of data. (x_{LRF}, y_{LRF}) is the position of LRF and $P_{new} = (x_{new}, y_{new})$ is the currently calculated center position (a, b). Equation 2.6 means that the LRF and center of target should be on a different side of the line determined by points from the data subset. This constraint can ensure the obtained model is convex.

If constraints Eq.2.5 and 2.6 are satisfied at the same time, then the currently calculated center position P_{new} is considered as a center candidate,

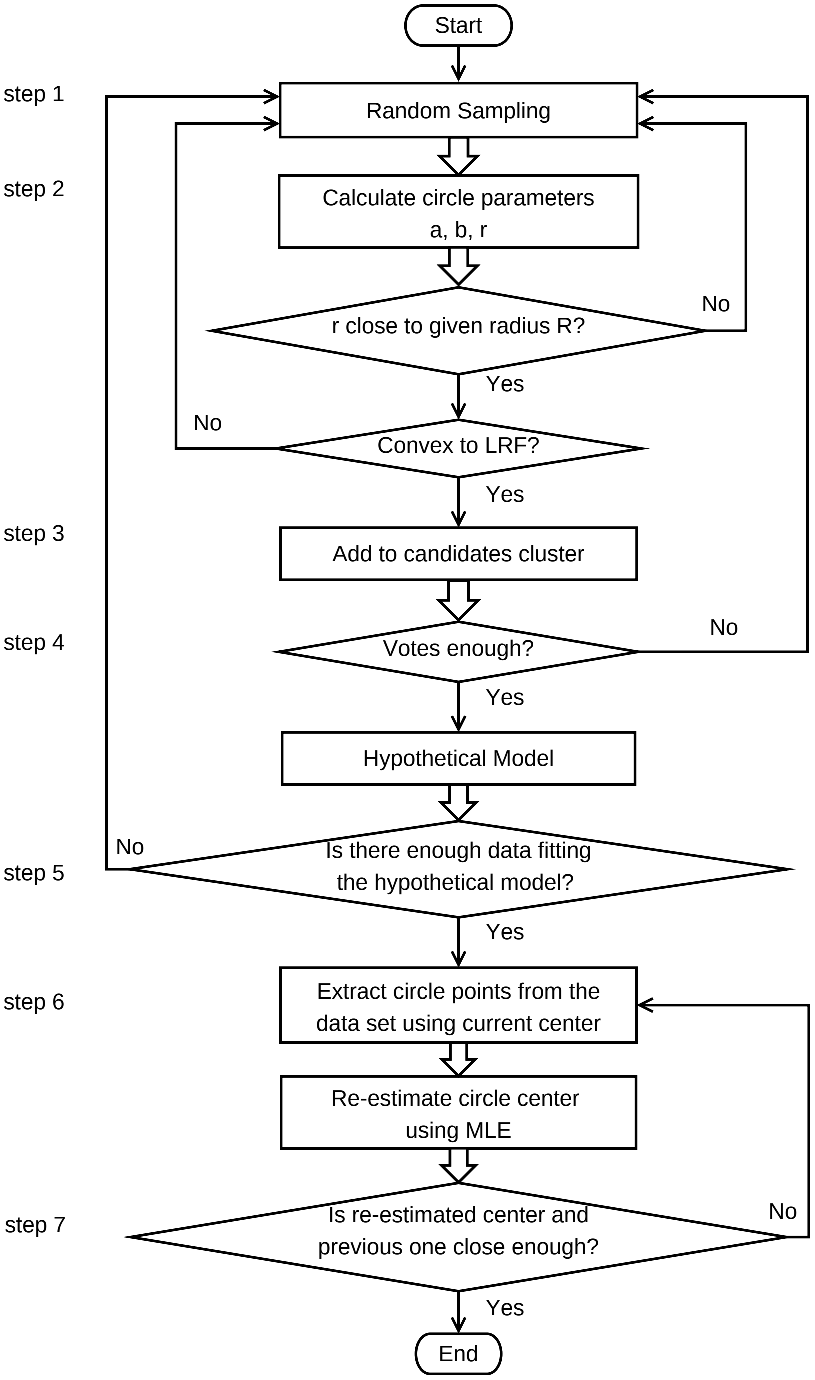

Start
step 1
Random Sampling
step 2
Calculate circle parameters
a, b, r
r close to given radius R?
No
Yes
No
Convex to LRF?
Yes
step 3
Add to candidates cluster
step 4
Votes enough?
No
Yes
Hypothetical Model
No
step 5
Is there enough data fitting
the hypothetical model?
Yes
step 6
Extract circle points from the
data set using current center
Re-estimate circle center
using MLE
step 7
Is re-estimated center and
previous one close enough?
No
Yes
End

Figure 2.5: Flow chart of the proposed positioning algorithm based on circle detection and fitting .

otherwise we just ignore this model and go back to Step 1.

Step 3. Add the new position candidate into candidates cluster by using the rule described below:

$$||P_{new} - P_{ij}|| \leq T_c, \tag{2.7}$$

where P_{ij} is arbitrary element of the i^{th} cluster, j is the index of the element in the i^{th} cluster, and T_c is the threshold to determine whether a position candidate belongs to a cluster or not.

If any element of the i^{th} cluster can satisfy Eq.2.7, then P_{new} is classified into the i^{th} candidate cluster.

Step 4. Each center candidate is considered as a vote to the candidate cluster. If the number of votes of current added cluster, N_v, has exceeded the threshold T_v, then this cluster is chosen to be a hypothetical model, as shown in Eq.2.8

$$N_v \geq T_v. \tag{2.8}$$

Here we define C_h, the center of the hypothetical model, by the average value of all centers in this cluster

$$C_h = \frac{1}{n}\sum_{j=1}^{n} P_{ij},$$

where n is the total number of votes in i^{th} cluster.

Step 5. All other points in this data set are tested by the hypothetical model determined by C_h and the given radius R. If a point fits the model well enough, which means Eq.2.9 is satisfied, it is considered as a hypothetical inlier.

$$\Big| ||C_h - P_i|| - R \Big| \leq T_d, \ P_i \in S. \tag{2.9}$$

If N_i, the number of points evaluated as inliers, is larger than the given threshold T_n, as described in Eq.2.10

$$N_i \geq T_n, \tag{2.10}$$

then this model is used as an initial model of the circle fitting by MLE algorithm, otherwise go back to step 1.

Step 6. All hypothetical inliers are used to re-estimate the model by MLE algorithm.

Step 7. After obtaining the estimated center position from MLE algorithm, the refined result C_r is compared to previous one C_p by using Eq.2.11

$$||C_p - C_r|| \leq T_i. \tag{2.11}$$

If distance between C_r and C_p is less than a given threshold T_i, then the whole algorithm is finished, otherwise the refined center position by MLE is used as a new center hypothesis C_h and go back to step 6.

2.4.2 Refinement for Accurate Position Estimation

An algorithm of circle fitting based on MLE is designed, after extracting the data points of circle from the total data set,.

The MLE for a circle is a method for deciding the parameters a, b, and r so that the observed data can most easily be obtained from the assumed noise model [30]. In other words, the MLE is a method for estimating the parameters which can maximize the likelihood of each data point $(x_i, y_i), i = 1, \ldots, K$. In this paper, we assume that each data point obtained from the LRF has an independent error described by a Gaussian distribution with mean 0 and standard deviation σ. The likelihood of each data point $(x_i, y_i), i = 1, \ldots, K$ can be expressed as

$$\begin{aligned} & p(x_1, \ldots, x_K, y_1, \ldots, y_K) \\ = & \prod_{i=1}^{K} \frac{\exp\left[-(x_i - \bar{x}_i)^2/2\sigma^2\right]}{\sqrt{2\pi\sigma^2}} \times \frac{\exp\left[-(y_i - \bar{y}_i)^2/2\sigma^2\right]}{\sqrt{2\pi\sigma^2}} \\ = & \frac{\exp\left[-\sum_{i=1}^{K}\left[(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2\right]/2\sigma^2\right]}{\sqrt{(2\pi\sigma^2)^{2K}}}, \end{aligned} \tag{2.12}$$

where $(\bar{x}_i, \bar{y}_i)$ is the true position of (x_i, y_i). To maximize this likelihood $p(x_1, \ldots, x_K, y_1, \ldots, y_K)$, we minimize $-\log[p(x_1, \ldots, x_K, y_1, \ldots, y_K)]$. We first take a logarithm of both sides of Eq.2.12 and then remove a constant term which does not contribute to minimization.

Here we have to fulfill Eq.2.4 as a constraint of x_i and y_i. We remove the restraint condition by using Lagrange's method of undetermined multipliers. Finally the MLE for a circle is equal to estimate parameters a, b, and r which will minimize J_{ML} expressed as Eq.2.13:

$$J_{ML} = \sum_{i=1}^{K} \frac{\left(x_i^2 + y_i^2 - 2ax_i - 2by_i + a^2 + b^2 - r^2\right)^2}{x_i^2 + y_i^2 - 2ax_i - 2by_i + a^2 + b^2}. \tag{2.13}$$

In our system, parameters which should be estimated are only a and b, since the radius of the reference target is given in advance. In this case, however, the MLE becomes a non-linear problem. To solve this non-linear equation, we apply the Newton-Raphson method since it is known to have a faster convergence rate than other Gradient methods, such as Conjugate gradient or Levenberg-Marquardt, if its initial value is close to the true value of a and b [57]. Here we use the detected center position from the previous detection result as the initial value of the Newton-Raphson method.

2.5 Experimental Results

To verify our proposed circle detection and fitting algorithm, we first test it by simulation experiments. Before the experiment at actual construction site, we conduct an indoor experiment with a small scale model of pile driver to test the proposed positioning system. At last, an outdoor experiment at actual construction site is performed.

2.5.1 Simulation experiment

Simulation setting

There are two basic simulated object shapes: circle and line. Considerable scenarios can be simulated by combining these two types. Here we assume that with the cylindrical reference target there are other objects

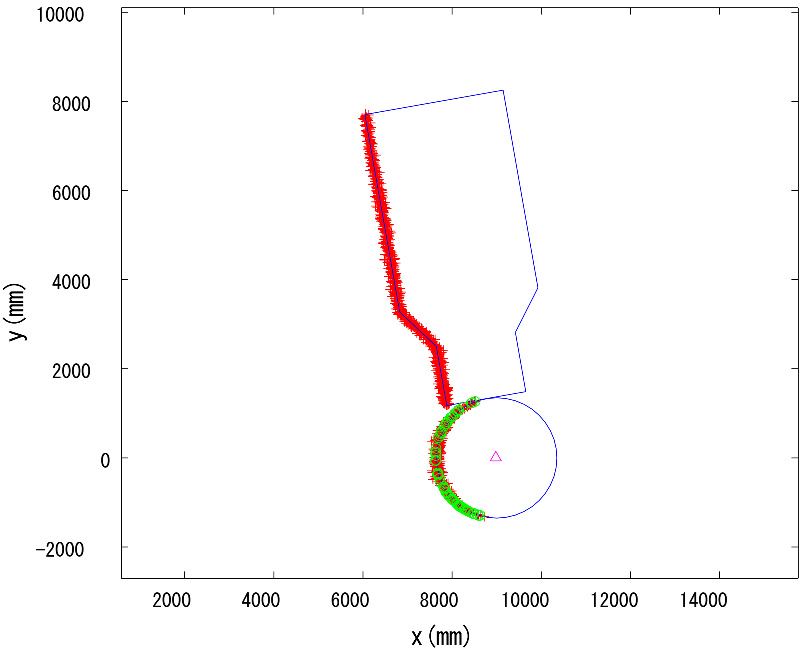


Figure 2.6: Simulation result. The blue lines stand for the actual contour of the objects. Red crosses stand for the simulated scanned data from LRF. Green circles stand for points which are recognized as circle by the proposed algorithm. The purple triangle stands for the estimated center of the detected reference target.

with cross-sections shaped like rectangle and trapezoid in the scenario whose contour is similar to arc especially with large sensor noise, as shown in Fig. 2.6.

In this simulation, we try to simulate the scene and devices of previous experiment at an actual construction site [22]. The noise of LRF data is assumed to be independent Gaussian noise with the standard deviation $\sigma = 50mm$ in the range of distance $d_{max} = 30 \times 10^3 mm$. The angle resolution of LRF is set to be $\theta_{reso} = 0.05°$. The radius of cylindrical reference target is set to be $R = 1350mm$. The distance from LRF to the center of cylinder is set to be $d = 9000mm$. Values of parameters T_r, T_c and T_d are empirically determined as σ. T_i is empirically determined as 0.3σ.

To find out the optimal number of votes, we test the algorithm by changing the parameter T_v from 1 to 10. 1000 times simulations are tested for each setting of parameters. All results are average value of 1000 times simulations. To make the system work in real time and always process the latest data, the maximum number of iterations of random sampling is set to be 10,000. If the algorithm cannot establish an acceptable model until maximum iteration time, it will abort and be considered as a failed detection.

Evaluation of the circle detection algorithm

The simulation result is shown in Fig.2.6. To evaluate the algorithm, detection rate, false detection rate, iteration times of random sampling and error of estimated center are used here.

There are three possible outcomes in the simulation result: (a) correct detection. (b) false detection, which is a wrong result but not rejected by detection algorithm. Here we define that a wrong result is $E > \sigma$. (c) failed detection, which is no result established in maximum iteration time.

Detection rate is defined as: $\frac{M_a}{M}$, and false detection rate is defined as: $\frac{M_b}{M}$, Where M is the total number of simulation, M_a is the number of correct detection and M_b is the number of false detection.

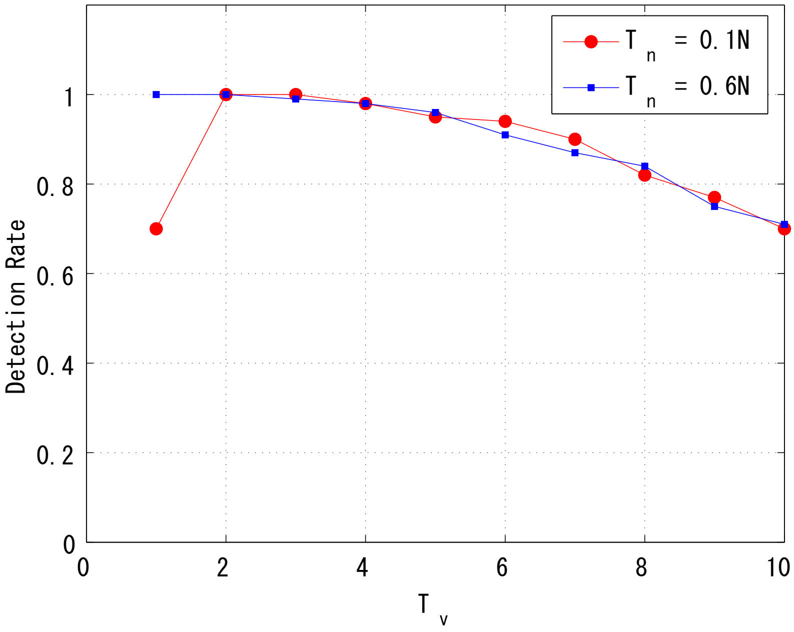


Figure 2.7: Detection rate

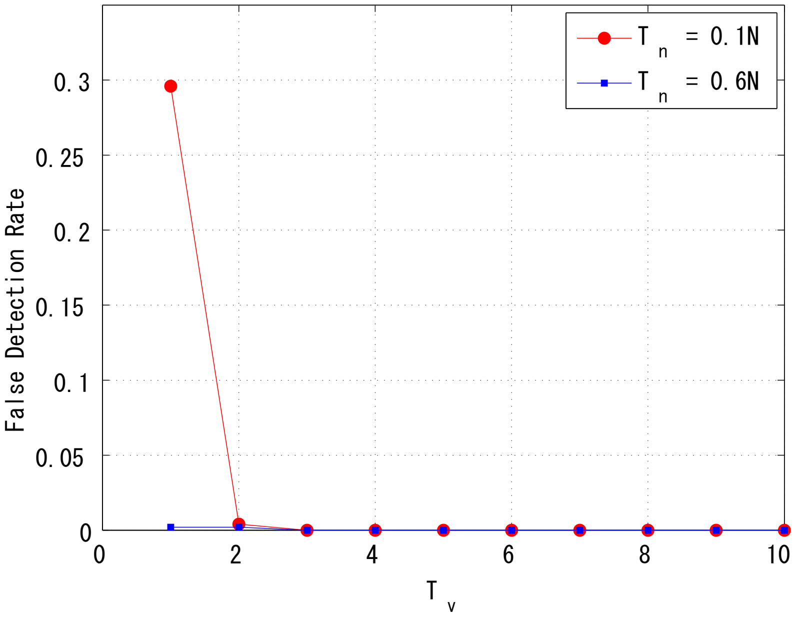


Figure 2.8: False detection rate

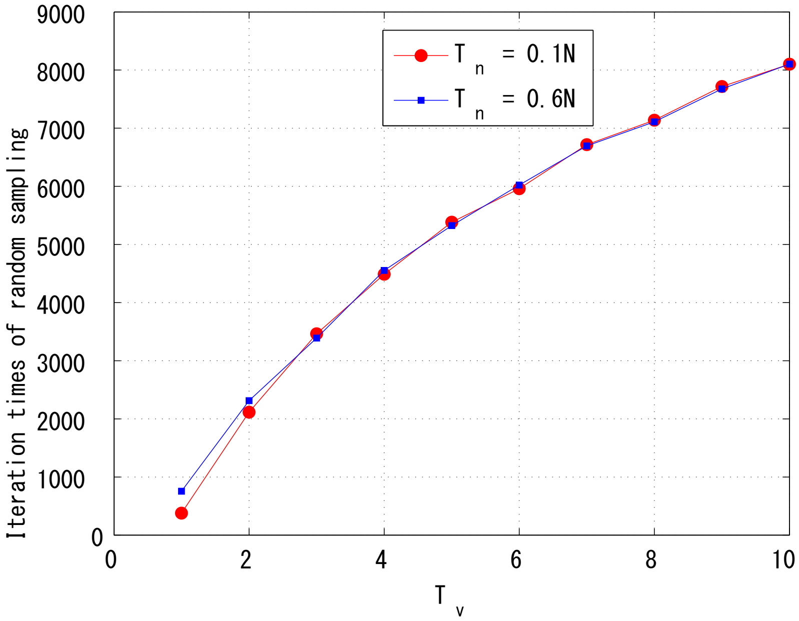


Figure 2.9: Iteration times of random sampling

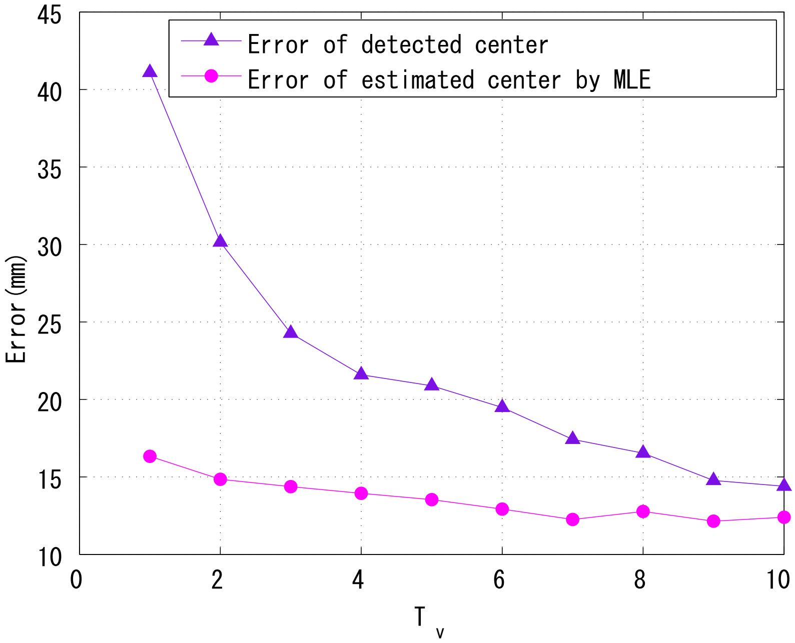


Figure 2.10: Error of estimated position and detected position ($T_n = 0.1N$)

Values of detection rate and false detection rate with parameter T_v changing from 1 to 10 are shown in Fig.2.7 and Fig.2.8. When parameter T_n is set to be small, the figures show that false positive rate is almost zero if the needed votes T_v is set larger than one.

Results with $T_n = 0.1N$ and $T_n = 0.6N$ are compared in figures, where N is the roughly estimated number of inliers as shown in:

$$N = \frac{2}{\theta_{reso}} \cdot (\arctan \frac{R}{D}),$$

where D is the distance from LRS to the center of the data cluster and θ_{reso} is the angular resolution of the positioning system. Parameter T_n is the number of data points required to claim that the hypothetical model can be accepted. It means that the voting procedure makes the proposed algorithm still work well even when occlusion or significant noise degrades the data. Here $T_n = 0.6N$ and $T_n = 0.1N$ are experimentally verified upper and lower bound.

Fig.2.7 shows that when votes are 2 and 3 the detection rate is almost 100%. But the value decreases when more votes are required. The main reason is the maximum iteration time limits the performance.

The average iteration times of random sampling increase while the number of votes becomes larger, as shown in Fig.2.9. The parameter T_n almost has no effect on the iteration time when $T_v > 1$. Fig.2.9 shows that choosing a small T_n would not increase the computation cost.

Position estimation precision

We evaluate estimation errors of the aforementioned two methods: the non-linear MLE and the proposed detection algorithm with known radius. Estimation error E is defined as:

$$E = \sqrt{(X_L - a)^2 + (Y_L - b)^2},$$

where (a, b) is the estimated center position of the cylindrical reference bar and (X_L, Y_L) is a position where we put the cylindrical reference bar.

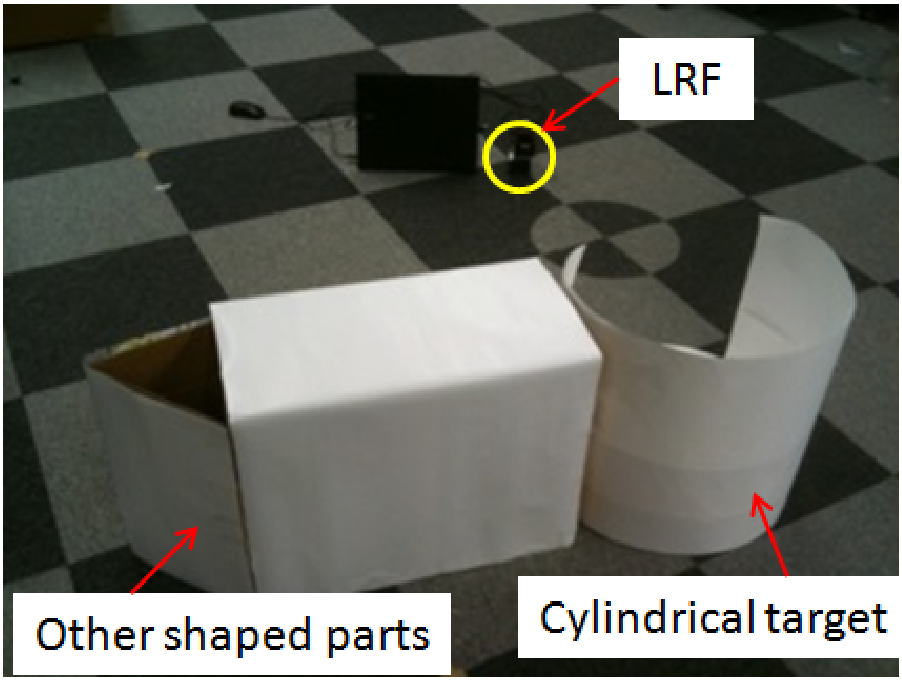


Figure 2.11: Indoor experiment scene

Fig.2.10 gives us the error of estimated center position using MLE and the error of detected center position, which is the result of Step 5. It shows that the number of votes T_v doesn't have a significant effect on the estimation result while the number of iterations increases dramatically with more votes. So a small T_v would be a good trade off between estimation accuracy and computation cost.

Considering detection rate, iteration time of random sampling, error of estimated center position and the flexibility of algorithm, $T_v = 2$ is the best choice of votes under the given conditions.

2.5.2 Indoor experiment

Indoor experiment setup

To verify the feasibility and estimation error of our proposed system, before implementing it in the actual construction field, we made the indoor experiment. As shown in Fig.2.11, the mesh on ground with the size of $500mm \times 500mm$, is used as prior given world coordinates. A cylinder with radius of $250mm$ is used as the reference bar. Two cardboard boxes, one with the rectangle shaped cross-section and one triangle shaped cross-section, are used to simulate the pile driver in the application scene. Fig.2.12 shows an example of the scanned data from LRF of the experiment scene.

To verify the feasibility of proposed circle detection algorithm, we firstly put the reference bar at a known position and estimate the center position only using MLE method. Then while keeping the cylinder at the same place, we put the other shaped objects near to it and estimate the position of cylinder again with the proposed positioning algorithm based on circle detection and fitting, to see if we can achieve the same estimation accuracy or not. Experiment with this set up is repeated at every $500mm$ from the distance $1000mm$ to $7500mm$.

Indoor experiment result

The experiment result proved that our proposed circle detection algorithm can effectively extract data points of circle from the scene with other shaped objects. The two curves of estimation error are almost the same, with the maximum difference of $3.4mm$, as shown in Fig.2.13.

To find out the most appropriate number of needed votes, we vary the parameter T_v from 1 to 10 to see the effect on detection rate. When votes are $T_v \geq 3$, as Fig.2.14 shows, the detection rate exceeds 90% even with small value of T_n. If there is no voting procedure, meaning $T_v = 1$, the detection rate is barely acceptable. This low detection rate is caused by the junction of triangle and rectangle boxes, whose contour is similar to an arc. This

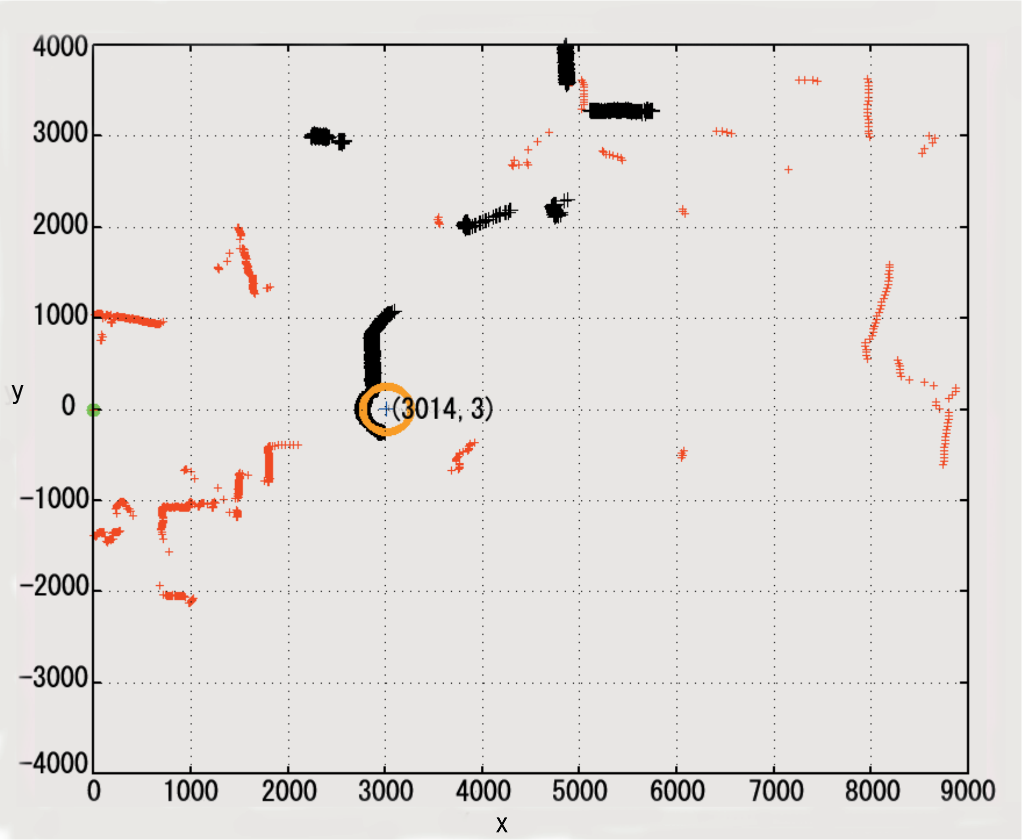


Figure 2.12: An example of measuring the center position of the cylindrical reference bar. The units of x and y are in [mm], and an LRF is set at the origin indicated by the green point. Red points represent background data, black represents the contour of foreground objects, orange represents the estimated circle, and blue cross represents the measured center position.

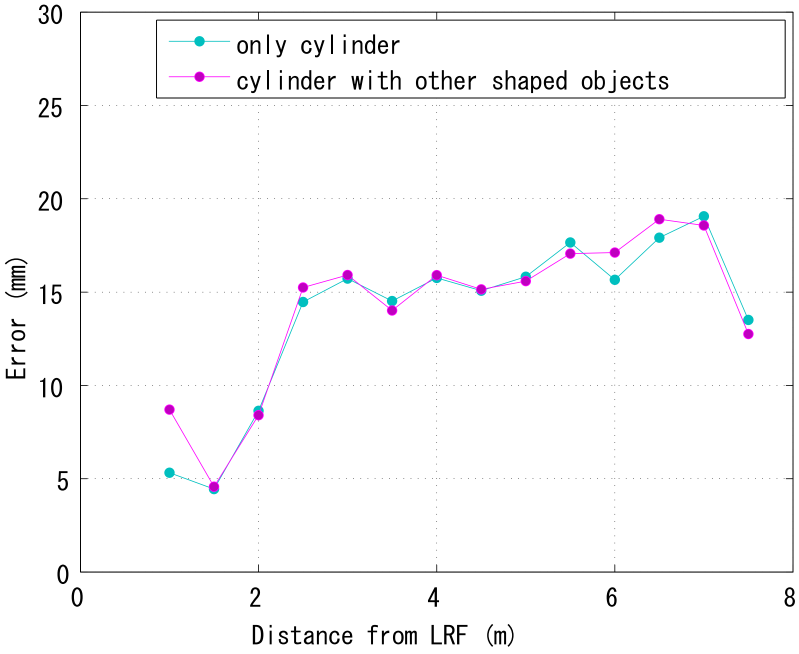


Figure 2.13: Estimation error of MLE and the proposed positioning algorithm of indoor experiment.

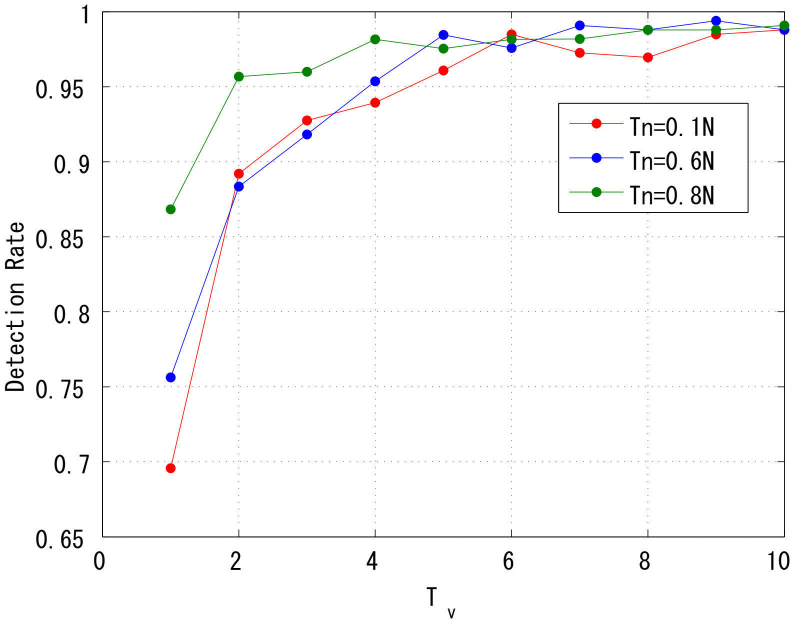


Figure 2.14: Detection Rate of indoor experiment

result proves the feasibility of the proposed circle detection algorithm and shows the possibility that it can still work well with occlusion or large noise present.

2.5.3 Experiment in the construction field

We tested the proposed system in an actual construction field. Our purpose is to measure the position of the pile which is being put into the drill hole. Tracking the position of pile can help to position it at the expected place, which requires an accuracy of 100*mm*.

The laser range finder is assembled with a pan unit, which is used to improve the angle resolution of LRF, as shown in Fig.2.4. LRF and pan unit are placed on the tripod which can be adjusted to keep the scan plane of LRF horizontal. The height of tripod also helps to keep other moving objects, like humans, from the scan range. The object to be measured is the pile shown in Fig.2.15, with a radius of 200*mm*. Limited by the arrangement of construction field, the distance between the measured object and LRF is about $15 \times 10^3 mm$.

Calibration of LRF

To calibrate the LRF, a thin metal stick (cross section: $30mm \times 1mm$) which has a highly reflectable surface is used. We first use total station to position several points with a precision around 2mm/km. We then put the calibration stick at those points and obtain the scan data of stick from LRF. We use an average value of scan data of stick to estimate its position. Transformation matrix between the world coordinate and the LRF coordinate can be calculated using least square estimation method.

Procedure of pile driving

As shown in Fig.2.15, the pile was lifted by the pile driver and being slowly put down into the hole. Currently in most construction fields, the pile is kept to the expected position by manual work. It needs three well trained workers using sticks to measure whether the pile is at the right place or not.

Result of the construction field experiment

The procedure of pile driving was recorded by the LRF. It started when the pile was moved near to the expected position about $1m$. After arriving the top of the hole, the pile was kept going down at a position with error around $50mm$. The placement of pile was adjusted to be more accurate at the final 10sec. The position adjustment by workers can be seen in Fig.2.16. Here error at X^L is defined as $(X_L - a)$ and error at Y^L as $(Y_L - b)$, where (a, b) is the designed position of pile. The errors between the estimated center position and the expected design position of pile are shown in Fig.2.17. The final position error given by proposed system is around $25mm$.

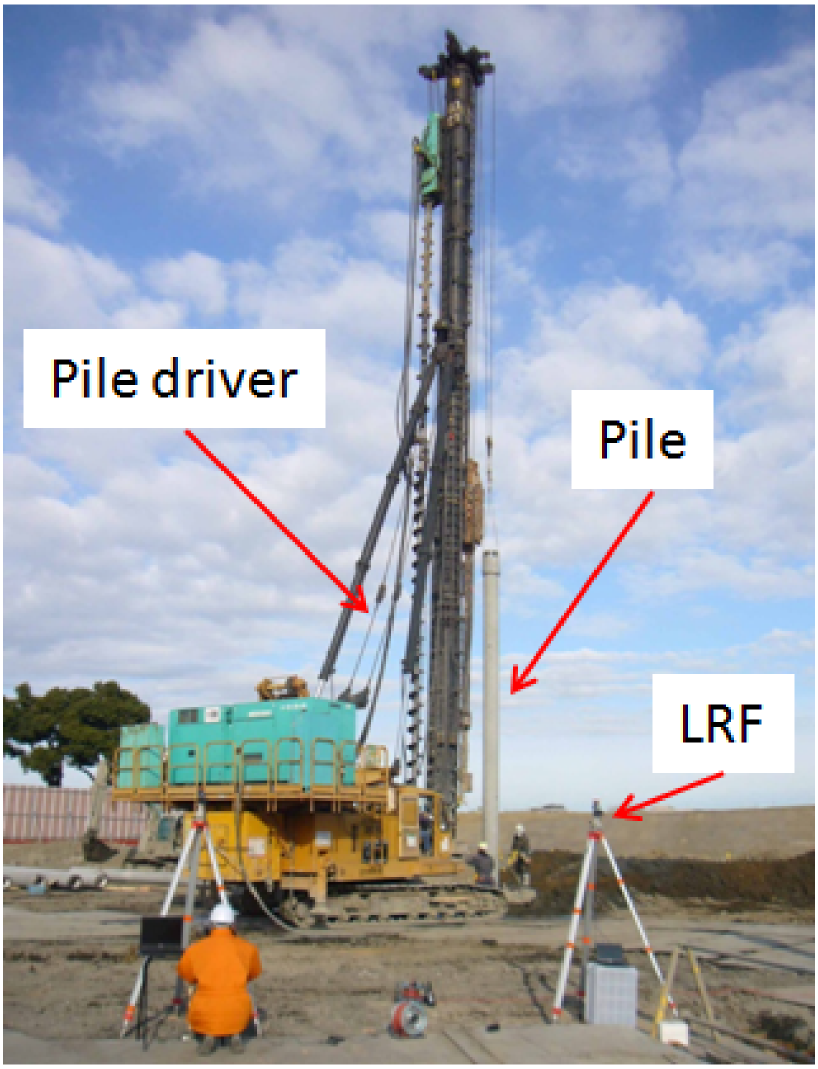


Figure 2.15: Scene of position measurement of pile driver

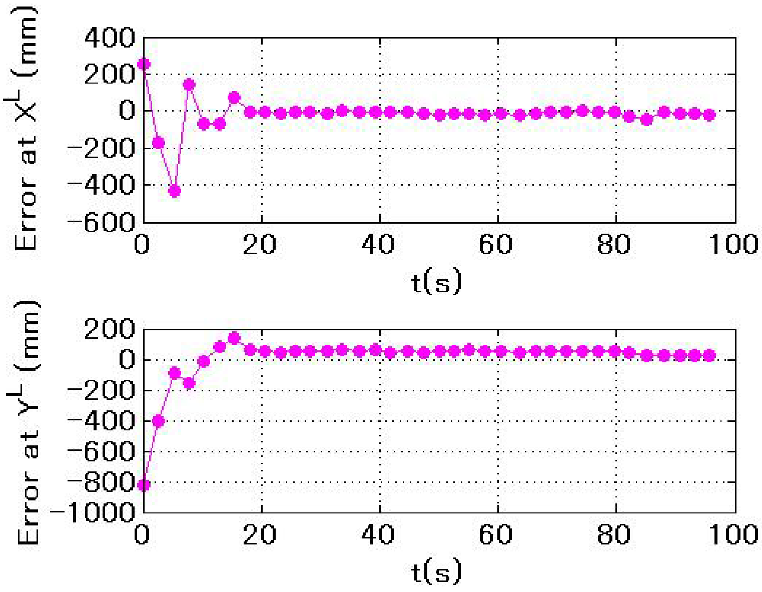


Figure 2.16: Measured pile position (X^L, Y^L) relative to the destination (a, b) of construction field experiment

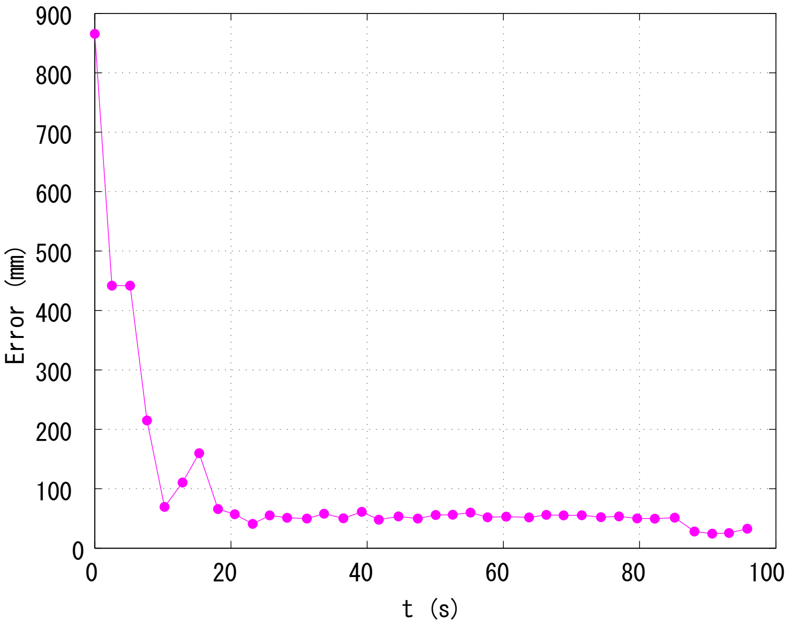


Figure 2.17: Distance to the pile destination

Currently there is no other direct way to measure the center position of pile while it moves. The accuracy of the traditional surveying system is $10mm$. In addition, from this experiment, it is certain that the pile was driven within the allowable range. We can estimate that the construction error is within the range of $15-35mm$. Considering the manual adjustment process, the result of our system is considered reasonable. The proposed system can measure the pile driving position in real-time which has not been possible by conventional surveying instruments.

2.6 Summary

In this chapter, based on circle detection and fitting, we propose an novel real-time pile driver positioning system using laser range finder. Taking advantages of LRF, such as high accuracy, fast data acquisition and large covering area, and utilizing the orientation-invariant property of the cylindrical target, a new surveying technique is presented. To extract the pile target from pile driver, we propose a circle detection algorithm. Then the MLE method is adopted to fit circle from extracted arc data. A pan unit is applied in the system to improve the angular resolution of LRF.

The simulation and indoor experiment prove the reliability and flexibility of the proposed circle detection algorithm. The experiment on the actual construction field shows that the proposed system can keep tracking the pile position in real-time while the pile driver works, which is impossible for the conventional surveying methods.

For the purpose of verifying the feasibility of proposed system, we only use data from a single LRF currently. Since multiple sensors could increase the amount of information and enlarge the positioning range, multiple sensor system will be investigated and implemented. In the current implementation, the number of needed votes is empirically determined. The robustness of this parameter should be investigated.

Chapter 3

3D Shape Reconstruction of Unknown Targets by Dynamic Sensing

This chapter describes how to rectify the distorted data of unknown targets from a moving range sensor.

3.1 Basic Idea

Since raw range data only have sensor-oriented coordinates, data from different view points are described in different coordinate system. In the continuously data gathering mode, each point has a individual coordinate system. To correctly align those points, we need to apply unique transformation to each point to make them described in the unified coordinate system. When there are some overlapped areas between two data sets, we apply transformations to make them coincide.

Those transformations can be derived from the sensor motion. We build a proper model to describe the sensor motion. Utilizing the correspondence of feature points in overlapped areas, we can construct a cost

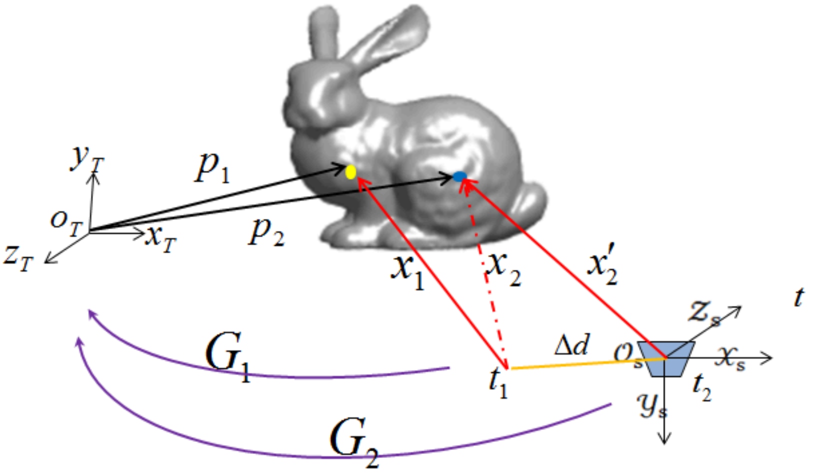


Figure 3.1: Coordinate system of moving range sensor and targets

function to evaluate the motion model. Sensor motion is estimated by minimizing the cost function. Finally the distorted data can be rectified based on the transformation derived from estimated sensor motion.

3.2 Sensor Motion Model

3.2.1 Common Representation of Rigid Body Motion

- Homogeneous Coordinates

  The pose of range sensor relative to object frame can be expressed as a rigid body transformation in R^3 using homogeneous coordinates as:

$$q_t = G \cdot q_0 \text{ ,with } G = \begin{bmatrix} r_{11} & r_{12} & r_{13} & d_x \\ r_{21} & r_{22} & r_{23} & d_y \\ r_{31} & r_{32} & r_{33} & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.1}$$

  $q_0 = [\ x_0 \quad y_0 \quad z_0 \quad 1\ ]^T$ is the initial pose of range sensor in the object frame, $q_t = [\ x_t \quad y_t \quad z_t \quad 1\ ]^T$ is the pose of range sensor at time t in the object frame.

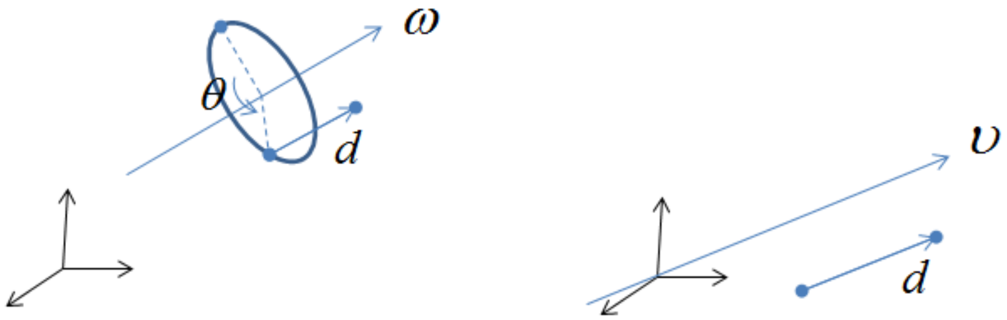


Figure 3.2: Definition of twist

The arbitrary 3D translation $[\begin{array}{ccc} d_x & d_y & d_z \end{array}]^T$ has 3 degrees of freedom. It looks like there are 9 unknowns in rotation matrix:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{3.2}$$

But constrained by the orthonormality property of rotation matrix, it actually has only 3 degrees of freedom. Therefore the transformation G has 6 degrees of freedom.

- Twist Coordinates

  According to Chasles' Theorem [39], every rigid body motion can be realized by a rotation about an axis combined with a translation parallel to that axis. As shown in Fig. 3.2, ω stands for the rotation axis. θ is the rotated angle about ω. d is the translation parallel to the axis. It can be represented using exponential of twist ξ, as shown in following equation:

$$G = e^{\hat{\xi}} = I + \hat{\xi} + \frac{(\hat{\xi})^2}{2!} + \frac{(\hat{\xi})^3}{3!} + \cdots \tag{3.3}$$

$$\text{where } \xi = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} \hat{\xi} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & v_2 \\ -\omega_2 & \omega_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.4}$$

where the rotated angel θ is the norm of ω. v describes the translation parallel to the axis ω.

- Euler angles
  One method of describing the orientation of a coordinate frame B relative to another coordinate frame A is as follows: start with frame B coincident with frame A. First, rotate the B frame about the z-axis of frame B (at this time coincident with frame A) by an angle α, then rotate about the (new) y-axis of frame B by an angle β, and then rotate about the (once again, new) z-axis of frame B by an angle γ. This yields a net orientation $R_{ab}(\alpha, \beta, \gamma)$ and the triple of angles (α, β, γ) is used to represent the rotation. The angles (α, β, γ) are called the ZYZ Euler angles. Since all rotations are performed about the principal axes of the moving frame, we define the following elementary rotations about the x-, y-, and z-axes:

$$R_x\left(\phi\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi & cos\phi \end{bmatrix}$$

$$R_y\left(\beta\right) = \begin{bmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{bmatrix}$$

$$R_z(\alpha) = \begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The final orientation of frame B, can be derived from the product of sequence of rotations of frame A, considering the B frame as fixed.

$$R_{ba} = R_z(-\gamma)R_y(-\beta)R_x(-\alpha) \tag{3.5}$$

- Quaternion

  Quaternions generalize complex numbers and can be used to represent rotations in much the same way as complex numbers on the unit circle can be used to represent planar rotations. Unlike Euler angles, quaternions give a global parameterization of $SO(3)$, at the cost of using four numbers instead of three to represent a rotation.

  Formally, a quaternion is a vector quantity of the form

  $$Q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k},\ q_i \in \mathbb{R}, i = 0, \cdots, 3$$

  where q_0 is the scalar component of Q and $\boldsymbol{q} = (q_1, q_2, q_3)$ is the vector component.

In this thesis, we use twist coordinate to represent sensor motion. There are two main advantages to using twists for describing rigid body kinematics. The first is that they allow a global description of rigid body motion which does not suffer from singularities due to the use of local coordinates. Such singularities are inevitable when one chooses to represent rotation via Euler angles, for example. The second advantage is that screw theory provides a very geometric description of rigid motion which greatly simplifies the analysis of mechanisms. [39] In the case of quaternion, an extra number of variable is needed. So we choose twist to describe sensor model.

3.2.2 Approximation of Sensor Motion

For convenience without loss of generality, we have made an assumption that the motion of sensor is smooth, which is appropriate in most practical applications. This means that the sensor motion can be represented by a smooth function whose continuity is higher than C^0. Such a smooth function can be represented by a sum of smooth basis functions, and the most simple case is the polynomial representation.

- Polynomial Representation

  While the sensor moves, the transformation changes in accordance with time t, and the twist coordinate is represented by functions of time t as $\xi(t)$

  $$\xi(t) = \xi_0 + \xi_1 \cdot t + \xi_2 \cdot t^2 + \xi_3 \cdot t^3 + \cdots \tag{3.6}$$

  which we call *polynomial motion*. Without loss of generality, we can assume that, initially at $t = 0$, the sensor and object coordinate systems are identical, and hence $\xi_0 = 0$. We denote the coefficients of polynomial motion of N-th order by $\xi(t)_N$, where $\xi_1, \xi_2, \cdots, \xi_N$ are 6×1 twist vectors. Here we assume that all parameters are constant in the whole sensor motion sequence of a time duration T_c.

  Therefore the trajectory of range sensor in the object frame $q(t)$ can be represented as:

  $$q(t) = G(t) \cdot q(0) = e^{\hat{\xi(t)}} \tag{3.7}$$

  Given the motion model of sensor $G(t)$, a point of the object measured at time t with local range sensor coordinate p_r, can be transformed to the object frame:

  $$q_o = G(t) \cdot p_r \tag{3.8}$$

  Here we give some examples to show that the use of polynomial approximation of sensor motion in twist coordinate system is suitable for common motions.

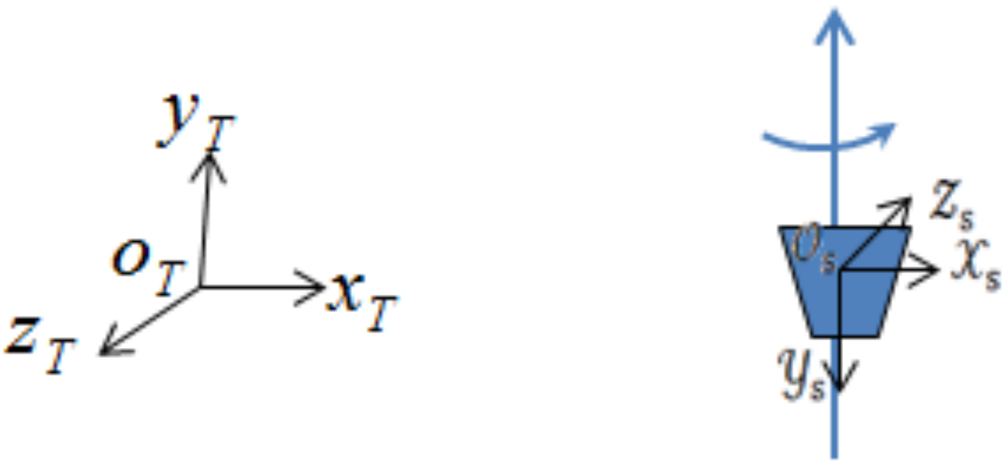


Figure 3.3: Pure rotation

- Pure Rotation

  In the case of pure rotation motion with a uniform angle velocity α rotating about y axis, it can be expressed as

$$\xi(t) = \begin{bmatrix} v(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \alpha \cdot t \\ 0 \end{bmatrix} = \xi_1 \cdot t \tag{3.9}$$

- Uniform Accelerated Rectilinear Motion

  In the case of uniform accelerated rectilinear motion along x axis, it can be described as

$$\xi(t) = \begin{bmatrix} v(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} a_1 \cdot t + a_2 \cdot t_2 \\ 0 \\ 0 \\ 0 \\ \alpha \cdot t \\ 0 \end{bmatrix} = \xi_1 \cdot t + \xi_2 \cdot t^2 \tag{3.10}$$

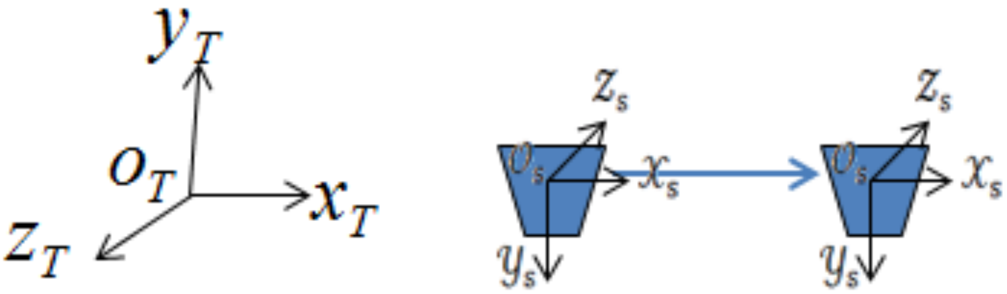


Figure 3.4: Uniform rectilinear motion

- Uniform Circle Motion

  In the case of uniform circle motion with a constant angle velocity β in $x - z$ plane, it can be written as

$$\xi(t) = \begin{bmatrix} v(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} \beta \cdot t \\ 0 \\ 0 \\ 0 \\ \alpha \cdot t \\ 0 \end{bmatrix} = \xi_1 \cdot t \tag{3.11}$$

- Cubic Spline Representation

  While the polynomial fitting of curve has the advantage of simplicity, it may become unstable when the order of polynomial is higher than 7. The actual trajectory of sensor could be very complex with respect to different applications. It may require high order polynomial representation which leads to unstable result. A more sophisticated way is to use spline curve. Here we assume that sensor trajectory is a combination of a set of common motions defined before. Since the order of polynomial representation of those common motions is lower than 3, cubic spline is enough for our applications.

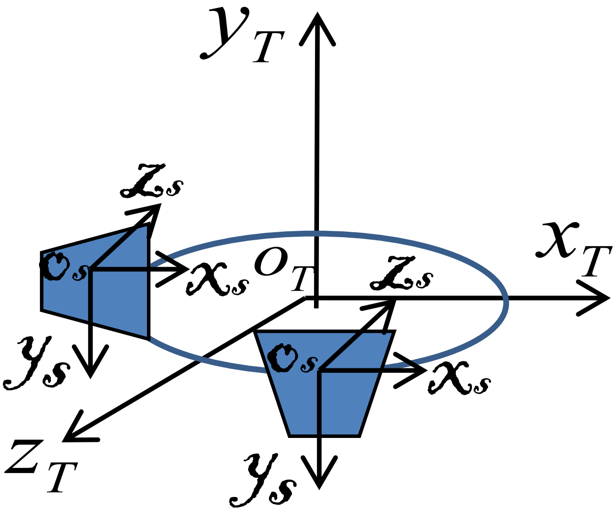


Figure 3.5: Uniform circle motion

Given $N + 1$ control points $[t_i, y_i]$, the function $\xi(t)$ can be re-written in the form of cubic spline as

$$\xi_i(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i \\ \text{for } t \in [t_i, t_{i+1}], i = 0, 1, \cdots, N - 1 \tag{3.12}$$

Since there are N intervals and four coefficients for each we require a total of $4N$ parameters to define the spline $\xi(t)$.

1. $\xi_i(t_i) = y_i$, where $i = 0, 1, \cdots, N - 1$
2. $\xi_i(t_{i+1}) = y_{i+1}$, where $i = 0, 1, \cdots, N - 1$
3. $\xi'_{i-1}(t_i) = \xi'_i(t_i)$, where $i = 1, 2, \cdots, N - 1$
4. $\xi''_{i-1}(t_i) = \xi''_i(t_i)$, where $i = 1, 2, \cdots, N - 1$

These conditions result in $4N - 2$ constraints. So we need two more conditions to completely fix the spline. According to different initial motion status of the sensor, we can add two conditions: $\xi'_0(t_0)$ and $\xi''_0(t_0)$.

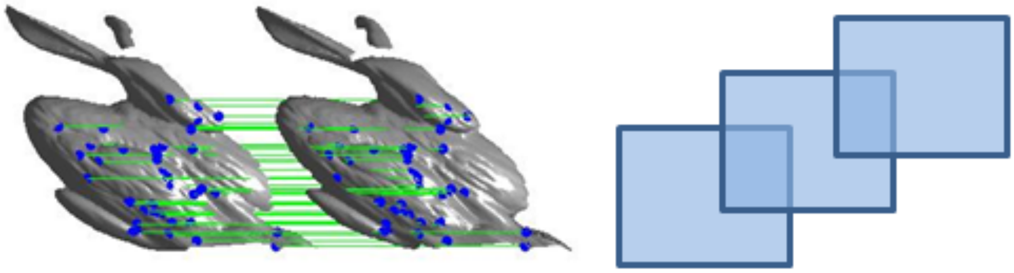


Figure 3.6: Overlapped areas between sequent range images.

3.3 Motion Model Estimation

3.3.1 Corresponding Constraints in Overlapped Areas

When we use a 3-D continuously scanning range sensor, there are overlap regions in sequent range images, through which we assume that the sensor motion is represented by a unique sensor motion model. If a point with its object coordinate p_o is measured in two range images indexed by 1 and 2, the sensor coordinates are related to the world coordinates by:

$$\begin{aligned} p_o &= G(t_1) \cdot p_{r1} \\ p_o &= G(t_2) \cdot p_{r2} \end{aligned}$$

Where p_{r1} and p_{r2} are the sensor coordinates and t_1 and t_2 are the measured time of point p_o in two range images successively. By eliminating the object coordinate p_o, the constraint equation describing the correspondence of feature points can be written as:

$$G(t_1) \cdot p_{r1} - G(t_2) \cdot p_{r2} = 0 \tag{3.13}$$

3.3.2 Cost Function

Each pair of corresponding points can give a constraint equation. Theoretically, when the number of equations is more than the number of unknown variables in $G(t)$, we can have a solution of sensor motion. In the actual case, we usually have redundant equations. To find an optimal solution from those constraints, we construct a cost function.

Given K points $p_1, \cdots, p_K$ with measurements $(p_{r1,i}, t_{1,i}, p_{r2,i}, t_{2,i}), 1 < i < K$, when the motion parameter of twist ξ is represented by N-th order polynomial, the optional solution is given by minimizing the cost function defined by the sum of squared errors:

$$E(\xi) = \sum_{i=1}^{K} ||G(t_{1,i}) \cdot p_{r1,i} - G(t_{2,i}) \cdot p_{r2,i}||^2 \tag{3.14}$$

with $G(t) = e^{\hat{\xi(t)}}$,where $\xi(t) = \xi_1 \cdot t + \xi_2 \cdot t^2 + \cdots + \xi_N \cdot t^N$.

In this cost function, there are $6N$ unknowns. Since each pair of feature points can give 3 constraint equations, theoretically, as long as $3K \geq 6N$, there would be a solution for the problem.

We minimize the cost function to find the optimal solution. Here we adopted Levenberg-Marquardt algorithm to solve it.

3.4 Simulation Experiments and Discussions

To verify and evaluate our proposed algorithm, we test it with simulated data which can afford ground truth. To show the robustness under different conditions, like the complexity of sensor motion and the shape of objects, we build a data set which contains the regular motion modes and typical targets in applications such as digitalization of culture heritages, autonomous vehicle.

Table 3.1: Virtual sensor specification

Vertical range	30°
Horizontal range	90°
Angular resolution	1/16°
fps	0.5
Scanning order	Raster

3.4.1 Setting up

The characteristic of virtual range sensor is set up as Table 3.1. We generate the range data by simulating the process that a straight line hits on the object surface in raster order with a given time sequence and given angular resolution. Thus we can obtain measured data, scanning time and corresponding ground truth.

We don't consider the feature detection procedure here and just make an assumption that the correspondence of feature points in different frames is known and established by methods like manual operation or feature matching. Another assumption is that all feature points are taken from static rigid objects. In our experiments, we generate feature points in a manual way, randomly choosing points from the overlapped areas of two successive frames.

3.4.2 Evaluation Methods

To evaluate the rectification performance, we compute the point distances between the rectified points $\{\mathbf{R}\}_1^N$ to ground truth $\{\mathbf{S}\}_1^N$. Here we define the absolute rectification error as:

$$e_i = ||\mathbf{R}_i - \mathbf{S}_i|| \tag{3.15}$$

The absolute error is affected by point resolution, which is related to sensor specifications and distances between sensors and observed objects. In order to evaluate the algorithm despite of those factors, we define the

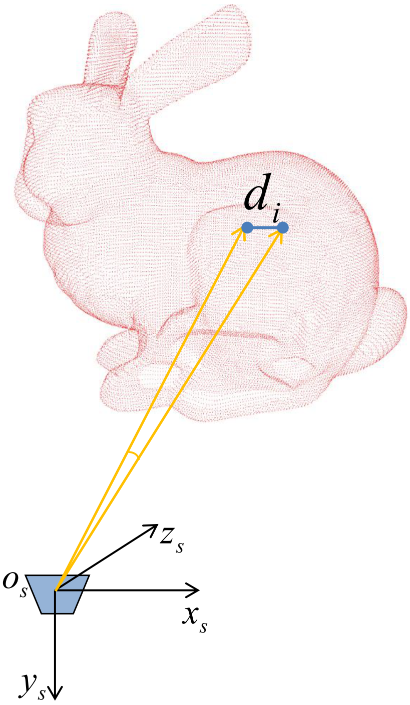


Figure 3.7: Point-wise distance.

normalized rectification error as:

$$E_i = \frac{e_i}{d_p} \tag{3.16}$$

where $d_p = \frac{1}{K}\sum_{i=1}^{K} d_i$ is the mean point-wise distance of the object, as shown in Fig.3.7.

3.4.3 Tests on Common Motion Modes

To evaluate our algorithm in different types of sensor motions, we test it under 5 general motion modes as shown in Fig.3.8. As shown in the first column of Fig.3.9, measurements are distorted by sensor motion. The second column of Fig.3.9 shows our rectified results. Table 3.2 shows the parameters of simulated motion modes in the data set.

Table 3.2: Regular motion modes

Motion description
1.Constant velocity
2.Constant acceleration
3.Pure rotation
4.Combination of 2 and 3
5.Uniform circular motion

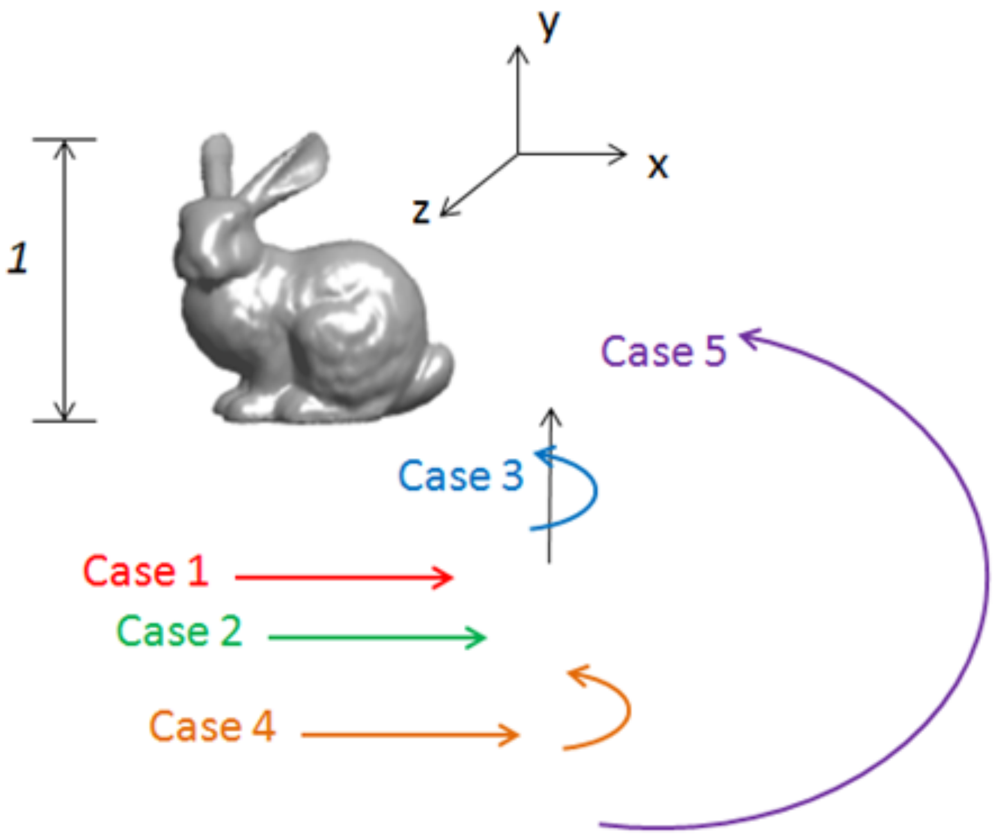


Figure 3.8: Simulated motion modes

3.4.4 Tests on Different Target Sets

We verify the proposed algorithm with different targets, as shown in Fig.3.10. As Fig.3.10 shows, as long as enough corresponding point are

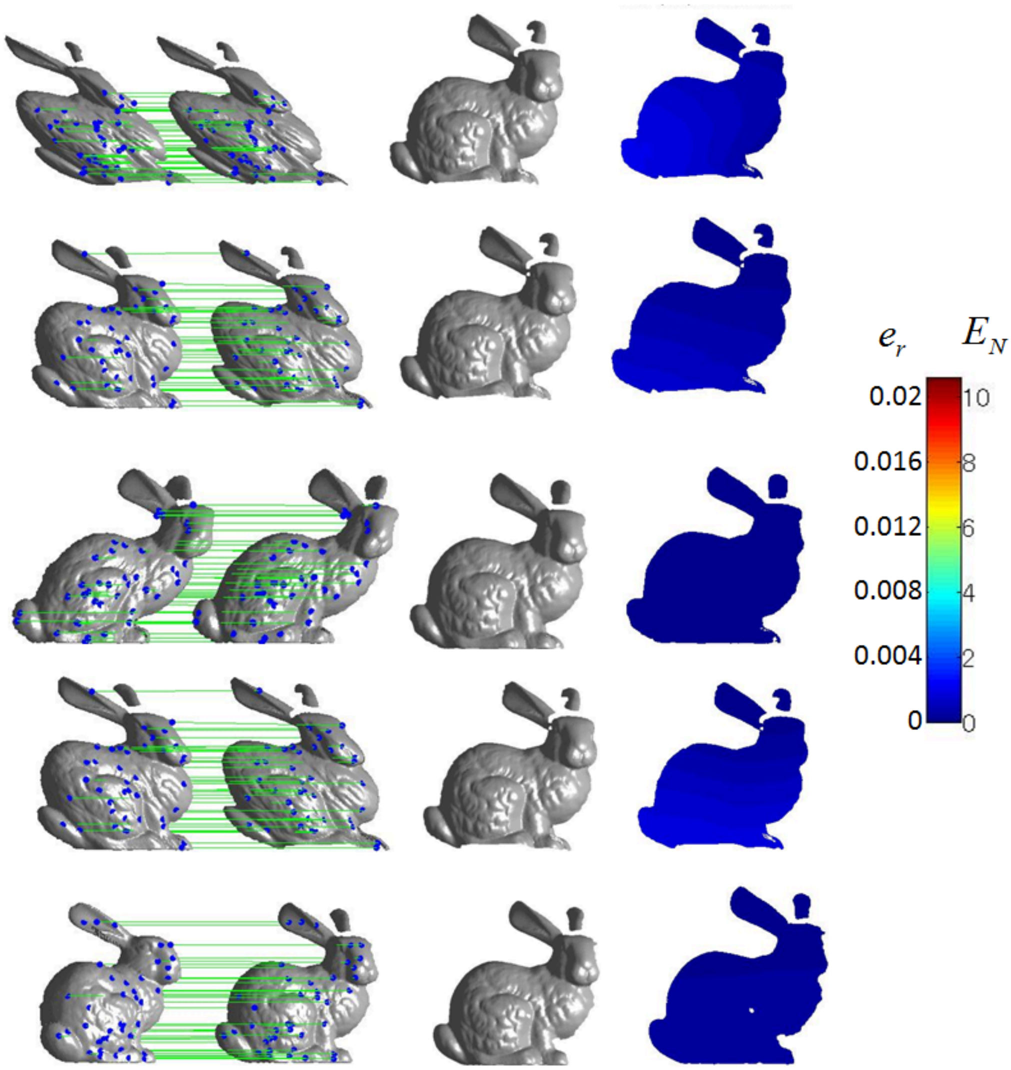


Figure 3.9: Simulation experiment on Stanford bunny model under 5 common motion modes. The first column: distorted data and correspondence between consecutive range image. The second column: rectified data using our method. The third column: rectification error.

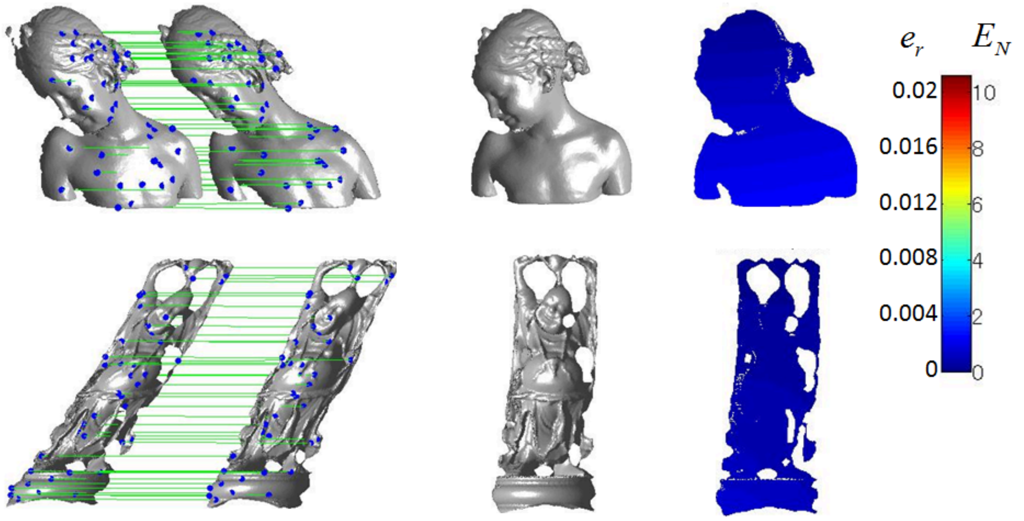


Figure 3.10: Simulation experiment with other objects under constant velocity motion. The first column: distorted data and correspondence between consecutive range image. The second column: rectified data using our method. The third column: rectification error.

sampled, the proposed algorithm works on different targets.

3.4.5 Robustness to Correspondence Noise

To verify the robustness of the proposed algorithm, we add noise to the correspondence constraints. The unit of noise is mesh resolution. As shown in Fig.3.12, the rectification error has a linear relation with correspondence noise.

3.4.6 Discussion on Optimization

In the experiment, we notice that for a given polynomial motion $G(t)_N$, the initial guess of the parameters in the Levenberg-Marquardt optimization algorithm has large effect on the result and iteration time. Empirically, we found that use the estimation result of low order polynomial motion

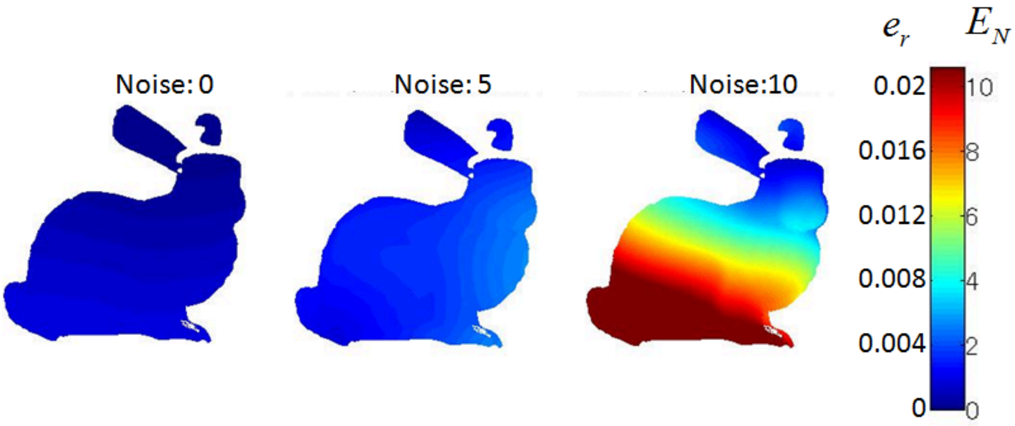


Figure 3.11: Rectification error with correspondence noise

$G(t)_{i-1}$ as the initial guess of $G(t)_i$ can give better result than directly estimating $G(t)_i$. In most cases we can achieve an acceptable solution by starting this iterative procedure from $G(t)_1$ with the initial guess $\xi_1 = 0$.

3.5 Application in 3D Reconstruction of Indoor Environment

The problem of fast 3D modeling of indoor environment attacks a lot of attention because of its wide applications to intelligent building, since 3D models can provide richer information about the environment. The state-of-the-art technologies of 3D modeling usually involve devices like camera, laser scanner, GPS with mobile platform. Laser scanner based systems have better accuracy than camera-based systems. Since high accurate localization device like GPS cannot receive satellite signals in indoor environment, information fusion with other devices like camera, IMU and encoder is a common solution.

Here we apply our proposed algorithm to build 3D model of an office environment, as shown in Fig.3.13. The purpose of this experiment is to verify that our algorithm can collect and self-calibrate data with moving

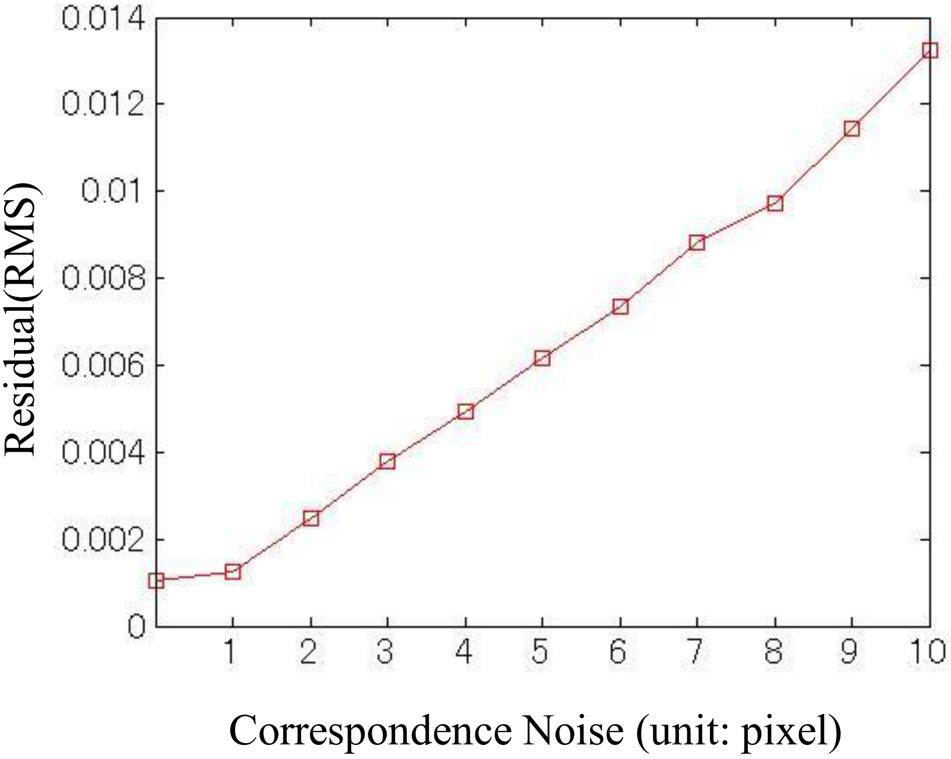


Figure 3.12: Rectification error with respect to correspondence noise.

platform using a single 3D laser range scanner. Such system has lower cost and works more efficiently.

3.5.1 Setting up

As shown in Fig.3.14, the laser range scanner is mounted on a cart. In order to cover more upper space, we set the scanning central axis a small upward angle with the horizontal plane. Details of sensor specification is in Table 3.3.

In the experiment, we mannually move the platform in the office to scan it in one round. Examples of data distortion caused by sensor movement are shown in Fig.3.15.

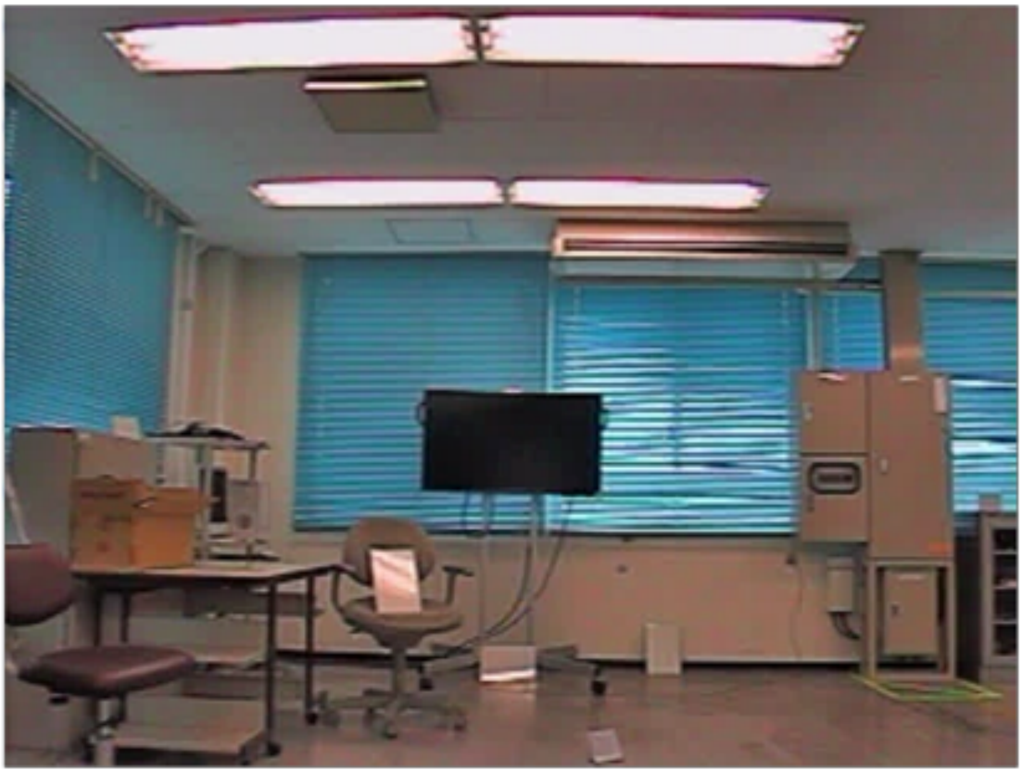


Figure 3.13: Scene of the indoor experiment.

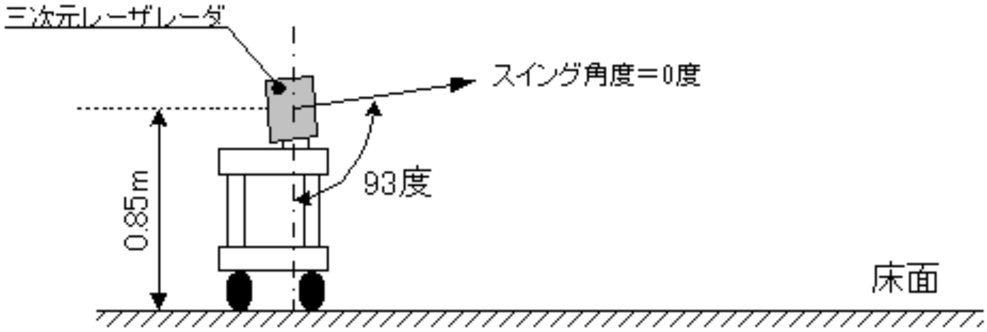


Figure 3.14: Setting up of range sensor in the indoor experiment.

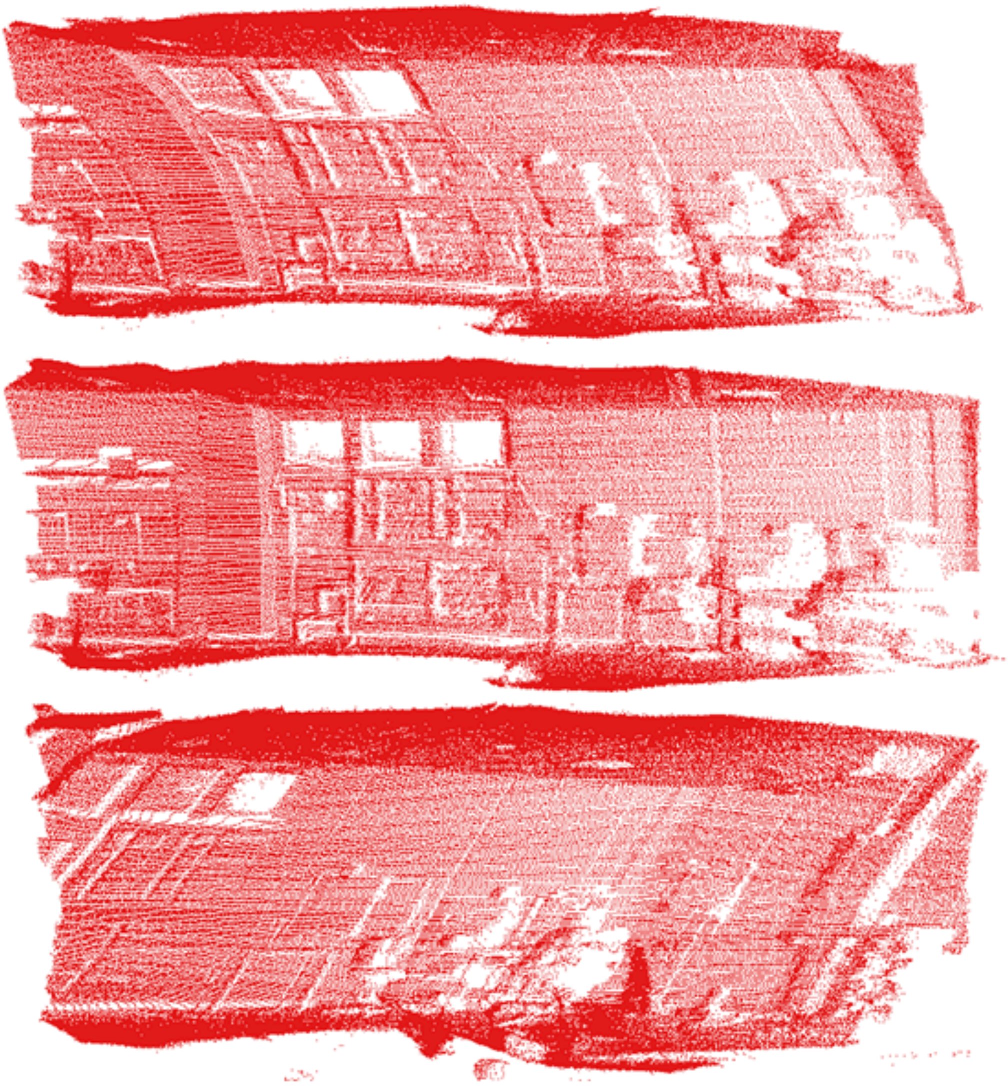


Figure 3.15: Distorted data of the indoor scene.

Table 3.3: Specification of IHI 3D laser range sensor

Model Number	IHI
Light source	JIS C 6802 Class1
Vertical field of view	60°
Horizontal field of view	90°
Measurable distance	$0.1[m] \sim 200[m]$
vertical angular resolution	0.01°
horizontal angular resolution	0.15°
Frame rate	$0.3 \sim 10$Hz

3.5.2 Results and Discussion

In order to evaluate our algorithm without the effect of performance of corresponding extraction step, we manually pick a set of corresponding points, as the green lines shown in Fig.3.16. Fig.3.17 shows the comparison between distorted data and rectified data by our algorithm.

One way to evaluate the data rectification is to check how well the point clouds can be aligned with ground truth, which is the point clouds $\{\mathbf{S}\}_1^N$ obtained by static range sensor. To align two point clouds, we adopt iterative closest point (ICP) method [69] to find the best alignment transformation, yeilding aligned point clouds $\{\mathbf{X}\}_1^M$ for raw distorted data, and $\{\mathbf{Y}\}_1^K$ for rectified data.

Now we can compute closest point distances for all points in $\{\mathbf{X}\}_1^M$ to $\{\mathbf{S}\}_1^N$:

$$d_m = d(\mathbf{X}_m, \{\mathbf{S}\}_1^N) = \min_{n \in [1,N]} ||\mathbf{X}_m - \mathbf{S}_n|| \tag{3.17}$$

Similarly, we can calculate closest point distance for point in rectified data set to the ground truth:

$$d_k = d(\mathbf{Y}_k, \{\mathbf{S}\}_1^N) = \min_{n \in [1,N]} ||\mathbf{Y}_k - \mathbf{S}_n|| \tag{3.18}$$

d_m and d_k are mapped as colors in Fig.3.18. Point distances increase from

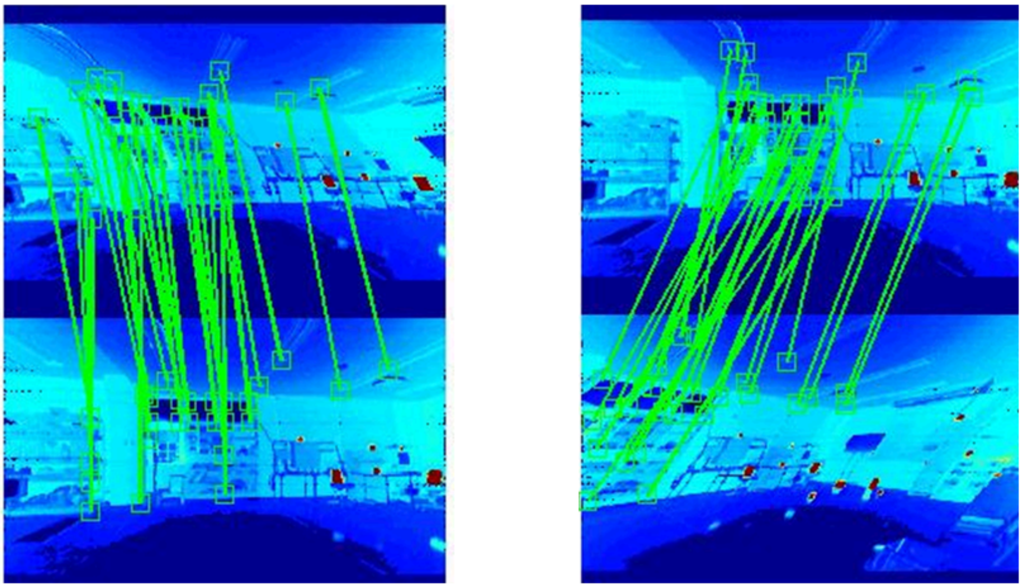


Figure 3.16: Manual correspondence of distorted data from the indoor scene.

blue to red. Using these distances, we can obtain two probability density error curves based on kernel density estimation (KDE) [61]. As shown in Fig.3.19, $p(d)$ of rectified data has more data with small error.

Another way to evaluate the point cloud rectification, is utilizing prior knowledge of indoor environment, that angles between walls, floor and ceiling should be 90 degree, as shown in Fig.3.20. We estimate plane normals of the subset of point clouds which only contain the wall,floor and ceiling parts using RANSAC. We then calculate the angle between two normals:

$$\theta_{i,j} = \arccos \hat{n}_i \cdot \hat{n}_j, \tag{3.19}$$

where $\hat{n}_i$ and $\hat{n}_j$ are normals for the two planes.

As shown in Table 3.4, the angles after rectification are much closer to the ground truth.

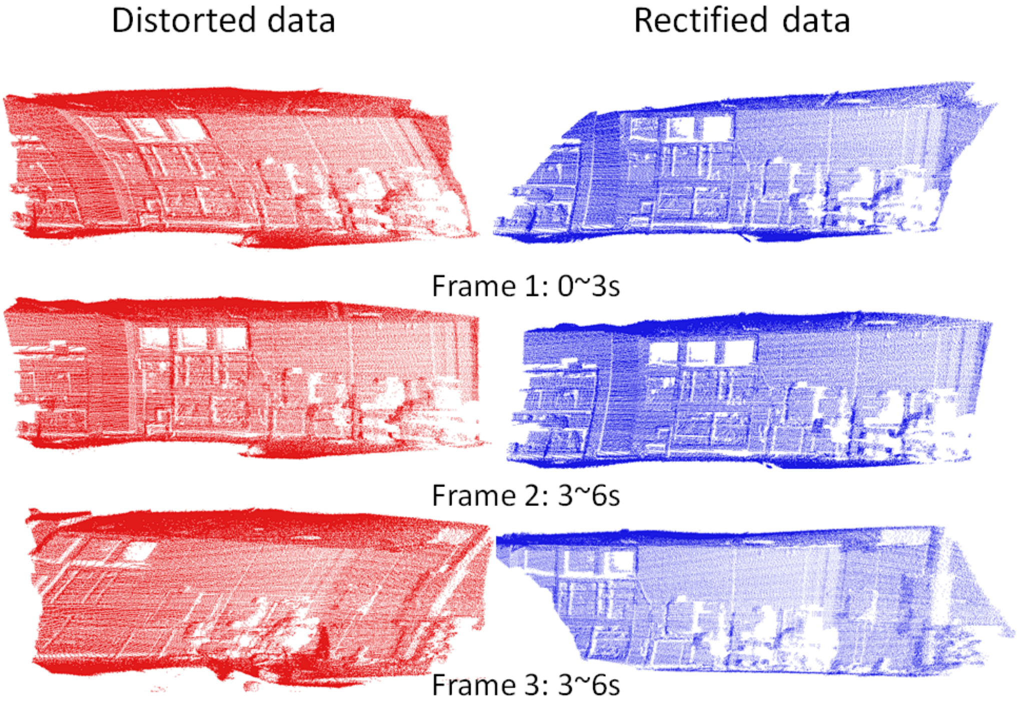


Figure 3.17: Rectification result of the indoor scene.

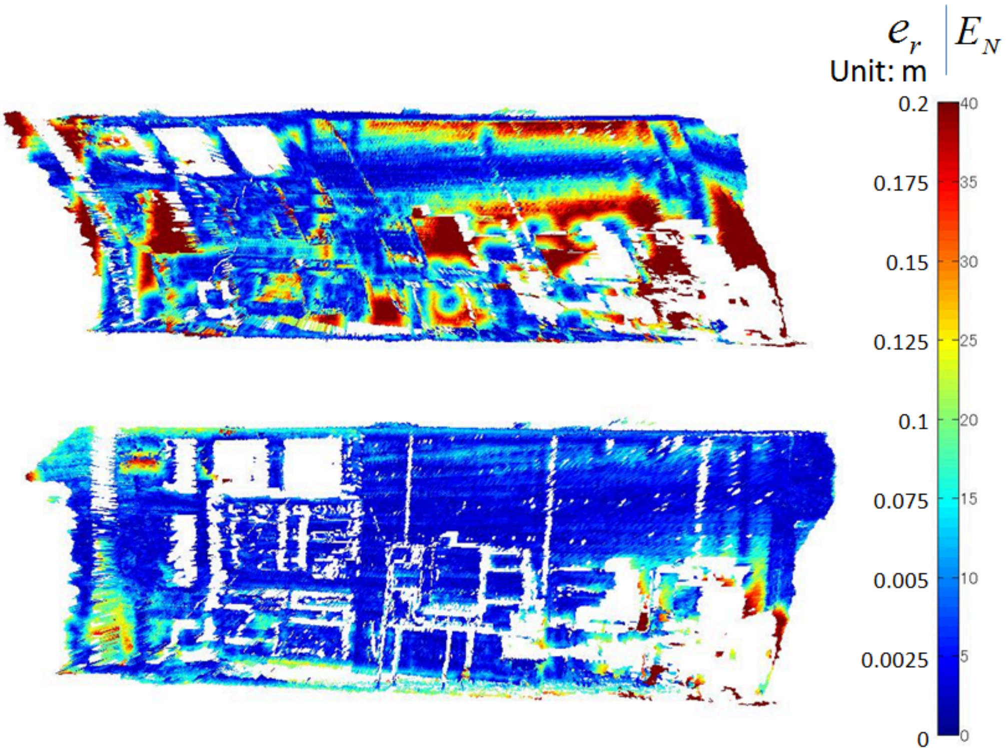


Figure 3.18: Rectification error of the indoor scene.

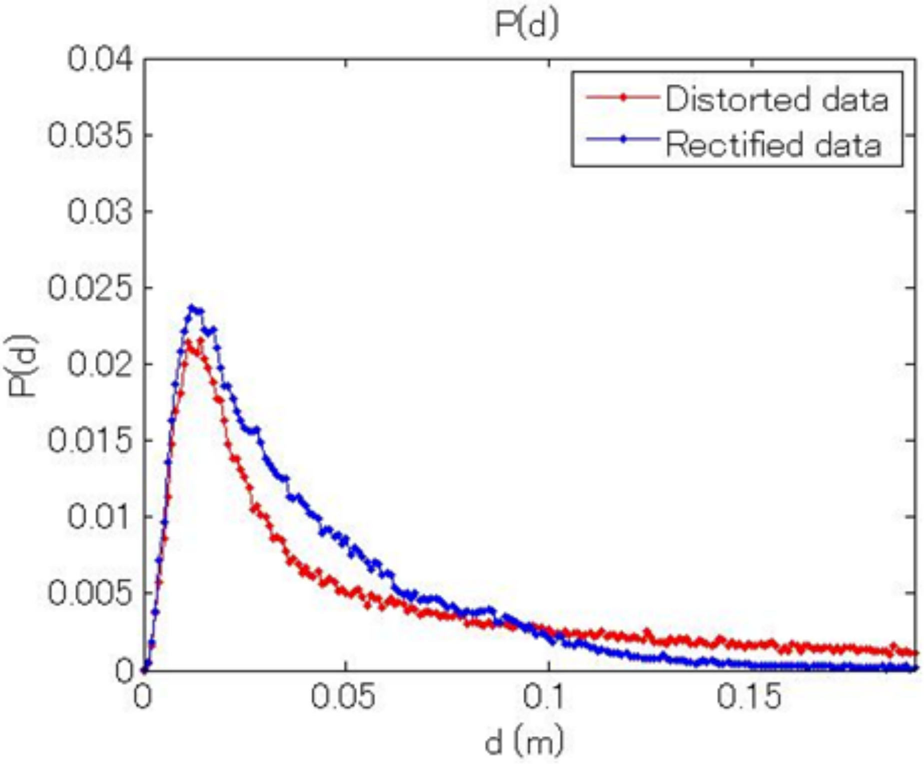


Figure 3.19: Comparison of error distribution curves between distorted data and rectified data of the indoor scene.

3.6 Application in Large Scale Culture Heritage Digitization

3.6.1 Bayon Project

One of the most important and comprehensive applications of 3D technologies is modeling cultural heritage objects. It has great significance in many aspects. Modeling technologies can provide digital archive of object shapes of culture heritage. Digital data enables us to restore the original shapes of the heritage objects, even if some unfortunate disasters destroy them, like natural weathering, fire, or wars. Moreover, researchers can do further analysis on digital data applying modern computer vision and graphics technologies. In addition, normal people can

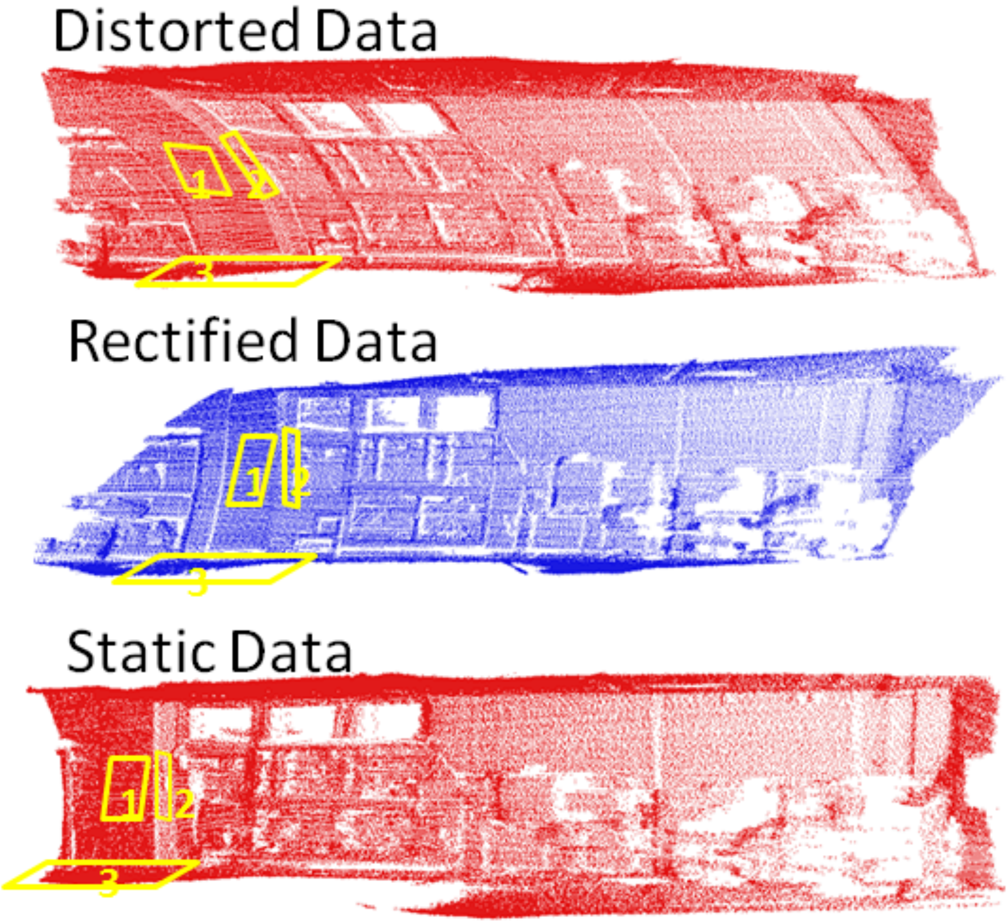


Figure 3.20: Comparison of plane angles between distorted data and rectified data of the indoor scene.

get access to digital culture heritages through the Internet from any corner around the world.

In order to preserve and study one important UNESCO World Heritage, the Bayon temple, which is located in the center of Angkor Thom in Cambodia, the Bayon Digital Archive Project [43] is started.

Table 3.4: Evaluation based on plane angles.

	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{2,3}$
Distorted data	87.0°	78.4°	52.2°
Rectified data	86.1°	84.4°	87.3°
Static data	88.9°	89.9°	88.8°

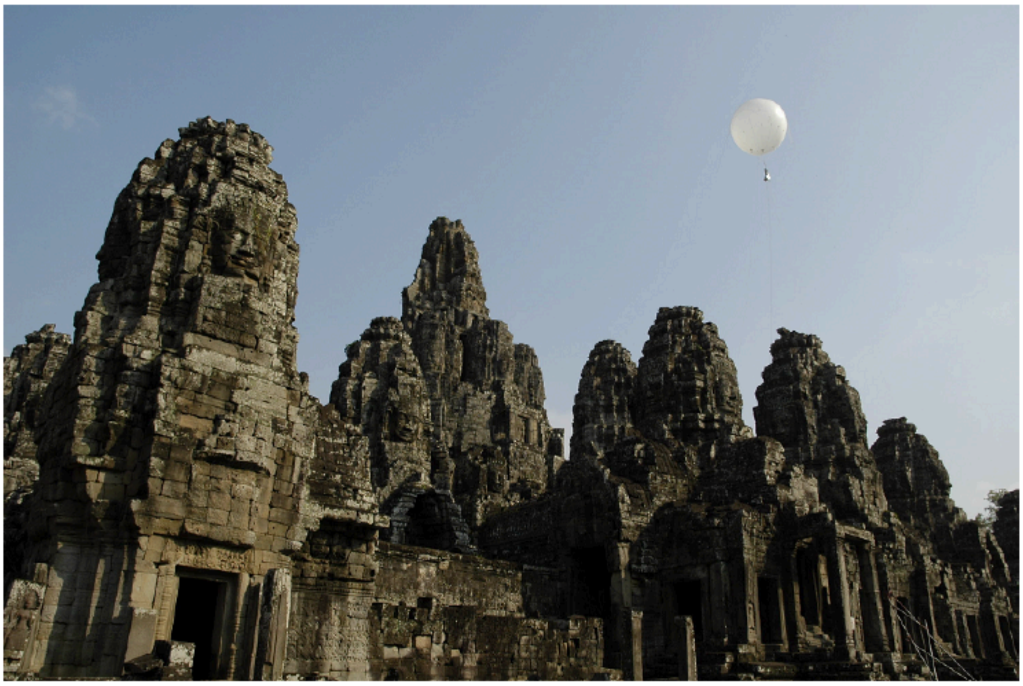


Figure 3.21: The FLRS and the Bayon Temple

3.6.2 Flying Laser Range Sensor

Several novel sensors are developed to observe this huge ($160m \times 140m \times 45m$) and complex structure. One of them is flying laser range sensor (FLRS), as shown in Fig.3.21. Using a balloon platform instead of previous helicopter platform, FLRS is safer and more flexible to measure large culture heritages in outdoor condition.

3.6.3 Results and Discussion

This system is certainly free from high frequency vibration such as that of a helicopter engine. However, the obtained range data are still distorted because the sensor itself is moving during the scanning processes, as shown in Fig.3.22.

We apply our method to the distorted FLRS data. Rectified data is shown in Fig.3.23.

3.7 Summary

In this chapter, we have presented a method to rectify distorted 3D data acquiring from moving range sensor based on approximating continuous sensor motion function using polynomial fitting of the twist representation of transformation. To verify and evaluate the proposed algorithm, we conducted experiments on simulated models which can provide ground truth. We test the algorithm with 5 basic motion modes: constant velocity, constant acceleration, pure constant rotation about fixed axis, uniform circle motion and combination of case 2 and 3. The results show that the distorted data is rectified properly only from the intrinsic property of the data itself. We also apply the proposed algorithm to actual data. One application is 3D reconstruction of indoor environment. Another application is 3D digitization of large scale world culture heritage.

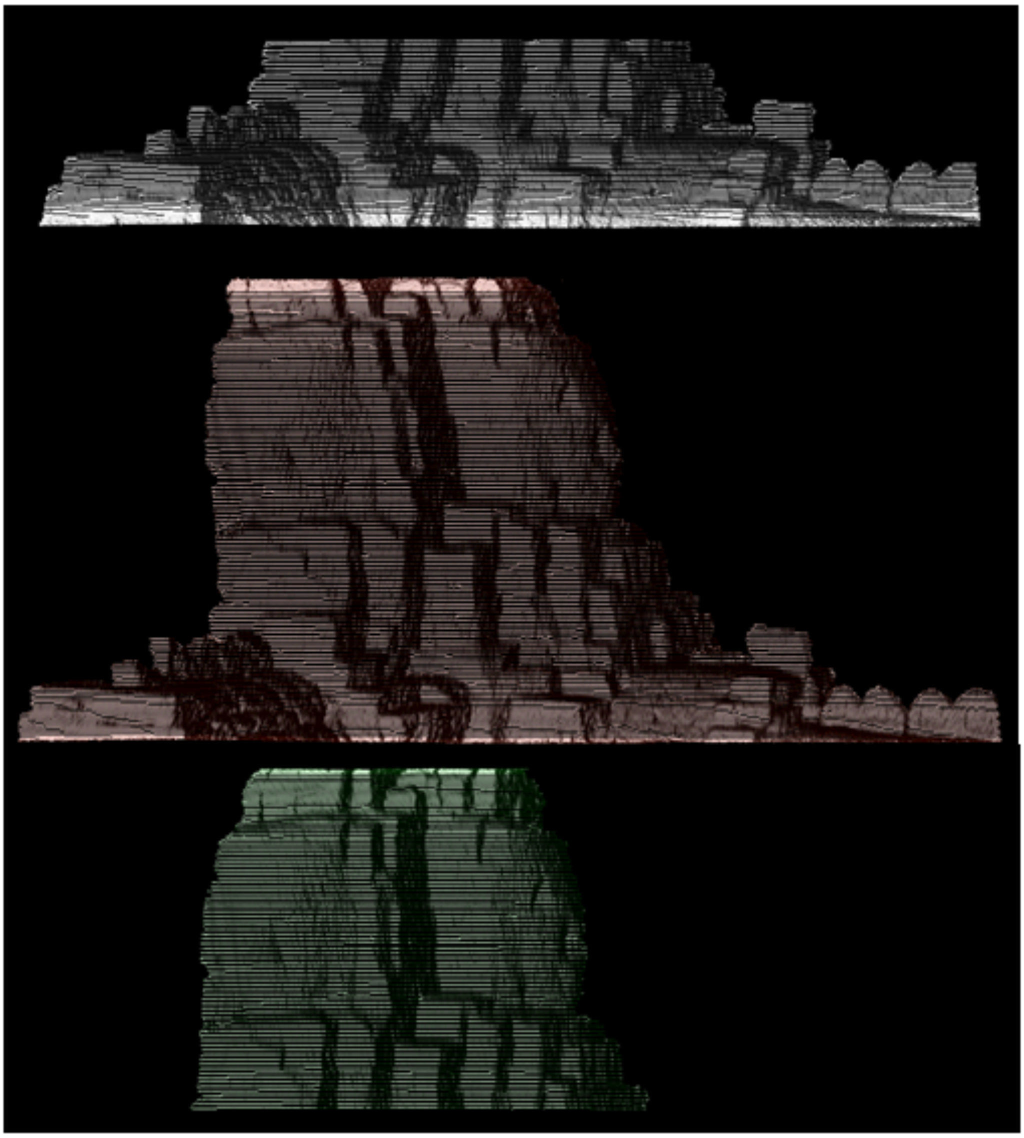


Figure 3.22: Distorted data of Bayon temple gathered by flying laser range sensor.

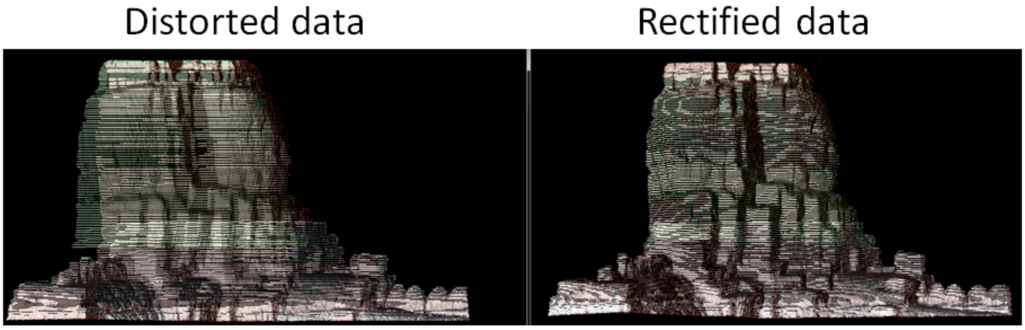


Figure 3.23: Rectified data of Bayon temple.

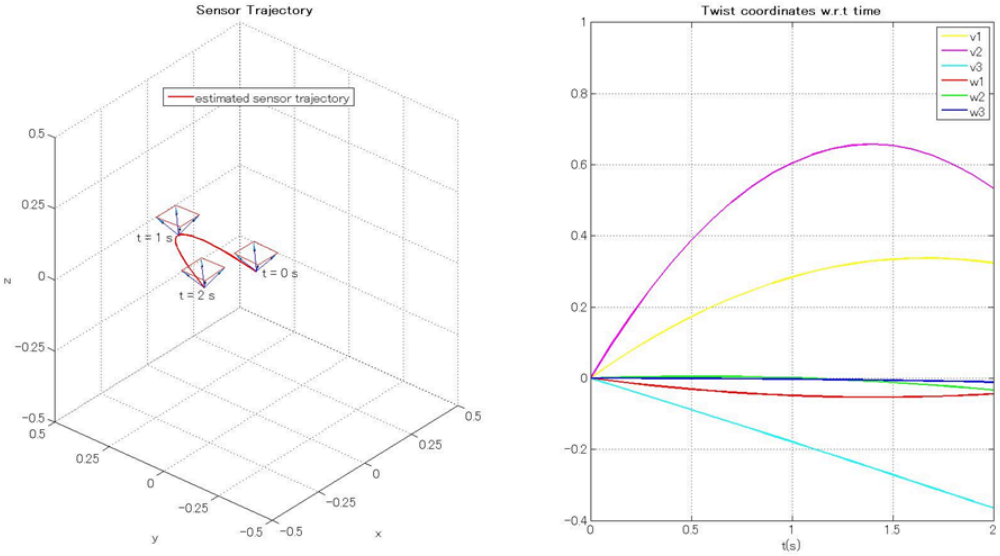


Figure 3.24: Estimated trajectory of flying laser range sensor.

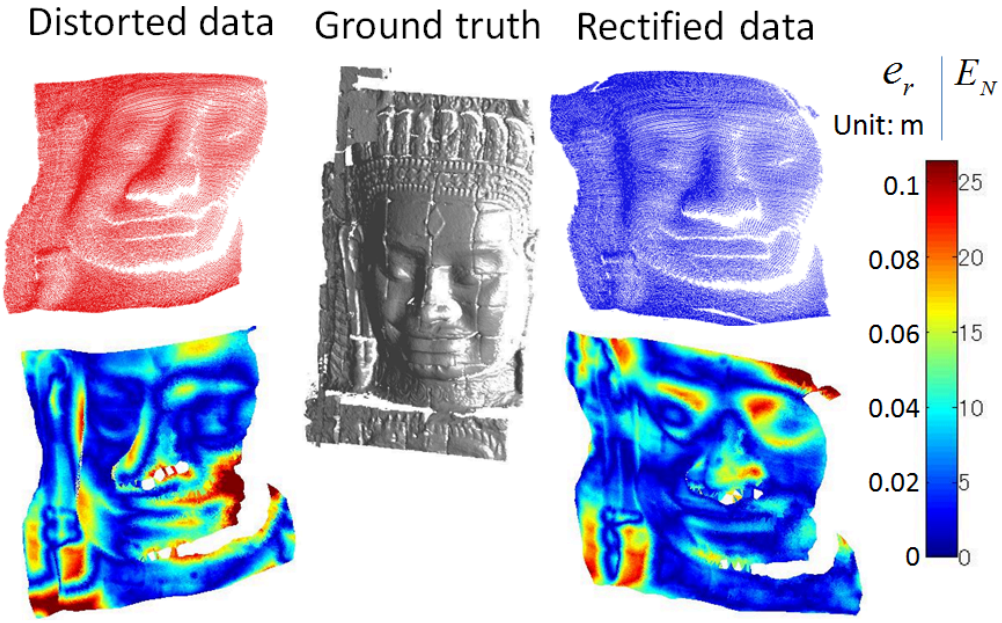


Figure 3.25: Test on a Buddha face.

Chapter 4

3D Correspondence Based on An Affine Invariant Shape Feature

4.1 Introduction

Finding meaningful correspondences between shapes is an important and fundamental task in shape analysis. It can also be referred to as alignment, registration, or matching under different contexts.

The definition of meaningful correspondences depends on the task you have. It could differ from identifying geometrically similar parts between shapes, to the problem of connecting elements which have the same function on the shapes.

In our case, the purpose of correspondence extraction is to offer geometrical constraints for the space-time reconstruction, especially for the 3D shapes distorted by sensor motion. Deformation coming from relative motion between range sensor and objects, has its own properties. It usually includes large amount of missing data, due to occlusion or change of scene or view points. The shape might be deformed significantly between scans and the deformation doesn't keep the geodesic distance invariant. Thus we consider local shape feature is better choice for this problem.

In this chapter, we propose a novel algorithm of extraction of 3D shape correspondences. The basic idea is utilizing Morse theory to extract topological information from information about critical points of a function. We design two types of Morse function for different conditions. We then find the maximal stable energy basins from the extracted topological tree by introducing energy landscape and disconnected graph here. To describe the features, we apply an affine invariant regularization to the extracted regions. We then attach a multiple-scale description to each critical point since there are multiple extracted regions supporting the point. To increase the identification of feature, we design a critical net structure to construct more descriptive features. To apply this algorithm to our problem that 3D shape reconstruction from data distorted by sensor motion, we assume that local data deformation satisfies affine transformations. Details will be described in latter sections.

4.2 Feature Detection

4.2.1 Morse Theory

In this sub-section, we give a brief review of basic definitions and concepts of Morse theory we will use here.

Definition of critical point: given a manifold $\mathcal{M}$, suppose $f : \mathcal{M} \rightarrow R$ is a smooth function. Then $x_0 \in \mathcal{M}$ is critical point of f if and only if $df|_{x_0} \in T^{\star}_{x_0}\mathcal{M}$. The value of $f(x_0)$ is called *critical value*.

A critical point is called *nondegenerate* if its Hessian is nondegenerate.

Definition of Morse function: a smooth function f is called a Morse function if its all critical points are nongenerate.

Morse Lemma: If p_0 is nondegenerate critical point of index λ of a smooth function $f : \mathcal{M} \rightarrow R$, then there exist local coordinates $(x_i)_{1 \leq i \leq m}$ near p_0 such that $x^i(p_0) = 0$, $\forall i$, and in these coordinates we have the equality

$$f = f(p_0) - \sum_{i=1}^{\lambda} (x^i)^2 + \sum_{j=\lambda+1}^{m} (x^j)^2 \tag{4.1}$$

where m is $dim(\mathcal{M})$.

It can be deduced from Morse Lemma that for a Morse function $f : \mathcal{M} \longrightarrow R$ with a critical point x_0, one can perturb f by composing with an isotopy of the manifold $\mathcal{M}$, that the new function has the same critical point x_0.

We assume that the deformation caused by sensor motion satisfies the perturbation mentioned above. In other words, we assume that critical points of Morse function on 3D mesh are kept after the distortion by sensor motion. Based on this assumption, we design a algorithm slicing the 3D mesh by critical values to extract feature regions which are invariant to the deformation.

In the case of 3D data scanned by moving range sensor, the coordinates of objects always change with the movement of sensor. Thus we need to choose appropriate Morse functions which are not affected by the definition of reference frame.

4.2.2 Implicit Polynomial Morse Function

One choice is to utilize implicit polynomial (IP) which fits the object surfaces.

Given a 3D data set, the 3D IP of degree n is defined as:

$$f_n(\mathbf{x}) = \sum_{0 \le i,j,k; i+j+k \le n} a_{ijk} x^i y^j z^k \tag{4.2}$$

where $\mathbf{x} = (x, y, z)$ is one point in the data set.

An IP fitting of 3D surface is to find an IP whose zero set $\{\mathbf{x} | f_n(\mathbf{x} = 0)\}$ can represent the given 3D data set best.

Ideally, it means:

$$f_n(\mathbf{x}) = \sum_{0 \le i,j,k; i+j+k \le n} a_{ijk} x^i y^j z^k = 0 \tag{4.3}$$

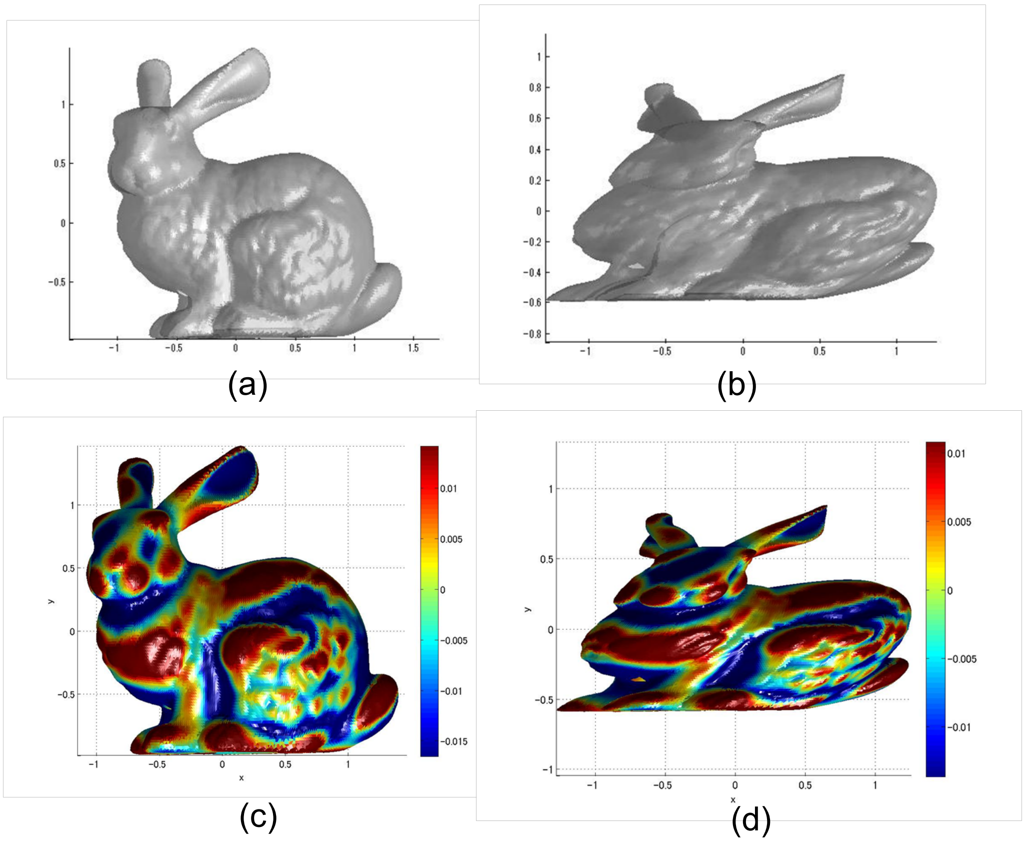


Figure 4.1: Stanford bunny colored by IP values. (a) Original model. (b) The one distorted by an affine transformation. (c) Original model colored by its IP values. (d) Distorted model colored by its IP values.

Technically, it can be considered as the IP which minimizes the square error:

$$\sum (f_n(\mathbf{x}))^2 \tag{4.4}$$

The degree n can be adaptively determined by adopting a fast adaptive fitting method developed by Zheng [71]. It achieves $O(Nk)$ complexity where N is the number of data points and k is the number of polynomial coefficients. Denote the found IP as $\tilde{f}_n(\cdot)$:

$$\sum (\tilde{f}_n(\mathbf{x}))^2 = \min_{f_n(\cdot)} \sum (f_n(\mathbf{x}))^2 \tag{4.5}$$

According to [71], the value of IP fitting won't be affected by the change of coordinate system. Another advantage is that IP is invariant to affine transformation. As shown in Fig. 4.1, left top figure is the original Stanford bunny model and the left bottom one is the model colored by the value of its IP function $\tilde{f}_n(\mathbf{x})$. The right top figure is the one distorted by an affine transformation and the right bottom is the one colored by its IP value. We can see that the corresponding parts on the two models share the same color distributions.

IP is a global fitting method where all points are involved in the calculation. It works well on the data sets which contain the same parts of objects. But in the case of large change of the observed targets, like view point change or occlusion, it will be affected a lot by the data change.

4.2.3 Beta-stable Laplacian Morse Function

To overcome the problem of IP in the case of data change, we design another Morse function based on Laplacian smoothing.

Laplacian is a differential operator defined as:

$$\Delta f = \nabla^2 f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} \tag{4.6}$$

Laplacian smoothing is one of the common methods of mesh smoothing. The basic idea is moving the vertices of mesh incrementally in the direction of the Laplacian. It can be written as a weighted sum of the one-ring neighbors of a vertex, which is actually a discrete approximation of Laplacian [10]:

$$L(x_i) = \sum_{j \in N_1(i)} \omega_{ij}(x_j - x_i) \tag{4.7}$$

One of the advantages of selecting Laplacian smoothing for our Morse function is that it uses relative positions of neighborhood vertices, which don't change with the reference frame. Another advantage is that the computational complexity of Laplacian smoothing is linear in time and space which is more practical on large meshes.

Different from the conventional Laplacian smoothing, we use the difference of the vertex and the one after smoothing as the Morse function:

$$I_k = ||L_{k+1}(x) - L_k(x)|| \cdot sign(n \cdot (L_{k+1}(x) - L_k(x))) \quad (4.8)$$

where k is the iteration times of Laplacian, n is the surface normal at point x.

Geometrically, this Morse function describes the local shape of objects. It emphasizes the parts where the shape changes fast. The more iterations of Laplacian done, the flatter the function would be, and more small changes on the shape are filtered. To determine an appropriate number of k, inspired by the concept of feature's stability in [20], we define a Beta-stable Laplacian Morse function.

Let N_k be the norm of I_k:

$$N_k = ||I_k|| = \sum ||L_{k+1}(x_i) - L_k(x_i)|| \quad (4.9)$$

The variation speed δ_k of the Morse function at k is:

$$\delta_k = N_{k+1} - N_k \quad (4.10)$$

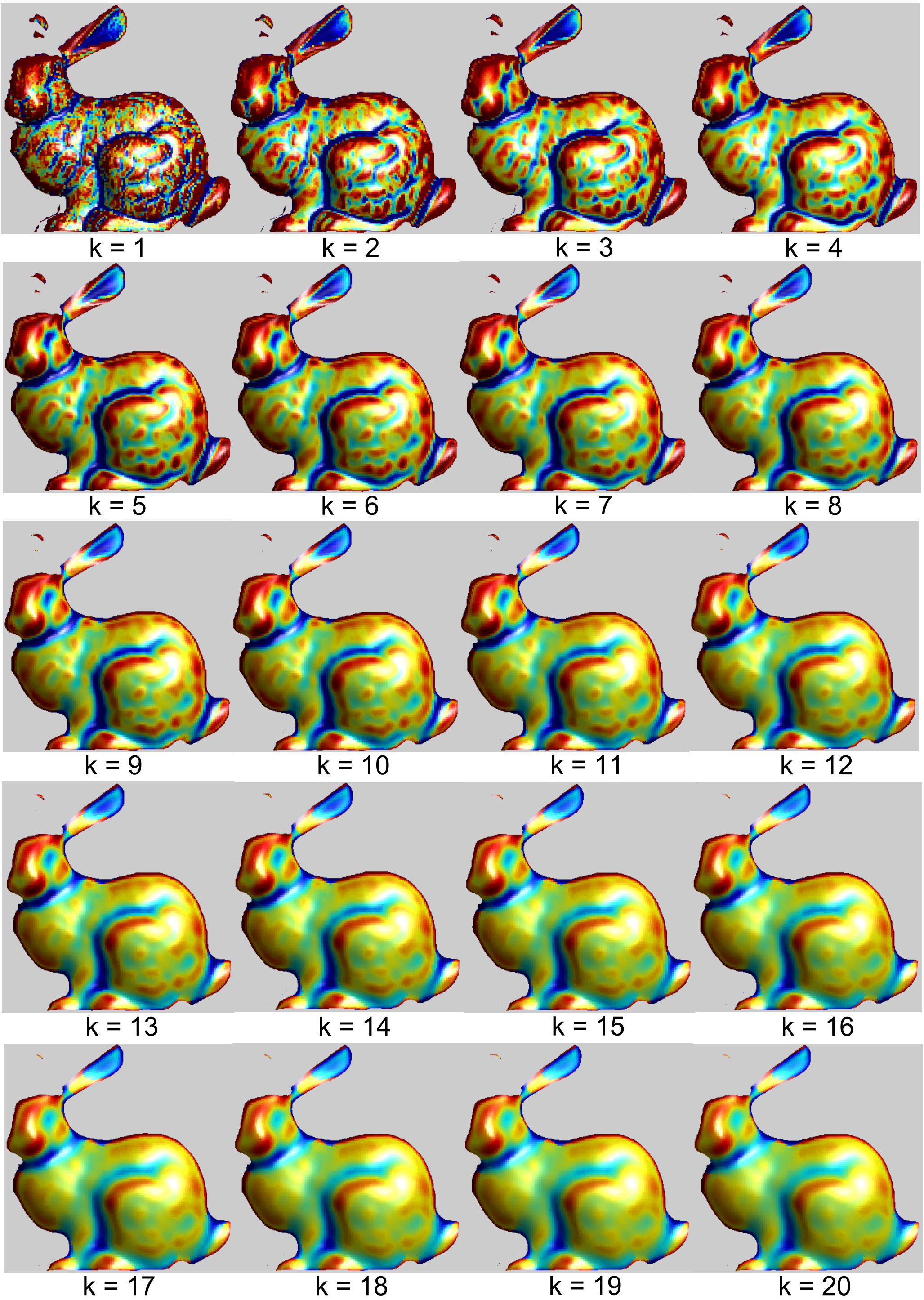
When δ_k is far away from zero, we say that the Morse function is not stable. A small change of k will lead to a significant change of N_k. From the point of view of geometric shape, it means a small hump on the surface also has large Morse function value, which is undesired. In contrast, when $\delta_k \approx 0$, we say that I_k is stable.

I_k is β-stable if k is the smallest integer for which $\delta_\epsilon = 0$ for all $\epsilon \in [k, k + \beta]$.

Fig. 4.2 shows the value of Laplacian Morse function changing with the iteration times k. Fig. 4.3 shows the variation speed δ_k versus k. It can be seen that the Morse function gets stable when $k \geq 15$. Fig. 4.4 shows the final Morse function value with $k = 15$ on the original 3D mesh.

4.2.4 Maximal Stable Energy Basin

According to Morse theory, the topology of a smooth manifold is very closely related to the critical points of the Morse function defined on the



k = 1 k = 2 k = 3 k = 4

k = 5 k = 6 k = 7 k = 8

k = 9 k = 10 k = 11 k = 12

k = 13 k = 14 k = 15 k = 16

k = 17 k = 18 k = 19 k = 20

Figure 4.2: Difference of Laplacian on Stanford bunny model

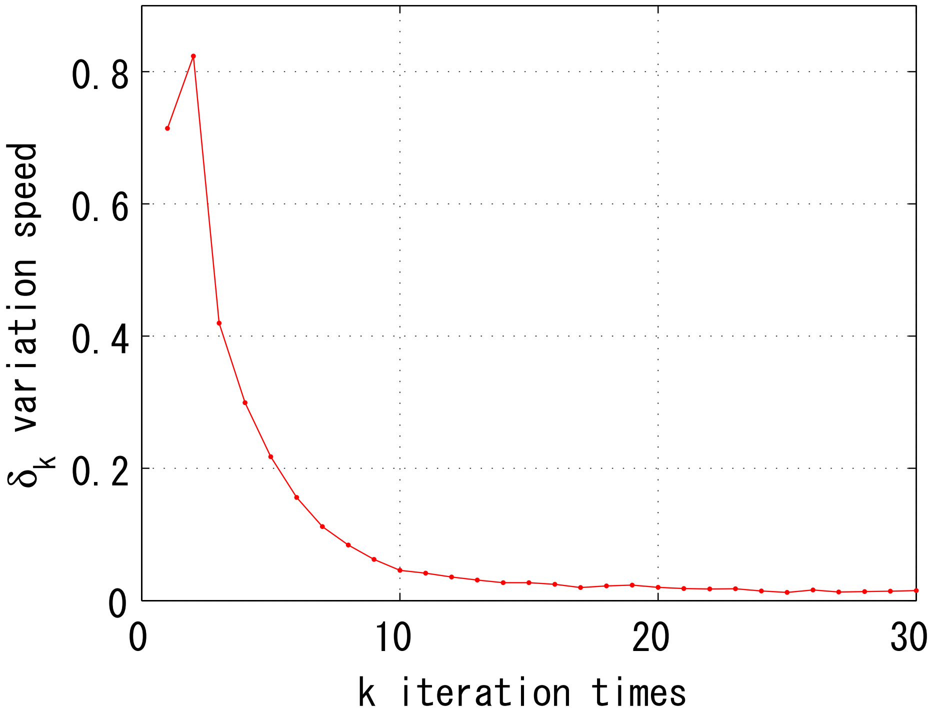


Figure 4.3: Shows the speed δ_k versus the iteration times k of Laplacian.

manifold. Given a smooth manifold M, sublevel set M^t is defined as:

$$M^t = \{x \in M; f(x) \leq t\} \tag{4.11}$$

The changes in the topology of M^t is an indicator of the presence of a critical point.

Our main interest in using this property is to slice the the manifold based on critical points.

3D mesh can be considered as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the vertices and $\mathcal{E}$ is referring to the undirected edges connecting the vertices.

First we define a connected component C in 3D mesh as a subgraph of $\mathcal{G}$, in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in $\mathcal{G}$.

Assume the boundary of the given Morse function $f(x)$ is $[a, b]$. When

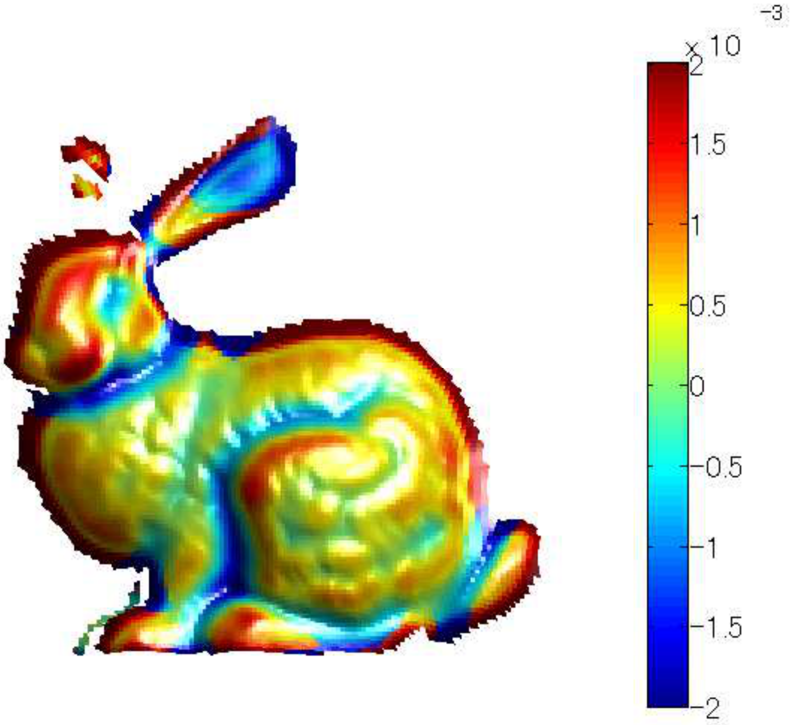


Figure 4.4: Laplacian Morse function with $k = 15$ iterations.

the value t gradually changes from a to b, the connectivities between local minimas in the sublevel set M^t change at critical values. As shown in Fig. 4.5, P_2 and P_3 is not connected until $t = f(p_4)$ and P_1 is not connected to P_2 and P_3 until $t = f(p_5)$.

We thus can determine the set of critical values $\{f^P\}$ for key point P by checking the change of connectivities of local minimas in M^t, like $f(P_4)$ and $f(P_5)$ in Fig. 4.5.

We use local minimas as key points. The support regions of local minima P, denoted as SR_P, is defined as the connected components containing P in the set of M^{f^P}. As shown in Fig. 4.5, the support regions of P_2 are the green and blue parts, and the support regions of P_3 are the pink and blue parts.

The change of connectivities between local minimas can be visualized

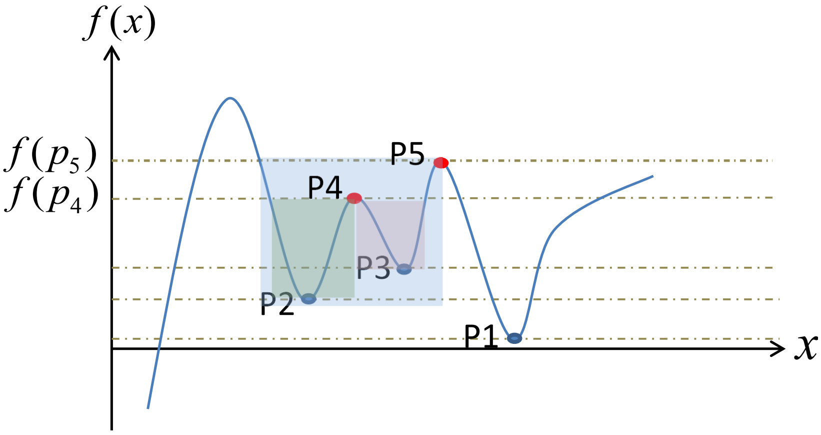


Figure 4.5: Extract support regions from Morse function based on critical points.

using a disconnectivity graph [11]. This idea is from the field of physics used for visualizing potential energy. As shown in Fig. 4.6, all local minimas are leaves nodes represented by blue circles. The critical values f^P where connectivities change are branch nodes represented by yellow triangles. Actually the critical values f^P of key point P are all branch nodes on the path from leaf P to the root.

We adopt the depth first search algorithm to calculate the connected components in M^t. The detection of local minima can also be done at the same step. We only need to find the point with the minimal value in the newly appeared connected components while t changes.

The detail of how to extract the feature on 3D mesh is described in Algorithm 4.1. Geometrically the extracted regions look like basins with the bottom at key points. So we name it energy basin extraction algorithm. We only mentioned the support regions around local minimas when we describe the algorithm. Similarly, it only need to take the minus value of Morse function f to get the support regions around local maximas.

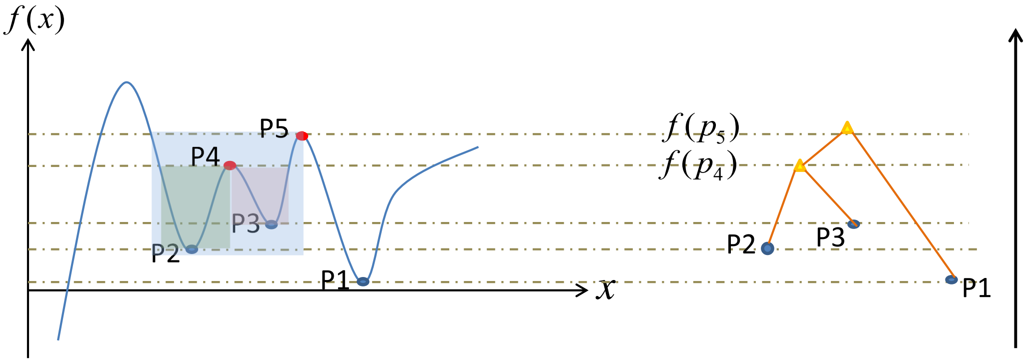


Figure 4.6: Construct a disconnected graph from Morse function based on critical points.

Algorithm 4.1: Energy Basin Extraction

```
Data: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$; $f(\mathcal{V}) \in [a, b]$
Result: a set of key points $\{P_i\}$ with corresponding multi-scale
        support regions $SR_{P_i}$
for $t = a \rightarrow b$ do
    Get the sublevel set $G^t$;
    Check connectivity of $G^t$ and give component label $C$ to $\mathcal{V}$;
    Update the component label $C_t^i$ for each key point $P_i$;
    Update the list of key points $\{P_i\}$;
    if Connectivities between $\{P_i\}$ change then
        Record $t$ as a parent node for connected $P_i$;
        Update the disconnected graph ($\mathcal{DG}$);
    end
end
for $i = 1 \rightarrow |P|$ do
    Get the path $\{t_j^i\}$ from $P_i$ to the root of $\mathcal{DG}$;
    for $j = 1 \rightarrow |t^i|$ do
        Get support region $SR_{P_i}^j$: the connected component
        containing $P_i$ in sublevel $G^{t_j^i}$;
    end
end
```

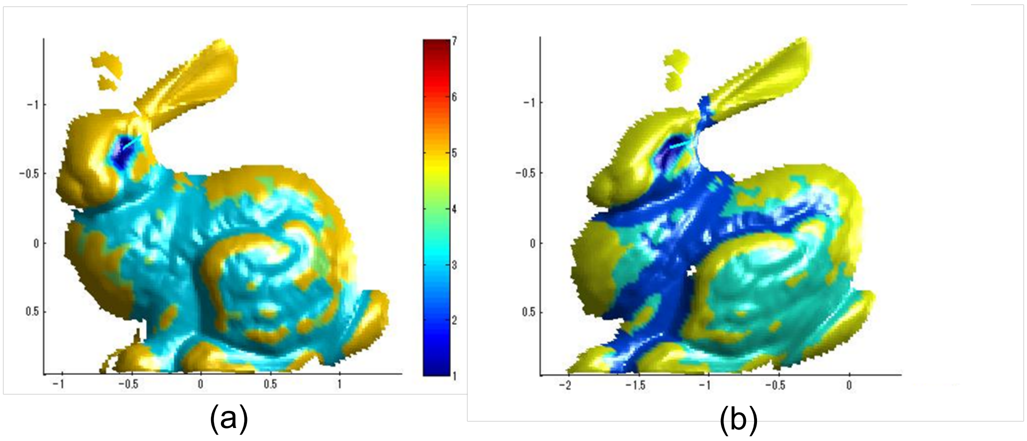


Figure 4.7: Example of multi-scale feature. The key point is the one on the eye denoted by a blue arrow. (a) Support regions extracted from data scanned by static sensor. (b) Support regions extracted from data scanned by moving sensor. Different colors stand for the different scales.

Fig. 4.7 shows an example of the multi-scale feature. The different colors stand for the support regions in different scales. Although they are not all the same in the two data sets because of the change of data, still they share the same region in the small scales. Thus we can find their similarity in different scales.

4.3 Feature Description

By far, we have extracted the multi-scale energy regions cut by critical points. The next step is to find an appropriate way to describe them.

Here we assume that the local deformations between the corresponding regions in range data from moving sensor approximately satisfy affine transformation, which is appropriate in normal moving vehicle. We thus apply an affine normalization to the extracted regions. Since there are multiple regions in different scales for each key point, we construct a multi-scale descriptor made by a set of a basic descriptor. Here we adopt the spin image to be the basic descriptor.

4.3.1 Affine Normalization

We use the shape normalization method based on the inertia matrix normalization presented by Cohignac [15].

The algorithm described in [15] is for 2D image data. Here we analogously expand it to 3D case.

Denote $\mathbb{I}_{\mathcal{F}}$ as the indicator function of a solid shape $\mathcal{F}$, assuming that $\mathcal{F}$ is previously translated so that its barycenter is at the origin of the 3D space. The moment of order (p, q, k) (p, q and k are natural integers) of $\mathcal{F}$ is defined by:

$$\mu_{p,q,k}(\mathcal{F}) = \int_{\mathbb{R}^3} x^p y^q z^k \mathbb{I}_{\mathcal{F}}(x, y, z) dxdydz \tag{4.12}$$

The discrete form for 3D mesh $\mathcal{M}$ can be rewritten as:

$$\mu_{p,q,k}(\mathcal{M}) = \sum_{\mathbb{R}^3} x^p y^q z^k \tag{4.13}$$
$$\text{where } (x, y, z) \in \mathcal{M}$$

Let $\mathcal{S}_{\mathcal{F}}$ be the following 3×3 positive-definite, symmetric matrix:

$$\mathcal{S}_{\mathcal{F}} = \frac{1}{\mu_{0,0,0}} \begin{bmatrix} \mu_{2,0,0} & \mu_{1,1,0} & \mu_{1,0,1} \\ \mu_{1,1,0} & \mu_{0,2,0} & \mu_{0,1,1} \\ \mu_{1,0,1} & \mu_{0,1,1} & \mu_{0,0,2} \end{bmatrix} \tag{4.14}$$

where $\mu_{i,j,k} = \mu_{i,j,k}(\mathcal{M})$. According to the uniqueness of Cholesky factorization, the decomposition of $\mathcal{S}_{\mathcal{F}}$: $\mathcal{S}_{\mathcal{F}} = B_{\mathcal{F}} B_{\mathcal{F}}^T$ may be unique, where $B_{\mathcal{F}}$ is a lower-triangular real matrix with positive diagonal entries.

Let $\mathcal{A}$ be a non-singular 3×3 matrix. The normalized shape associated to $\mathcal{F}$ is the shape $\mathcal{F}' = B_{\mathcal{F}}^{-1}(\mathcal{F})$. It can be proved that the normalized shape $\mathcal{F}'$ is invariant to affine transformation $\mathcal{A}$, up to a rotation $\mathcal{Q}$. (See details in [15])

4.3.2 Multi-scale Description

Since our normalized shapes have rotation change varying with the coordinate system, invariance to rotation should be a desired property of the feature descriptor. Spin image is chosen as the basic descriptor for its rotation invariance and fast computation. We propose an variation algorithm of spin image.

First we simply review the original spin image algorithm proposed by Andrew E. Johnson [27]. A spin image is created for an oriented vertex on 3D mesh surface. First a partial, object-centered coordinate system is defined with respect to an oriented point: the radial coordinate α, defined as the distance to the surface normal at the point, and the elevation coordinate β, defined as the signed distance to the tangent plane at the point. A 2-D accumulator indexed by α and β is incremented in the surface mesh which is within the support region of this vertex.

To use spin image in our algorithm, for each support region SR_P^j belonging to the key point P, a spin image is calculated. Since the size of each support region SR_P^j is different, the coordinates range of α and β of the spin image is set as the maximal distance along each coordinate axis from the key point to the furthest point in the region. The resolution of bins in the accumulator is set the same for all regions.

Given the spin image SI_j for each support region SR_P^j, We construct our descriptor of the key point P as:

$$D(P) = \{SI_1, SI_2, \cdots, SI_N\} \tag{4.15}$$

where $N = |SR_P|$ is the cardinality of the set $\{SR_P\}$.

The similarity of two key points P and Q is defined as the minimal pairwise distance of spin images between their supporters:

$$d(P,Q) = \min_{SI_i \in \{D(P)\}, SI_j \in \{D(Q)\}} ||SI_i - SI_j|| \tag{4.16}$$

This descriptor compares features in different scales and is more robust to occlusions and the change of view points.

4.4 Matching Criteria and Evaluation

Consider two 3D meshes $\mathcal{M}$, $\mathcal{N}$ to be matched. Let $K^{\mathcal{M}}$ and $K^{\mathcal{N}}$ be the detected key points set of $\mathcal{M}$ and $\mathcal{N}$ respectively. As described in Eq. 4.16, let $d(p, q)$ be the similarity of two key points p and q.

We say key point $p \in K^{\mathcal{M}}$ is matched to q_1:

$$q_1 = \arg\min_{q \in K^{\mathcal{N}}} d(p, q) \tag{4.17}$$

$$\text{if } \frac{d(p, q_2)}{d(p, q_1)} > 1.5, \text{ where } q_2 = \arg \min_{q \in K^{\mathcal{N}} \setminus q_1} d(p, q)$$

It means if the best match is at least 50% better than the second best match for p, then p has a matched key point.

To evaluate the performance of the proposed algorithm, we use *accuracy* and *repeatability* [52]:

$$\text{Accuracy} = \frac{\text{correct matches found in } \mathcal{M} \text{ and } \mathcal{N}}{\text{total matches found in } \mathcal{M} \text{ and } \mathcal{N}} \tag{4.18}$$

$$\text{Absolute Repeatability} = \text{repeatable key points} \tag{4.19}$$

$$\text{Relative Repeatability} = \frac{\text{repeatable key points}}{\min\{|K^{\mathcal{M}}|, |K^{\mathcal{N}}|\}} \tag{4.20}$$

A key point $p \in K^{\mathcal{M}}$ is said to be *repeatable* if the distance from its nearest neighbor, a key point $q \in K^{\mathcal{N}}$, after transformed according to the ground truth transformation matrix G_T, is less than a threshold ϵ:

$$||p - G_T q|| \leq \epsilon \tag{4.21}$$

4.5 Experiment and Discussion

To verify and evaluate the proposed 3D corresponding extraction algorithm, we use models from the Stanford 3D Scanning Repository [47] and AIM@Shape Shape Repository [46]. First test is extracting correspondence between original model and the one distorted by affine transformations. The second test is between the static data and the synthesized data gathered by virtual moving sensor.

Fig. 4.8 and Fig. 4.9 are examples of extracted feature regions and their affine normalizations. To show the performance of affine normalization, only a single-scale corresponding region is shown for one key point. As shown in the sub figures (a) and (b), the exact corresponding region is extracted from the distorted data. After the affine normalization, the normalized regions have the same shapes, up to a rotation.

4.5.1 Evaluation on Affine Distortion

To verify the performance of proposed algorithm to affine distortion, a series set of affine transformations are applied to original object models.

As shown in Fig. 4.10 and Fig. 4.11, we apply different affine transformation to the bunny model and compare the extracted feature points with the original model. The results show that although beta-stable Laplacian Morse function can detect feature points twice than IP Morse function, IP is more stable on different distortions.

We apply the same set of affine transformations to other models. The repeatability curves using IP Morse function and beta-stable Morse function are shown in Fig. 4.12 and Fig. 4.13, respectively. The results show that IP does have a more stable performance than Laplacian in the affine distortion. Fig. 4.14 (a) and (b) show the matching accuracy of using IP and Laplacian, respectively. Fig. 4.15 and Fig. 4.15 show the matching result of using IP and Laplacian on different models with the same affine transformation, respectively.

4.5.2 Evaluation on Synthetic Data from Moving Range Sensor

Fig. 4.17 and Fig. 4.18 are the matching results between the synthetic static data and distorted data from moving sensor with different velocities, using beta-stable Laplacian Morse function.

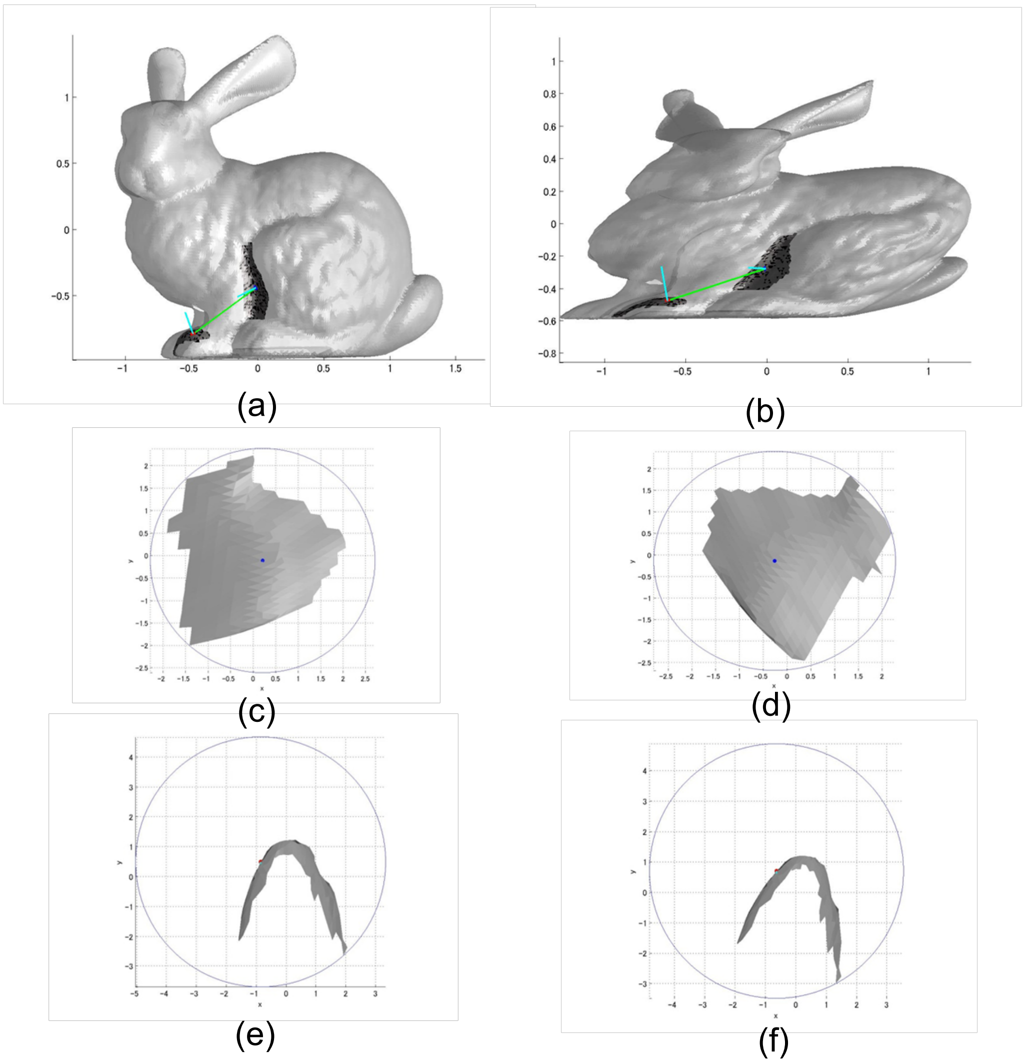
4.6 Summary

In this chapter, we propose a novel algorithm of extraction of 3D shape correspondences. The basic idea is utilizing Morse theory to extract topological information from information about critical points of a function.

We design two types of Morse function for different conditions. The one based on implicit polynomial fitting of 3D mesh is more robust to data noise and suitable for sparse data set. The other one based on difference of Laplacian has the advantages to the problem of data changes, like occlusion or the change of view point and scene.

We then find the maximal stable energy basins from the extracted topological tree by introducing energy landscape and disconnected graph here. To describe the features, we apply an affine invariant normalization to the extracted regions. We then attach a multiple-scale description based on spin image to each critical point thus their similarities are compared in different scales.

The simulation experiments prove the verification and robustness of the proposed algorithm, especially for the distorted data gathered from moving sensor.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.8: Example of extracted feature regions and their affine normalizations. (a) Two key points with their support regions extracted from the original data. (b) Corresponding key points with their support regions extracted from the data distorted by affine transformations. (c) and (e) are the normalized regions from the original data. (d) and (f) are the corresponding normalized region from distorted data.

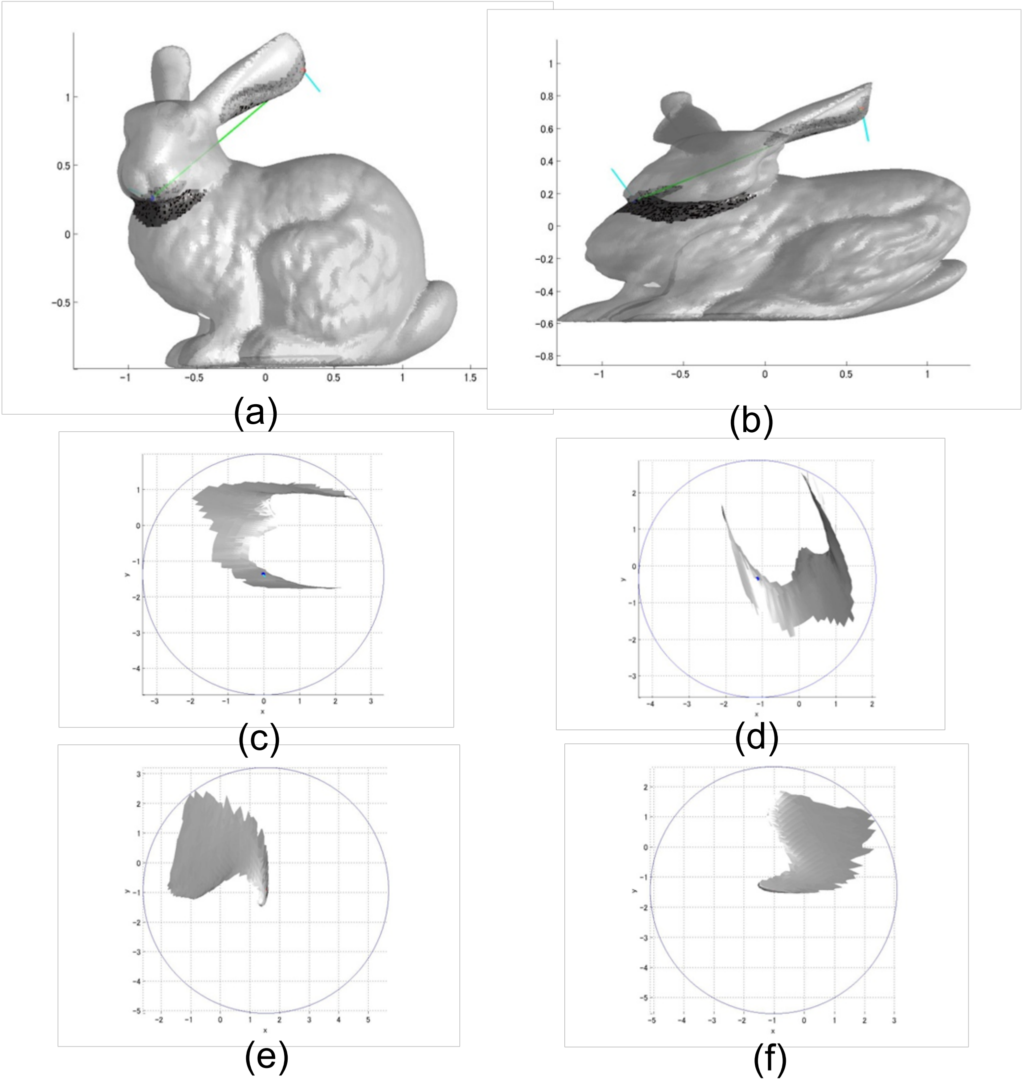


Figure 4.9: Another example of extracted feature regions and their normalizations. (a) Two key points with their support regions extracted from the original data. (b) Corresponding key points with their support regions extracted from the data distorted by affine transformations. (c) and (e) are the normalized regions from the original data. (d) and (f) are the corresponding normalized region from distorted data.

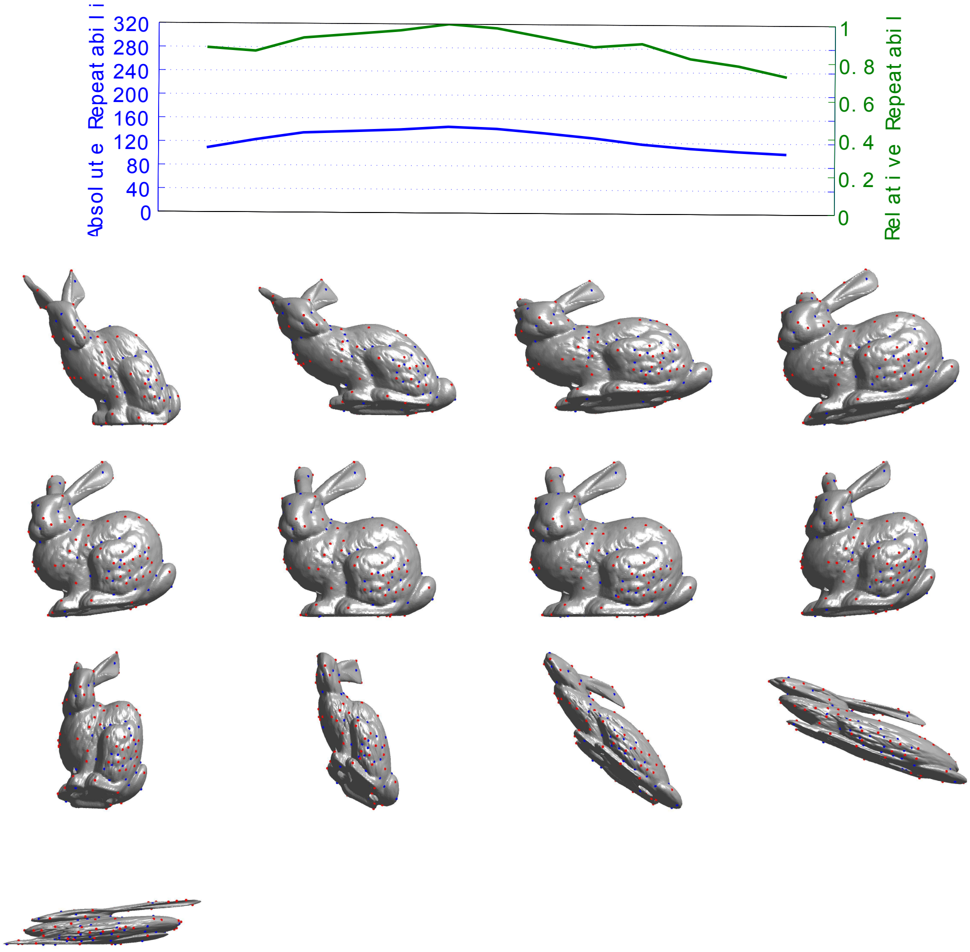


Figure 4.10: Repeatability curve of Stanford bunny model applied a set of affine distortion, using IP Morse function.

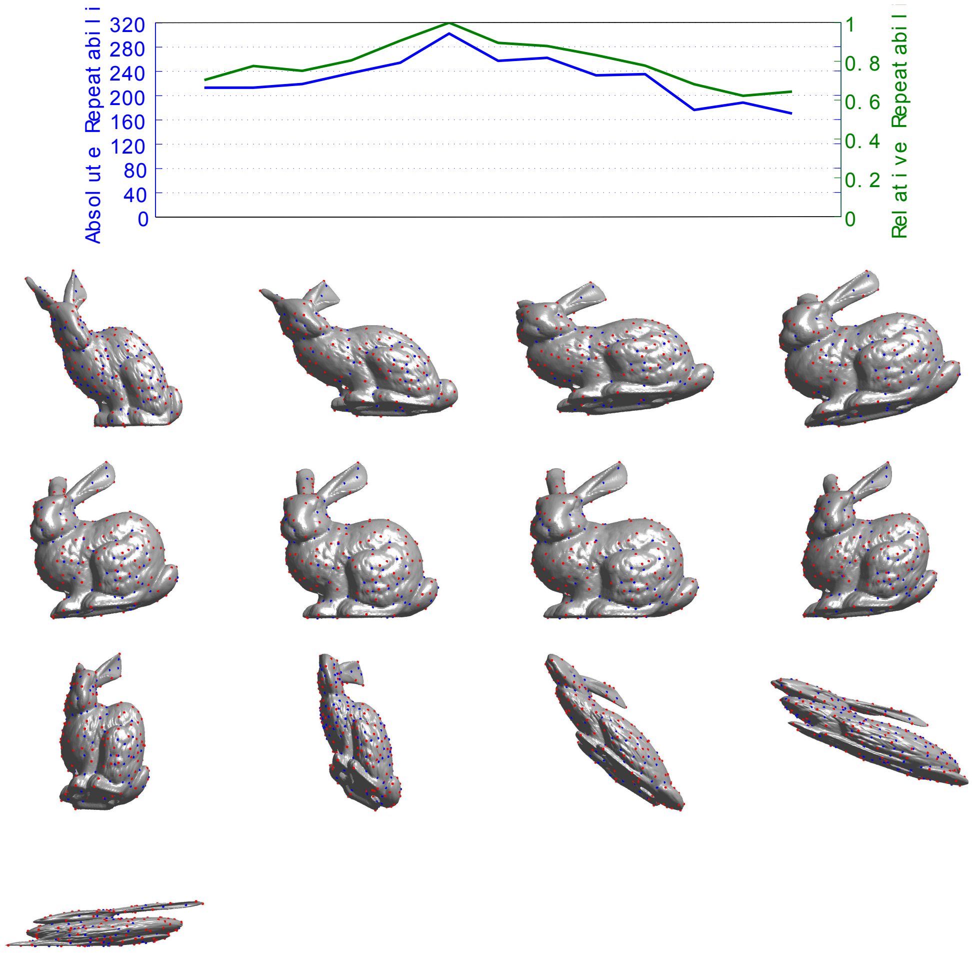
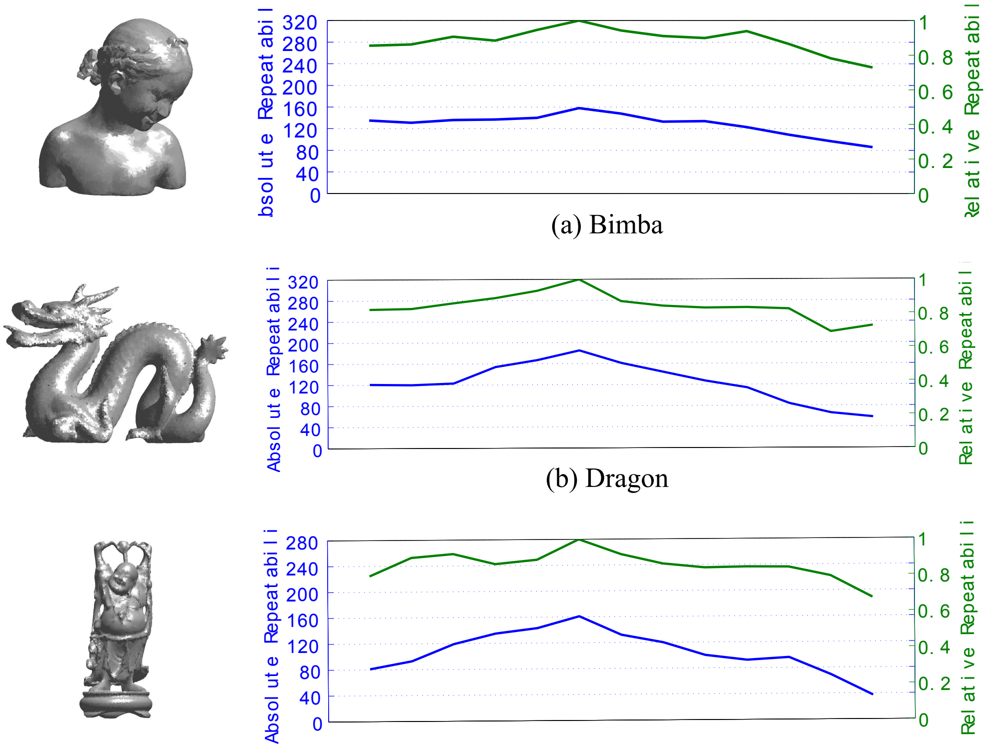


Figure 4.11: Repeatability curve of Stanford bunny model applied a set of affine distortion, using beta-stable Laplacian Morse function.

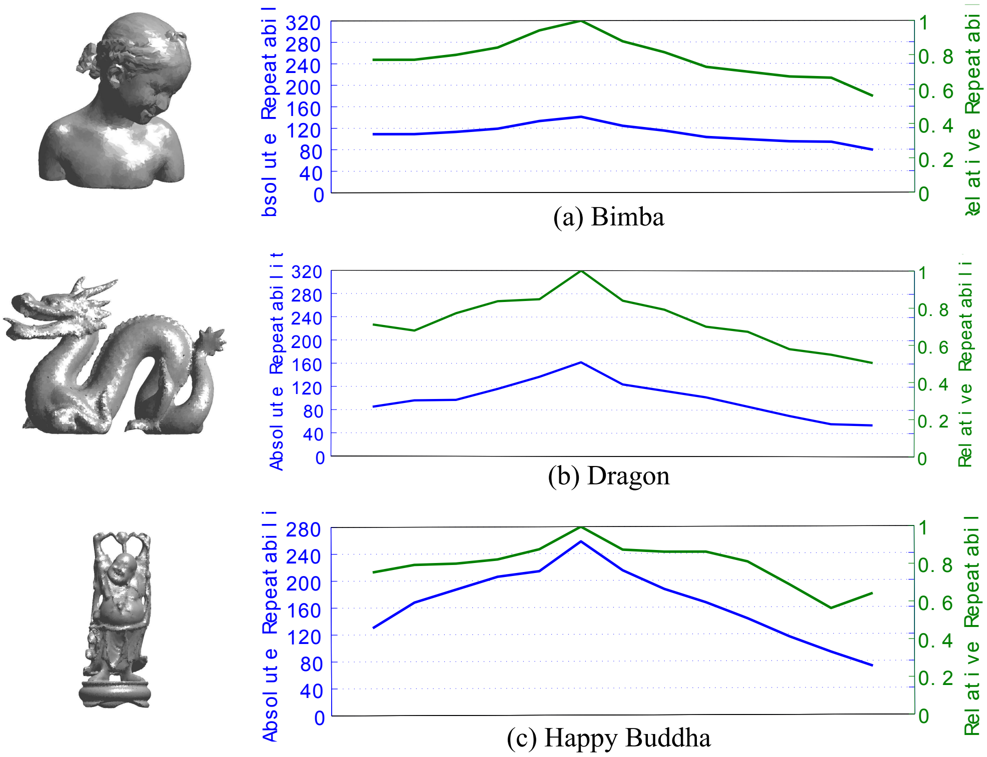


(a) Bimba

(b) Dragon

(c) Happy Buddha

Figure 4.12: Repeatability curve of other models applied the same set of affine distortion, using IP Morse function.

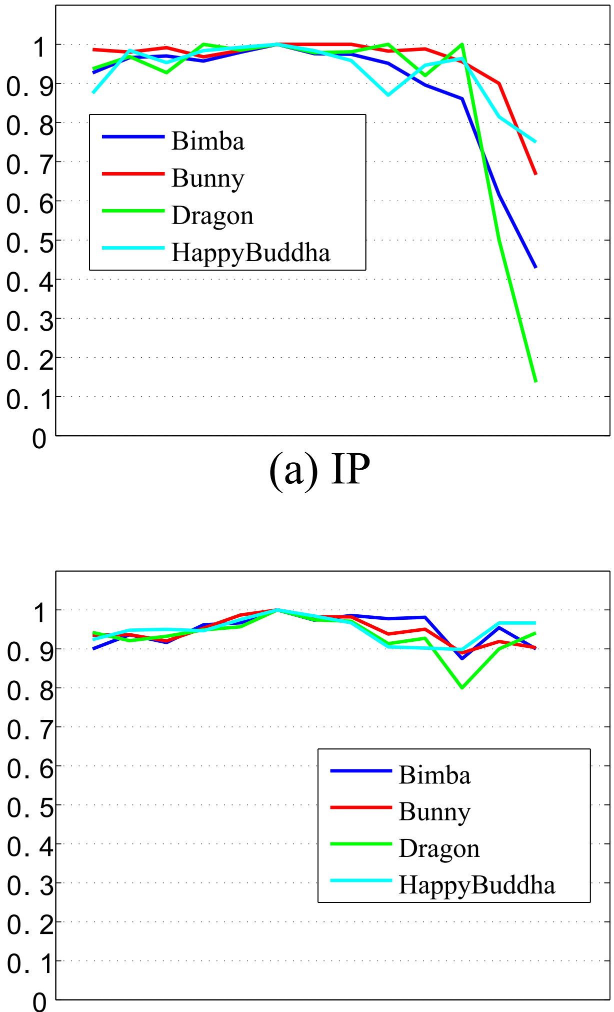


(a) Bimba

(b) Dragon

(c) Happy Buddha

Figure 4.13: Repeatability curve of other models applied the same set of affine distortion, using beta-stable Laplacian Morse function.



(a) IP

(b) Laplacian

Figure 4.14: (a) Match accuracy using IP Morse function. (b) Match accuracy using beta-stable Laplacian Morse function.

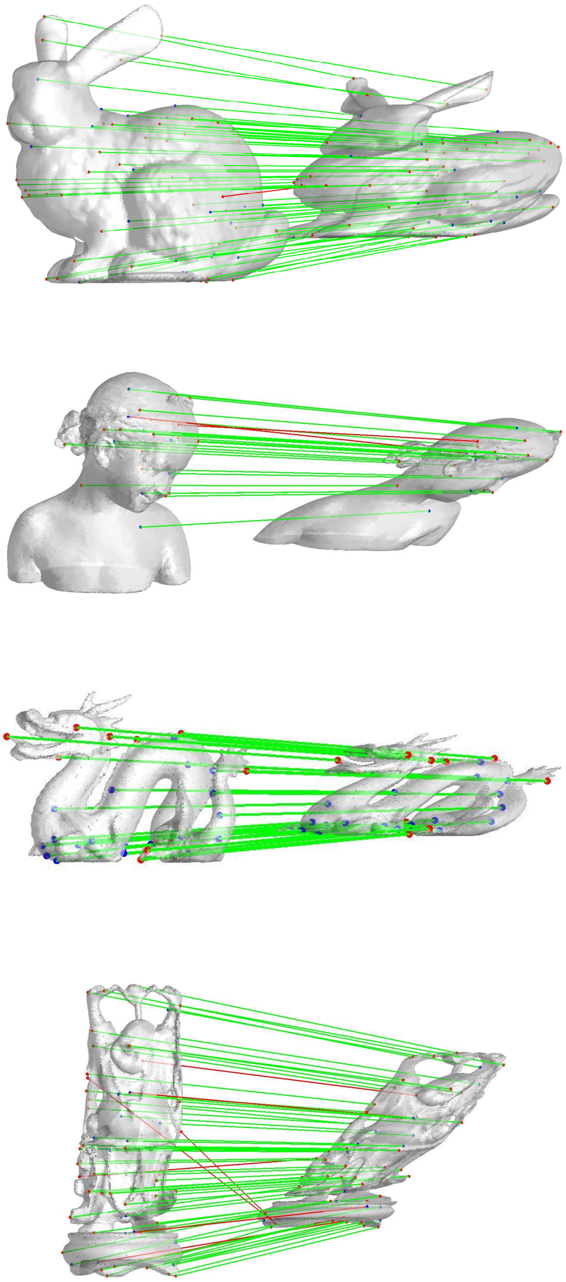


Figure 4.15: One group of matching result of the ones using IP Morse function.

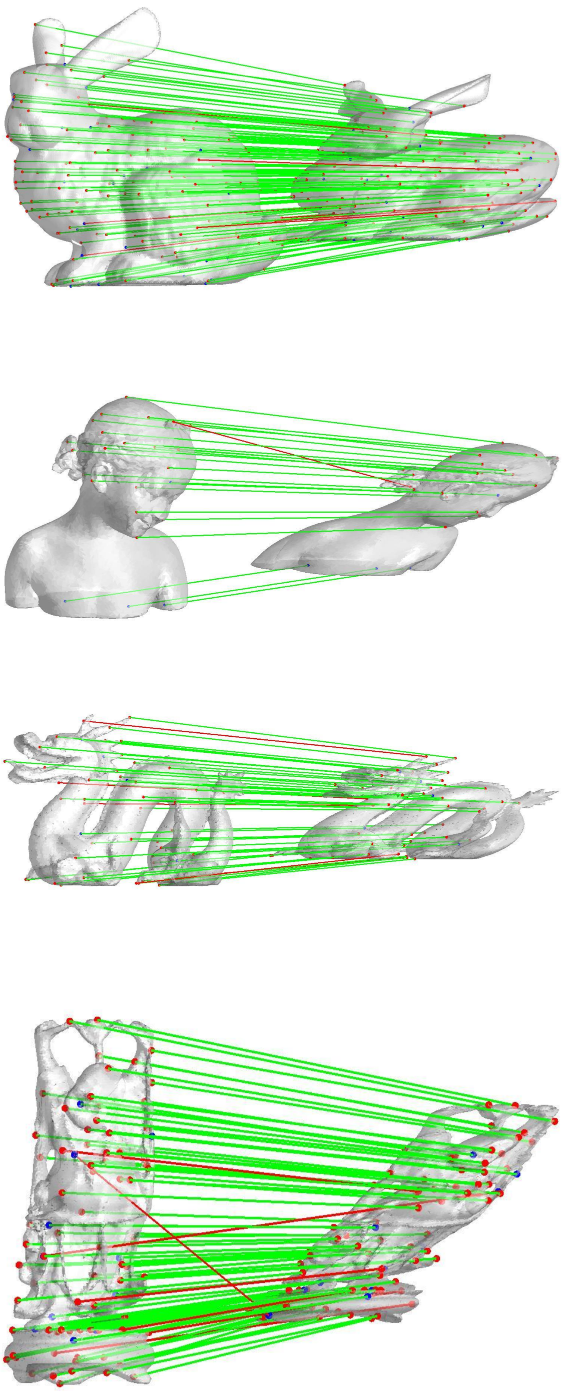


Figure 4.16: One group of matching result of the ones using beta-stable Laplacian Morse function.

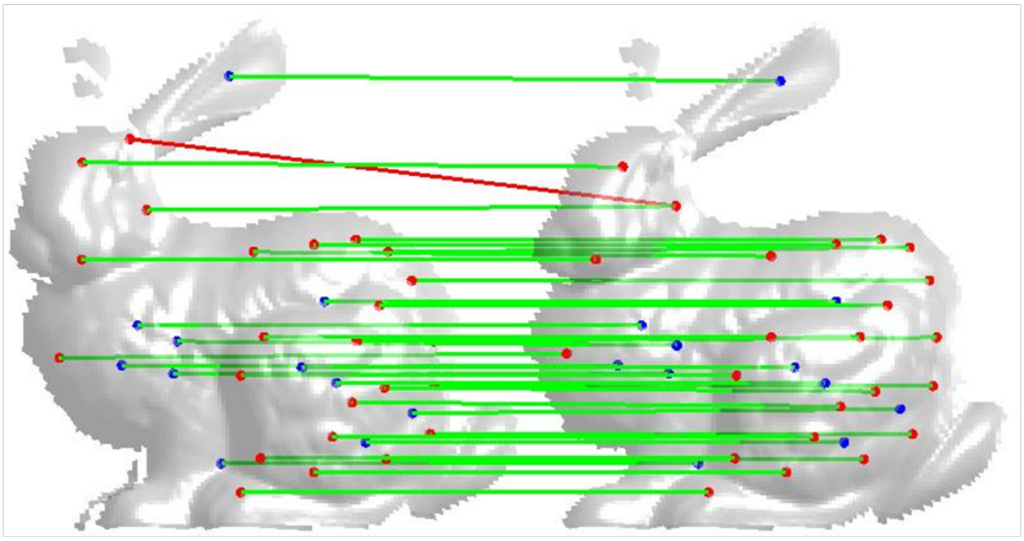


Figure 4.17: Correspondence between synthetic static data and constant velocity (C_1) data using our proposed algorithm.

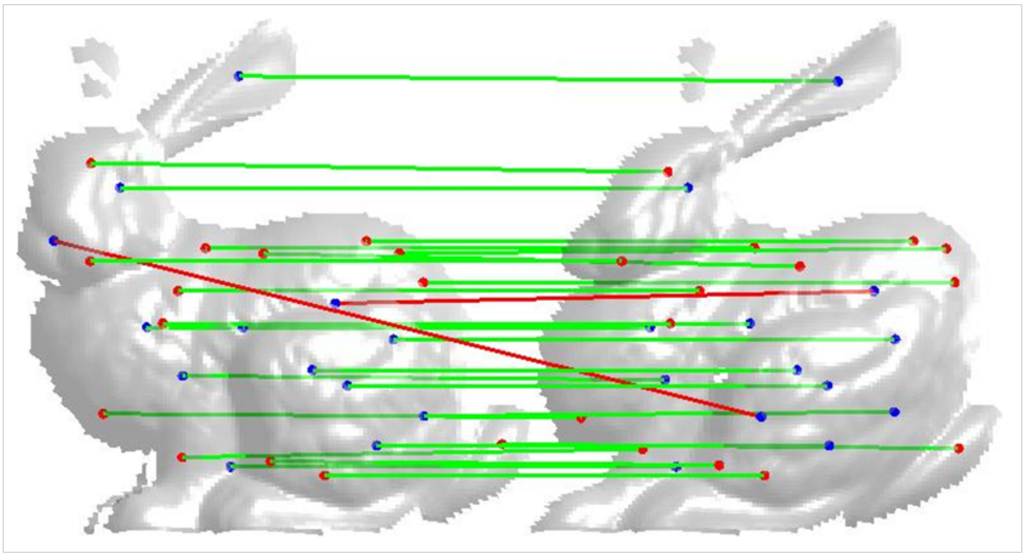


Figure 4.18: Correspondence between synthetic static data and constant velocity (C_2) data using our proposed algorithm.

Chapter 5

Conclusions

5.1 Summary

- A real-time pile driver positioning system using laser range finder We propose a method which utilizes the prior shape model to estimate sensor motion. Based on the estimated motion model, distorted measurements can be rectified and accurate localization of targets can be achieved. As an application, we build a real-time pile driver positioning system using laser range finder. The detail is described is Chapter 2.

- 3D shape reconstruction using a moving range sensor based on estimation of sensor motion — We propose a feature based polynomial fitting method to estimate 6 DOF motion parameters of moving range sensor. The 3D model can be reconstructed according to a highly accurate sensor motion estimation. When using a 2D scanning sensor instead of a 1D scanning sensor, the same region is measured in multiple times when the sensor moves. We will show that we can reconstruct the sensor motion and the scene from only the measured coordinates and times of the same set of points. This reconstruction is intrinsic, which relies on only the intrinsic properties of the distortion, and not relying on the extrinsic information from other sensors. Firstly, for robust estimating sensor movement, we model the sensor motion using polynomial with respect to time. Secondly, to estimate the parameters of sensor motion model, we utilize the 3D corresponding points extracted from the overlapped parts between consecutive frames. This method doesn't need the secondary sensor and is not limited with specific environment features. Without linearization of constraint and discretization of trajectory, distorted data is accurately rectified. The detail is described is Chapter 3.

- An 3D affine invariant shape feature
  To obtain the corresponding constraints described in chapter 3, we propose a novel 3D affine invariant feature detection and matching method which is designed for the deformed 3D data collected by moving range sensor. The basic idea is utilizing Morse theory to extract topological information from information about critical points of a function. We design two types of Morse function for different conditions. The one based on implicit polynomial fitting of 3D mesh is more robust to data noise and suitable for sparse data set. The other one based on difference of Laplacian has the advantages to the problem of data changes, like occlusion or the change of view point and scene. We then find the maximal stable energy basins from the extracted topological tree by introducing energy landscape and disconnected graph here. To describe the features, we apply an affine invariant normalization to the extracted regions. We then attach a multiple-scale description based on spin image to each critical point thus their similarities are compared in different scales. The simulation experiments prove the verification and robustness of the proposed algorithm, especially for the distorted data gathered from moving sensor.

5.2 Contributions

We have explored the problem of 3D reconstruction under a moving sensing system only with a single range sensor. There are three contributions of the thesis.

1. Development of a real-time pile driver positioning system using laser range finder.
2. Development of a 3D shape reconstruction method using a single moving range sensor.
3. Development of a 3D affine invariant shape feature.

In the first contribution, we propose a highly efficient positioning system using a Laser Range Finder. Over the traditional surveying systems ours is superior to automatically detect the position of the pile or pile driver in real time with high accuracy. we first develop LRF based surveying system to scan the construction site in real time and gather the 2D laser point data. Then we detect target object such as pile or pile driver by fast fitting a circle-like geometric model to the data based on Maximum Likelihood Estimation (MLE) inference. The performance of the algorithm is validated by both synthesized and real data set. The results demonstrate the potentials on feasibility of our method in future construction field.

In the second contribution, we propose a method of 3D reconstruction under a moving sensing system only with a single range sensor. Different from ″stop-scan-go″ strategy, our method works in an efficiently continuous manner. There is no need to stop the platform to obtain an stationary scan. Sensor can move under a reasonable motion mode and simultaneously scan the target or environment. Data distortion caused by continuous movement will be rectified. This continuous manner is much more efficient and attractive in practical applications.

In the third contribution, we propose a novel 3D affine invariant shape feature. It is designed for finding correspondence between distorted data deformed by sensor motion. Due to its multi-scale character, it is robust to data change caused by occlusion or change of view point. The fact that it extracts local affine invariant regions allows the possibility of dealing with significant deformation.

5.3 Future Directions

- Shape structure analysis using Morse theory
  To enhance the identification of feature, it is common to make a combination of features. Considering the structure of disconnected graph created from the Morse function on 3D mesh, it's possible to construct a more descriptive feature which is made by a set of critical points. As mentioned before, $\mathcal{DG}$ is generated by critical points and their connectivities relations. A sub-graph of $\mathcal{DG}$ which contains more than one critical point, can apply a more static feature.

- Deformation measurement for comparing similarity between objects
  Comparing similarity between a group of similar objects is an important task in a lot applications. Since our proposed shape feature actually detect local regions whose deformation satisfies affine transformation between objects. Those local transformation can be calculated and used as a deformation measurement of similarity between objects.

References

[1] Peter Abeles. Robust local localization for indoor environments with uneven floors and inaccurate maps. In *Intelligent Robots and Systems (IROS)*, pages 475 –481, sept. 2011.

[2] I. Abuhadrous, S. Ammoun, F. Nashashibi, F. Goulette, and C. Laurgeau. Digitizing and 3d modeling of urban environments and roads using vehicle-borne laser scanner system. In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 1, pages 76 – 81 vol.1, sept.-2 oct. 2004.

[3] S. Baker, E. Bennett, Sing Bing Kang, and R. Szeliski. Removing rolling shutter wobble. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2392 –2399, june 2010.

[4] A. Banno, T. Masuda, T. Oishi, and K. Ikeuchi. Flying laser range sensor for large-scale site-modeling and its applications in bayon digital archival project. *International Journal of Computer Vision*, 78:207–222, 2008. 10.1007/s11263-007-0104-6.

[5] Atsuhiko Banno and Katsushi Ikeuchi. Determination of motion parameters of a moving range sensor approximated by polynomials for rectification of distorted 3d data. *Machine Vision and Applications*, 22:889–897, 2011. 10.1007/s00138-010-0318-4.

[6] A. Birk, N. Vaskevicius, K. Pathak, S. Schwertfeger, J. Poppinga, and H. Buelow. 3-d perception and modeling. *Robotics Automation Magazine, IEEE*, 16(4):53 –60, december 2009.

[7] Michael Bosse and Robert Zlot. Continuous 3d scan-matching with a spinning 2d laser. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 4312 –4319, may 2009.

[8] Haili Chui and Anand Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2a? " 3):114 – 141, 2003. ¡ce:title¿Nonrigid Image Registration¡/ce:title¿.

[9] S. Das, C. Gleason, S. Shen, S. Goddard, and L.C. Perez. 2d tracking performance evaluation using the cricket location-support system. In *Electro Information Technology, 2005 IEEE International Conference on*, pages 6 pp. –6, may 2005.

[10] Mathieu Desbrun, Mark Meyer, Peter Schröder, and Alan H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 317–324, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[11] Jonathan P. K. Doye, Mark A. Miller, and David J. Wales. Evolution of the potential energy surface with size for lennard-jones clusters. *The Journal of Chemical Physics*, 111(18):8417–8428, 1999.

[12] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, January 1972.

[13] M. Dunbabin and Peter Corke. Autonomous excavation using a rope shovel. *Journal of Field Robotics*, 23(6-7):379–394, 2006.

[14] Jani Even, Panikos Heracleous, Carlos Ishi, and Norihiro Hagita. Multi-modal front-end for speaker activity detection in small meetings. In *Intelligent Robots and Systems (IROS)*, pages 536 –541, sept. 2011.

[15] Cao. F, Lisani. J, Muse. J, Morel. P, and Sur F. *A Theory of Shape Identification*. Springer Verlag, 2008.

[16] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[17] P. Forssen and E. Ringaby. Rectifying rolling shutter video from hand-held devices. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 507 –514, june 2010.

[18] K. Fujiwara, K. Nishino, J. Takamatsu, Bo Zheng, and K. Ikeuchi. Locally rigid globally non-rigid surface registration. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1527 – 1534, nov. 2011.

[19] Natasha Gelfand, Niloy J. Mitra, Leonidas J. Guibas, and Helmut Pottmann. Robust global registration. In *Proceedings of the third Eurographics symposium on Geometry processing*, SGP '05, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[20] Steve Gu, Ying Zheng, and Carlo Tomasi. Critical nets and beta-stable features for image matching. In *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, ECCV'10, pages 663–676, Berlin, Heidelberg, 2010. Springer-Verlag.

[21] Yanying Gu, A. Lo, and I. Niemegeers. A survey of indoor positioning systems for wireless personal networks. *Communications Surveys Tutorials, IEEE*, 11(1):13 –32, quarter 2009.

[22] Tamura Hajime. Position measurement system using laser range finder in construction fields – application for pile position measurements. *Thesis*, pages 115–116, 2010.

[23] A. Harrison and P. Newman. High quality 3d laser ranging under general vehicle motion. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 7 –12, may 2008.

[24] Xiangqi Huang, T. Sasaki, H. Hashimoto, and F. Inoue. Circle detection and fitting based positioning system using laser range finder. In *System Integration (SII), 2010 IEEE/SICE International Symposium on*, pages 442 –447, dec. 2010.

[25] Xiangqi Huang, T. Sasaki, H. Hashimoto, and F. Inoue. Circle detection and fitting using laser range finder for positioning system. In *Control Automation and Systems (ICCAS), 2010 International Conference on*, pages 1366 –1370, oct. 2010.

[26] Yekeun Jeong, Yunsu Bok, Jun-Sik Kim, and In-So Kweon. Complementation of cameras and lasers for accurate 6d slam: From correspondences to bundle adjustment. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3581 –3588, may 2011.

[27] Andrew Johnson. *Spin-Images: A Representation for 3-D Surface Matching*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.

[28] Timor Kadir, Andrew Zisserman, and Michael Brady. An affine invariant salient region detector. In Tomas Pajdla and Jiri Matas, editors, *Computer Vision - ECCV 2004*, volume 3021 of *Lecture Notes in Computer Science*, pages 228–241. Springer Berlin / Heidelberg, 2004.

[29] S. Kagami, R. Hanai, N. Hatao, and M. Inaba. Outdoor 3d map generation based on planar feature for autonomous vehicle navigation in urban environment. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1526 –1531, oct. 2010.

[30] Kenichi Kanatani. Statistical optimization for geometric fitting: Theoreticalaccuracy bound and high order error analysis. *International Journal of Computer Vision*, 80:167–188, 2008.

[31] Ali H. Kashani, William S. Owen, Nicholas Himmelman, Peter D. Lawrence, and Robert A. Hall. Laser scanner-based end-effector

tracking and joint variable extraction for heavy machinery. *The International Journal of Robotics Research*, 29(10):1338–1352, 2010.

[32] Yaron Lipman and Thomas Funkhouser. Mobius voting for surface correspondence. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), August 2009.

[33] Y.L. Loke, A.A. Gopalai, B.H. Khoo, and S.M.N.A. Senanayake. Smart system for archery using ultrasound sensors. In *Advanced Intelligent Mechatronics, AIM 2009. IEEE/ASME International Conference on*, pages 1160 –1164, july 2009.

[34] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[35] D.G. Lowe. Object recognition from local scale-invariant features. *Proc. IEEE Conf. Int. Conf. on Comp. Visi.*, pages 1150–1157, 1999.

[36] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 384–393, London, 2002.

[37] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *Int. J. Comput. Vision*, 60(1):63–86, 2004.

[38] K. Morioka and H. Hashimoto. Appearance based object identification for distributed vision sensors in intelligent space. *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, 1:199 –204, sept. 2004.

[39] Richard M. Murray, S. Shankar Sastry, and Li Zexiang. *A Mathematical Introduction to Robotic Manipulation*, chapter 2. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1994.

[40] A. Myronenko, X. Song, and M. Carreira-Perpinan. Free-form nonrigid image registration using generalized elastic nets. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1 – 8, 2007.

[41] Ryutarou Ohbuchi, Tomo Otagiri, Masatoshi Ibato, and Tsuyoshi Takei. Shape-similarity search of three-dimensional models using parameterized statistics. In *Proc. Pacific Graphics 2002*, pages 265–274, 2002.

[42] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics*, 21:807–832, 2002.

[43] Bayon Project. $http : //www.cvl.iis.u - tokyo.ac.jp/research/bayon/index.shtml$.

[44] M.A. Rana, H. Setan, Z. Majid, and A.K. Chong. Computer assisted surgical planner for craniofacial reconstruction - imaging techniques. In *Geometric Modeling and Imaging–New Trends, 2006*, pages 215 –220, aug. 2006.

[45] D. Raviv, A. M. Bronstein, M. M. Bronstein, R. Kimmel, and N. Sochen. Affine-invariant diffusion geometry for the analysis of deformable 3d shapes. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 2361–2367, Washington, DC, USA, 2011. IEEE Computer Society.

[46] AIM@Shape Shape Repository. $http : //shapes.aimatshape.net/$. Accessed: 07/02/2013.

[47] Stanford 3D Scanning Repository. $http : //graphics.stanford.edu/data/3dscanrep/$. Accessed: 07/02/2013.

[48] Erik Ringaby and Per-Erik Forssén. Scan rectification for structured light range sensors with rolling shutters. In *IEEE International Conference on Computer Vision*, Barcelona, Spain, November 2011. IEEE, IEEE Computer Society.

[49] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145 –152, 2001.

[50] Julian Ryde and Huosheng Hu. 3d mapping with multi-resolution occupied voxel lists. *Autonomous Robots*, 28:169–185, 2010. 10.1007/s10514-009-9158-3.

[51] Y. Sahillioglu and Y. Yemez. 3d shape correspondence by isometry-driven greedy optimization. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 453 –458, june 2010.

[52] S. Salti, F. Tombari, and L.D. Stefano. A performance evaluation of 3d keypoint detectors. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2011 International Conference on*, pages 236 –243, may 2011.

[53] T. Sasaki, Xiangqi Huang, H. Hashimoto, and F. Inoue. Position measurement of piles using a laser range finder for accurate and efficient pile driving. In *Advanced Intelligent Mechatronics (AIM), 2011 IEEE/ASME International Conference on*, pages 241 –246, july 2011.

[54] Avinash Sharma, Radu P. Horaud, Jan Cech, and Edmond Boyer. Topologically-robust 3d shape matching based on diffusion geometry and seed growing. In *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, 2011.

[55] S.T. Shivappa, M.M. Trivedi, and B.D. Rao. Hierarchical audio-visual cue integration framework for activity analysis in intelligent meeting rooms. In *Computer Vision and Pattern Recognition Workshops, IEEE Computer Society Conference on*, pages 107 –114, june 2009.

[56] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Proceedings of the Symposium on Geometry Processing*, SGP '09, pages

1383–1392, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association.

[57] H. Tamura, T. Sasaki, H. Hashimoto, and F. Inoue. Circle fitting based position measurement system using laser range finder in construction fields. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 209 –214, oct. 2010.

[58] Johan W. H. Tangelder and Remco C. Veltkamp. A survey of content based 3d shape retrieval methods. In *In Shape Modeling International*, pages 145–156, 2004.

[59] G. Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(11):1115 –1138, nov 1991.

[60] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *Int. J. Comp. Visi.*, 1(59):61–58, 2004.

[61] M Wand. *Kernel smoothing*. Chapman & Hall, London New York, 1995.

[62] Huai-Yu Wu, Hongbin Zha, Tao Luo, Xu-Lei Wang, and Songde Ma. Global and local isometry-invariant descriptor for 3d shape comparison and partial matching. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 438 –445, june 2010.

[63] S. Wuhrer, Z.B. Azouz, and Chang Shu. Posture invariant surface description and feature extraction. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 374 –381, june 2010.

[64] H. Yano, Y. Miyamoto, and H. Iwata. Haptic interface for perceiving remote object using a laser range finder. In *EuroHaptics conference,*

2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint, pages 196 –201, march 2009.

[65] Andrei Zaharescu, Edmond Boyer, Kiran Varanasi, and Radu P. Horaud. Surface feature detection and description with applications to mesh matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, June 2009.

[66] Wei Zeng and X.D. Gu. Registration for 3d surfaces with large deformations using quasi-conformal curvature flow. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2457 –2464, june 2011.

[67] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, D. Samaras, and N. Paragios. Dense non-rigid surface registration using high-order graph matching. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 382 –389, june 2010.

[68] H. Zhang, A. Sheffer, D. Cohen-Or, Q. Zhou, O. Van Kaick, and A. Tagliasacchi. Deformation-driven shape correspondence. *Computer Graphics Forum*, 27(5):1431–1439, 2008.

[69] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vision*, 13(2):119–152, October 1994.

[70] Huijing Zhao and R. Shibasaki. A novel system for tracking pedestrians using multiple single-row laser-range scanners. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35(2):283 – 291, march 2005.

[71] Bo Zheng, J. Takamatsu, and K. Ikeuchi. An adaptive and stable method for fitting implicit polynomial curves and surfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):561 –568, march 2010.

List of Publications

Journal Papers

1. Fumihiro Inoue, Takeshi Sasaki, Xiangqi Huang, Hideki Hashimoto, "A study on position measurement system using Laser Range Finder and its application for construction work," *Journal of Robotics and Mechatronics*, Vol. 24, No. 1, pp. 226-234, 2012

International Conferences

1. Xiangqi Huang, Takeshi Sasaki, Hideki Hashimoto, Fumihiro Inoue, "Circle detection and fitting based positioning system using laser range finder," *IEEE/SICE International Symposium on System Integration*, pp. 442-447, 2010

2. Xiangqi Huang, Takeshi Sasaki, Hideki Hashimoto, Fumihiro Inoue, "Circle detection and fitting using laser range finder for positioning system," *International Conference on Control Automation and Systems*, pp. 1366-1370, 2010

3. Takeshi Sasaki, Xiangqi Huang, Hideki Hashimoto, Fumihiro Inoue, "Position measurement of piles using a laser range finder for accurate and efficient pile driving," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 241-246, 2011

4. Fumihiro Inoue, Takeshi Sasaki, Xiangqi Huang, Hideki Hashimoto, "Development of position measurement system for construction pile using laser range finder," *Proceedings of the 28th International Symposium on Automation and Robotics in Construction*, pp. 574-580, 2011

5. Xiangqi Huang, Takeshi Sasaki, Hideki Hashimoto, Fumihiro Inoue, Bo Zheng, Takeshi Masuda, and Katsushi Ikeuchi, "An Accurate and Efficient Pile Driver Positioning System Using Laser Range Finder", *2012 International Workshop on Depth Image Analysis*, submitted.