

Projector Calibration for Spatial Augmented Reality in
Large-Scale Space

March 2020

Xie Chun

Projector Calibration for Spatial Augmented Reality in
Large-Scale Space

School of Integrative and Global Majors
Ph.D. Program in Empowerment Informatics
University of Tsukuba

March 2020

Xie Chun

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Professor Itaru Kitahara for his support and supervision throughout my Ph.D. studies. Without his persistent help, the goal of this research would not have been realized. The many opportunities he provided me over the past five years have allowed me to grow and broaden my horizons. He also convincingly guided me to be positive even when the road got tough. Then, I wish to show my gratitude to Professor Yoshinari Kameda for the sharp and right-to-the-point advice he gave me. The vast experience he shared also benefited me in my research life. I would also like to thank Professor Kenji Suzuki. His interdisciplinary comments and advice helped me re-examine my research with a new perspective. I also want to thank Mr. Issei Takahashi for his selfless help as a co-author on my publication. I need to reserve a special recognition to Assistant Professor Hidehiko Shishido for his support, guidance, and friendship. Besides, I would like to pay my special regards to all the staff of EMP Office, especially Ms. Nobue Kobayashi and Miss. Emi Murakami, for their great help on my campus life.

Lastly, I wish to acknowledge the support and great love of my parents. They stranded behind me and kept me going on during a tough time in their lives.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
LIST OF FIGURES.....	v
LIST OF TABLES	vii
Abstract.....	viii
Chapter 1 Introduction	1
Chapter 2 Projector Calibration for Spatial Augmented Reality and Related Works ...	6
2.1 Geometric Calibration of Projector.....	6
2.2 Related Works	7
2.2.1 Offline Calibration.....	7
2.2.2 Online Calibration	8
2.2.3 Self-Calibration	8
2.3 Unsolved Issues of Projector Calibration for SAR in Large-Scale Space	10
2.3.1 Working Distance	10
2.3.2 Lens Distortion	10
2.3.3 Tradeoff Between Depth Accuracy and Scene Completeness.....	12
Chapter 3 An Accurate Projector Calibration Method for Large Planar Surface Using Camera and Laser-Line Pointer	15
3.1 Camera Lens Distortion Correction	17
3.2 Projector Lens Distortion Correction.....	21
3.3 Projective Distortions Correction.....	23
3.4 Evaluation	24

3.4.1	Effectiveness of the Method	24
3.4.2	Accuracy Evaluation.....	25
3.4.3	Projection Results	29
Chapter 4	A Projector Calibration Method for Arbitrary Shaped Surface Using a Mobile Camera.....	32
4.1	Structured Light Patterns	33
4.2	Fail-Safe Design of Rhombus Pattern.....	37
4.2.1	Fail-Safe Coding Scheme	37
4.2.2	Fail-Safe Decoding Algorithm	38
4.3	Robust Projector Calibration Using a Mobile Camera	40
4.3.1	Point Cloud Generation	41
4.3.2	Codeword Correction	43
4.3.3	Projector Calibration	44
4.4	Evaluation of Local Image Features	47
4.5	Evaluation in Close-Range Scenes	49
4.5.1	Datasets.....	49
4.5.2	Comparison to Checkerboard-based Calibration.....	51
4.5.3	Projection Surface Reconstruction Results.....	52
4.5.4	Reconstruction Precision	52
4.6	Evaluation in Large-Scale Scenes	54
4.6.1	Datasets.....	54
4.6.2	Surface Reconstruction Result	57
4.6.3	Comparison to Self-Calibration using Stationary Camera	57

Chapter 5 Conclusion.....	60
References	64

LIST OF FIGURES

Figure 2-1. The relationships between depth uncertainty range and depth, baseline length, and uncertainty of the corresponding point on the camera image	12
Figure 2-2. The tradeoff between the baseline length and occlusion area.....	13
Figure 3-1. A real-world, large-scale SAR application: The floor projection system for learning aid in a school gym.....	15
Figure 3-2. Laser line projection on the floor.....	20
Figure 3-3. An example of the extracted laser lines.	20
Figure 3-4. The initial estimation of projector principle point.	22
Figure 3-6. The parameters involved in the Landolt ring VA test.	26
Figure 3-5. The definition of max edge misalignment D.....	25
Figure 3-7. A comparison of max edge misalignment after calibration.	28
Figure 3-8. The projection results before(a,c) and after(b,d) calibration	31
Figure 4-1. An example of the spatial coding SL pattern used in the proposed calibration method.	36
Figure 4-2. The two feature types and their codewords.	38
Figure 4-3. The pair-point searching strategy and a failure example.....	39
Figure 4-4. The flow diagram of the proposed calibration algorithm	46
Figure 4-5. The images that are used to evaluate the local image features.	48
Figure 4-6. Experiment setup of dataset #1, #2 and #3.....	50
Figure 4-7. Sample images of the close-range experiment datasets.....	50

Figure 4-8. Reconstruction results of dataset #1, #2 and #3.....	53
Figure 4-9. Side views of the scene in dataset #4 and #5.....	54
Figure 4-10. Sample images of the large-scale experiment datasets.....	55
Figure 4-11. A comparison of reconstruction results using both SIFT and pattern features, and only pattern features for dataset #4 and #5.	56
Figure 4-12. The camera views of the stationary camera for self-calibration.....	58

LIST OF TABLES

Table 3-1. Estimations of projector lens distortion.....	24
Table 4-1. Elements of the pseudorandom array	35
Table 4-2. A comparison of the quality of SIFT features extracted from four spatial coding structured light patterns in terms of quantity and discriminability.....	48
Table 4-3. Summary of the close-range datasets and calibration results.....	49
Table 4-4. Comparison of the estimations of focal length and principle point of the projector made by the proposed method and a checkerboard method.	51
Table 4-5. Summary of the large-scale datasets and calibration results	55

Abstract

The aim of this thesis is to consider how computer vision technologies can be applied to perform geometric projector calibration for large-scale spatial augmented reality (SAR) systems. Specifically, how to use the projector-camera system (PROCAMS), which is an active stereo system for 3D sensing, to estimate the crucial information to achieve geometrically correct projection result on a projection surface. Indeed, over the past decade, a number of new methods and technologies that apply PROCAMS to geometric projector calibration have been proposed, and some of them are for SAR. These works have proven their effectiveness in desktop scale, or at most living room scale scenes. In recent years, SAR has shown its value in many fields like exhibition, art, education, and advertising, where it is used to display contents in large-scale spaces. Despite this, the feasibility of PROCAMS-based geometric projector calibration methods in large-scale spaces is rarely discussed or mentioned, and manual calibration is still a common approach on site. Given the fact that PROCAMS-based method can simplify the calibration process, this thesis addresses the research question of what issues lowered the feasibility of PROCAMS-based projector calibration method as the scale increases, and how to solve them.

This thesis gives a brief introduction to the SAR, followed by a description of two common calibration scenarios (known planar surface and unknown surface) and the corresponding calibration problems. Three unsolved issues that arise as the scale of the

calibration scene increases are discussed in this thesis: working distance, lens distortion, and the tradeoff between depth accuracy and scene completeness. Why these issues significantly affect the calibration process in large-scale space and why they are unsolved are also explained. An online projector calibration method using a laser line pointer is proposed to address the issues in the planar surface scenario. It applies “straight lines have to be straight” method to estimate the lens distortions by optimizing the straightness of lines on the surface. An evaluation criterion is designed and according to which the proposed method outperformed a typical self-calibration method in a real-world, large-scale SAR scene. Furthermore, a method that uses a mobile camera to calibrate the projector is proposed to address the issues in the unknown surface scenario. The camera’s mobility, which makes the method unique among other existing methods, allows the use of multiview images without introducing extra cameras. The experiment results showed that the method achieved high-quality calibration results in several different scenes. It is also shown that the method solved the problem of tradeoff depth accuracy and scene completeness despite the long working distance thanks to the multiview image.

Chapter 1

Introduction

This thesis discusses the research on projector calibration methods for large-scale Spatial Augmented Reality (SAR) systems. In particular, it discusses several unsolved issues arise in the calibration process as the scale of the SAR scene increases and proposes corresponding solutions to address them using computer vision technologies.

In recent years, a concept named Augmented Reality, or AR, has been getting more and more popular in both research and application fields. Different from another concept with a similar name called Virtual Reality or VR, which displays visual information that is entirely artificial, AR extents human's vision by overlapping extra artificial information onto the real world, thus augments reality.

According to how artificial visual information is displayed, AR technologies can be divided into HMD-based and projection-based. The latter is also known as SAR. It displays artificial visual information by directly projecting images onto the surfaces in the real world. Some typical applications of SAR have been developed: Shader Lamps [1] allows virtual painting on real 3D objects. Deformation Lamp [2] makes static printed images perceptually moving by adding motion information. Some applications like [3] can change the perceived physical proprieties of the projection surface. Also, there are several large-scale immersive SAR systems like Large Space [4].

The significant advantages of SAR over HMD-based AR are: first, it puts no burden on the users as it does not require any wearable device; second, the projected information is shared by every user in the scene; and third, it is relatively simple to display large content in a wide space. These advantages make SAR a rather good choice for large-scale, public application in fields like exhibition, art, education and advertising.

However, the disadvantages of SAR are also obvious. As the display devices (i.e., the projectors) and the display surfaces are typically stationary, SAR applications are usually required to be run in a spatialized space where the display system has been installed in advance. Consequently, deploying a projection system for SAR can be extremely cumbersome. First, users have to measure the geometry of the projection surface, and then run geometric calibration for all the projectors in the system so that the projection results are geometrically correct. Moreover, color and luminance calibrations are also required in many situations, such as when multi-projector alignment and blending is involved, or the physical properties of the surface like color and albedo significantly affect the projection results. Apart from that, users also need to deal with many other tasks and challenges, including but not limited to projector installation, projector control, and frame synchronization. As the scale of the space increases, it increases the deployment difficulty and raises new issues that may not be significant in a small-scale scenario. Therefore, research on new technologies that simplify the deployment process as well as investigate and resolve issues that arise as the scale increases is of great value.

This thesis focused on the topic of projector geometric calibration in large-scale SAR. Traditionally this can be achieved by manual methods. In short, the user needs to

deform the projector image manually while observing the projection until an acceptable result is obtained. This method is time-consuming and laborious. To simplify the process, one can use projector-camera systems (PROCAMS), which is an active stereo system. In PROCAMS, the camera is used to capture the projection results. In order to calibrate the projector, correspondences between the camera image and the projector image are required. The correspondences are usually generated by structured light (SL) projection. Although there has been much research on PROCAMS in the last decade, they mostly assume a close-range scenario. The issues when using it to calibrate the projector in a large-scale space were, unfortunately, rarely mentioned and discussed.

The contributions of this work lie in proposing new PROCAMS based projector calibration methods to solve the issues in large-scale SAR scenarios. In particular, this research focuses on the following issues:

- In a large-scale space, the working distance of the projector is usually too large to apply conventional offline and online calibration methods. The reason is that these methods usually require a calibration board. However, as the working distance increases, the size of the required calibration board also increases to an impractical level. To solve this issue, methods that do not require any calibration board in this research.
- The long working distance and large projection area also enlarge the effect of lens distortion on a planar projection surface when the users watch the projected images in a relatively close range, or projector stitching is required in subsequent processing, or both. Therefore, it needs to be calibrated more accurately. In this research, the high

accuracy is achieved by using line straightness, which is more viewer-oriented, as an objective function to be optimized.

- The typical setup of PROCAMS, where the camera is installed very close to the projector, causes higher triangulation error in large-scale space as the baseline between the projector and camera is too short comparing to the distance between the target object and the projector. The triangulation error significantly affects the calibration accuracy. Merely increasing the baseline length, however, potentially causes occlusions and thereby affect scene completeness. In this research, this tradeoff is solved by using a mobile camera and take advantage of multiview images.

By solving the above problems, this study makes PROCAMS a more feasible method for projector calibration in large-scale SAR, thereby helping to simplify the calibration process. These three issues are explained in further detail in Section 2.3.

Depending on the SAR scenario, two evaluation methods are used in this research. For a known planar projection surface, the pixel misalignment in real-world scale is calculated. Then it is compared to the desired calibration accuracy calculated from viewing distance, projector resolution, projection area, and visual acuity. For more complex unknown surfaces, the effectiveness of the proposed method is evaluated by 3D reconstruction accuracy, scene completeness, and reprojection error.

In Chapter 2, the common scenarios of projector geometric calibration are introduced at first. This chapter then shows some conventional camera-based projector calibration methods and discuss their limitations when applied to SAR in large-scale

space. Chapter 3 introduces an online calibration method using a camera and laser line pointer. This method was proposed to calibrate a large-scale projection system for a planar surface. The calibration process is also described in detail, and the calibration accuracy the method achieved is shown in the experiments. Chapter 4 introduces an innovative method that allows calibrating a projector using a mobile camera. This method shows how projector calibration in large-scale space can benefit from multiview images taken by the camera. Chapter 5 makes a conclusion to this research.

Chapter 2

Projector Calibration for Spatial Augmented Reality and Related Works

2.1 Geometric Calibration of Projector

Usually, the projector projection is regarded as an inverse projection of a pin-hole camera, and therefore, it is the perspective projection similar to the camera models

Two scenarios for projector calibration in SAR are considered in this research. The first is that the projection target is planar, and its coordinate system has been defined. In this case, we need to estimate the 3×3 homography \mathbf{H} that satisfies:

$$\boldsymbol{x} \sim \mathbf{H}\mathbf{X} \quad (2-1)$$

where \boldsymbol{x} and \mathbf{X} are the homogenous coordinates of the corresponding 2D points on the projector plane and the projection target, respectively. Warping the projector buffer by \mathbf{H} , we can get the geometrically correct projection result on the target plane. The second scenario is when the shape of the projection surface is unknown. In this case, we need to

perform 3D reconstruction to estimate the geometry of the projection surface X' , and also estimate the intrinsic matrix K and extrinsic matrix $[R | t]$ of the projector that satisfies:

$$x \sim K[R|t]X' \quad (2-2)$$

where X' is the homogenous coordinate of the corresponding 3D point of x on the projection surface, R and t are the 3×3 rotation matrix and the 3×1 translate vector, respectively.

In either scenario, the acquisition of correspondences between the projector and the projection target is essential, and it requires the assistance of a camera. Therefore, camera calibration is also necessary.

2.2 Related Works

2.2.1 Offline Calibration

There are many methods to calibrate the projector-camera system in a separated manner, such as [5] [6] and [7]. They first calibrate the camera intrinsics parameters offline via a checkerboard-based method. Then use the calibrated cameras to estimate the projector intrinsics and extrinsics. The most commonly used one for camera calibration is Zhang's method [8]. This method requires a set of images of a checkerboard pattern captured from several different views. Its accuracy highly depends on the number of samples, and the checkerboard orientations in the images thus require expert knowledge to achieve accurate calibration results. In [9], A. Richardson et al. proposed a method to

assist users in selecting the next checkerboard pose. It iteratively estimates the most valuable pose that improves the calibration accuracy and shows the pose to the user.

2.2.2 Online Calibration

To simplify the PROCAMS calibration process, researchers proposed several methods to calibrate the projector and camera simultaneously by projecting some pattern images onto a planar checkerboard [10] [11]. As the pattern images encoded the coordinate information of projector pixels, their corresponding positions on the checkerboard plane can be detected from camera images. With these correspondences, a projector can also be calibrated using the same manner as Zhang's camera calibration method. In [10], pixel coordinates are encoded via Gray code patterns [12]. In [11], for each checkerboard orientation, after capturing the real checkerboard, a sheet is covered on top of the checkerboard, changing it into a projection screen, and then project a checkerboard image onto it.

2.2.3 Self-Calibration

Yamazaki et al. [13] proposed a fully automatic calibration method for a generic projector-camera pair. The method is based on the decomposition of a radial fundamental matrix into intrinsics and extrinsics of the two devices. Deglint et al. [14] tried to improve the self-calibration method by using a modified cost function and optimization strategy. These methods use (2-3), also known as Bougnoux's formula [15], to calculate the focal

lengths of the projector and camera from their principle points p_p and p_c , and an estimation of the fundamental matrix F .

$$f_p^2 = -\frac{p_c^T [e_c]_{\times} \hat{I}_3 F p_p p_p^T F^T p_c}{p_c^T [e_c]_{\times} \hat{I}_3 F \hat{I}_3 F^T p_c} \quad (2-3)$$

$$f_c^2 = -\frac{p_p^T [e_p]_{\times} \hat{I}_3 F^T p_c p_c^T F p_p}{p_p^T [e_p]_{\times} \hat{I}_3 F^T \hat{I}_3 F p_p}$$

$[e_c]_{\times}$ and $[e_p]_{\times}$ are the skew-matrix of the left and right null-vector of F , respectively, and $\hat{I}_3 = \text{diag}(1,1,0)$. It is known that (2-3) is very sensitive to the assumed position of principle points and even small errors in F . In the worst cases, the right side of (2-3) can become negative, which makes the solution impossible [16]. In such a case, the calibration completely fails. Therefore, these methods require a relatively accurate initial guess for the principal point. However, this is very difficult to achieve because, unlike the camera, the principal point of a projector is usually not close to the image center, and its position is affected by the projector's lens shift function. S. Will and A. Grundhofer proposed a generic self-calibration method for multi-projector-camera systems [17]. It automatically selects a pair of cameras that produce the most reliable calibration result, and then iteratively integrate the rest cameras and projectors. It makes sure that only camera pairs that give reasonable results in (2-3) are considered as candidates of the initial pair. This method requires at least two stationary cameras for the initial pair selection. It also requires expert knowledge to choose the appropriate number and locations of camera views to calibrate the system successfully.

2.3 Unsolved Issues of Projector Calibration for SAR in Large-Scale Space

2.3.1 Working Distance

Checkerboard based online and offline calibration methods are viable only when the projection target is relatively close to the projector and camera. It is because the calibration is usually required to be performed on the working distance. As the working distance increases, the projection area also increases, resulting in an increase in the size of the checkerboard to be used. The calibration distance is therefore limited by the available size of the checkerboard in practical. In a large-scale space, the working distance is usually considerable, which makes it impossible to use a regular size checkerboard for calibration. Furthermore, zoom, focus, and offset may be adjusted after installation, making offline calibration method more impractical in large-scale spaces.

2.3.2 Lens Distortion

Lens distortion correction is essential, especially in multiple projector applications where projector stitching is involved. Some calibration methods like [5] treat the projector's lens distortion as ignorable parameters. It is acceptable if the projectors are in close range to the screen. However, when the distance between the projector and screen increases to achieve a larger projection area, lens distortion significantly affects the alignment accuracy, and the misalignment becomes more obvious to users as they are relatively close to the screen. Therefore, it should be corrected by calibration.

Checkerboard-based online and offline calibration method can estimate the lens distortion accurately if enough checkerboard poses are given. However, they are not feasible in large-scale space due to the reasons given in section 2.3.1. Self-calibration methods can estimate the lens distortion parameters by non-linear optimization that minimize reprojection error [13] [14] [17]. However, the projection results are not guaranteed to be visually correct to the users, so new approaches are needed to address this problem.

2.3.3 Tradeoff Between Depth Accuracy and Scene Completeness

A PROCAMS is an active stereo system; therefore, it suffers from the problem of the tradeoff between occlusion and depth error like other stereo systems. Consider a simple scene as being shown in Figure 2-1: the point q , which is the projection of projector feature p on the camera image, is known inexactly: $x+\delta x$, $x-\delta x$. The depth accuracy δZ can be expressed as:

$$\delta Z = \frac{Z^2}{Bf} \delta x \quad (2-4)$$

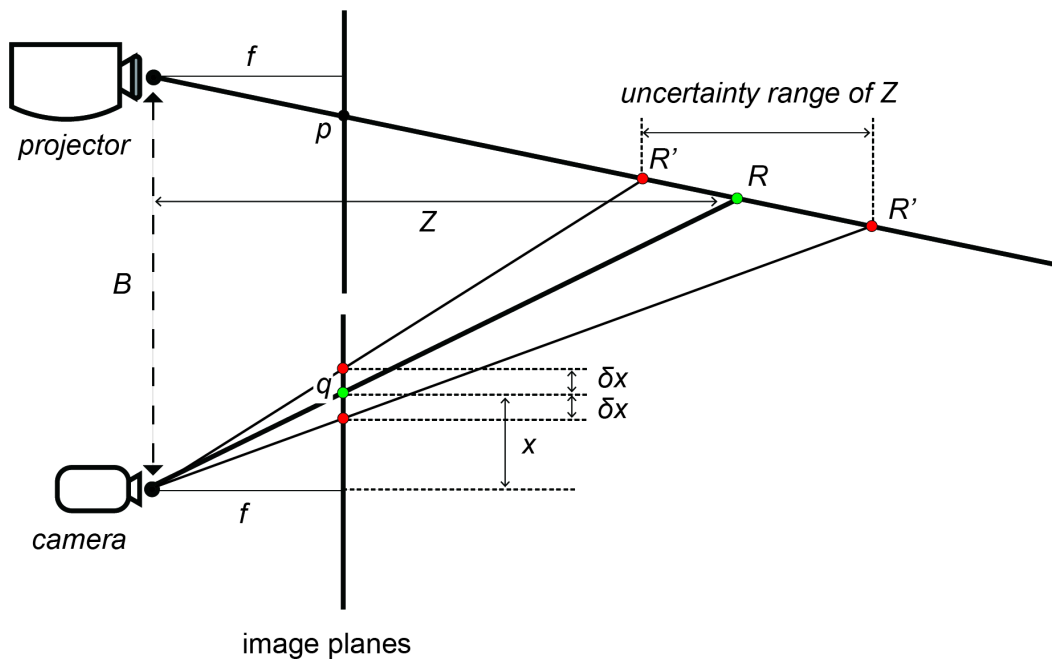


Figure 2-1. The relationships between depth uncertainty range and depth (Z), baseline length (B), and uncertainty of the corresponding point on the camera image (δx).

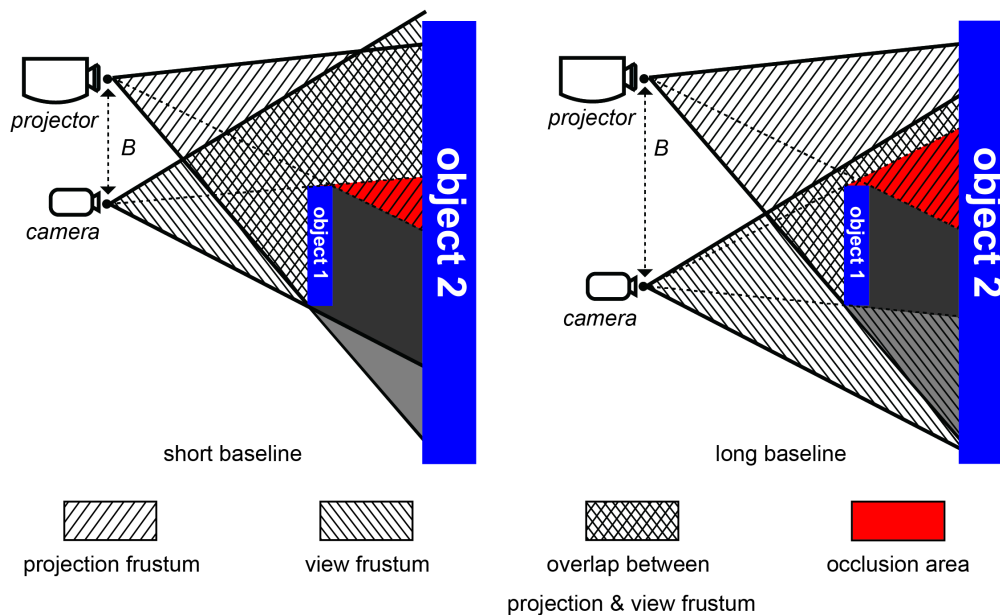


Figure 2-2. The tradeoff between the baseline length and occlusion area. The occlusion area is covered by the projection but not be able to be captured by the camera. It shows that as the baseline between the projector and camera becomes longer, the occlusion area becomes larger.

where B is the baseline, f is the focal length, and Z is the depth. It expresses how much depth Z can change given a change in x , and that the uncertainty of Z is proportional to the square of Z and inversely proportional to the length of baseline B . In other words, given an uncertainty range of x , the uncertainty range of depth becomes larger as B decreases and Z increases. On the other hand, as demonstrated in Figure 2-2, increasing the baseline can cause larger occlusion; therefore, lower the completeness of the scene. Moreover, it causes difficulties in finding correspondences between the projector and camera due to the large distortion and the change of intensity [18]. Due to the tradeoff described above, in a large-scale space where the distance between the projector and screen is long (i.e., the depth Z is large), finding an appropriate position to set the camera

for the projector calibration can be tricky. Setting the camera close to the projector ensures that the whole projection area is visible to the camera; however, it results in a short baseline, thus potentially causes large depth error. Conversely, moving the camera away from the projector such that the angle is sufficiently large will cause occlusions and also lower the number of their correspondences.

Chapter 3

An Accurate Projector Calibration Method for Large Planar Surface Using Camera and Laser-Line Pointer

In this chapter, we consider a real-world, large-scale SAR system, as shown in Figure 3-1: a floor projection system consists of four sets of projector-camera systems with each camera mounted on top of a projector. This system is installed on the ceiling of a school gym, and the projections are displayed on the floor. The resolutions of the projector and the camera are $1024 \text{ pixels} \times 768 \text{ pixels}$ and $1920 \text{ pixels} \times 1200 \text{ pixels}$, respectively. The angles of view of the projector and the camera are 60° and 70° , respectively. The projection area covered by four projectors is about 90.2 m^2 ($11 \text{ m} \times 8.2 \text{ m}$). The distance from the projectors to the floor is about 6m.

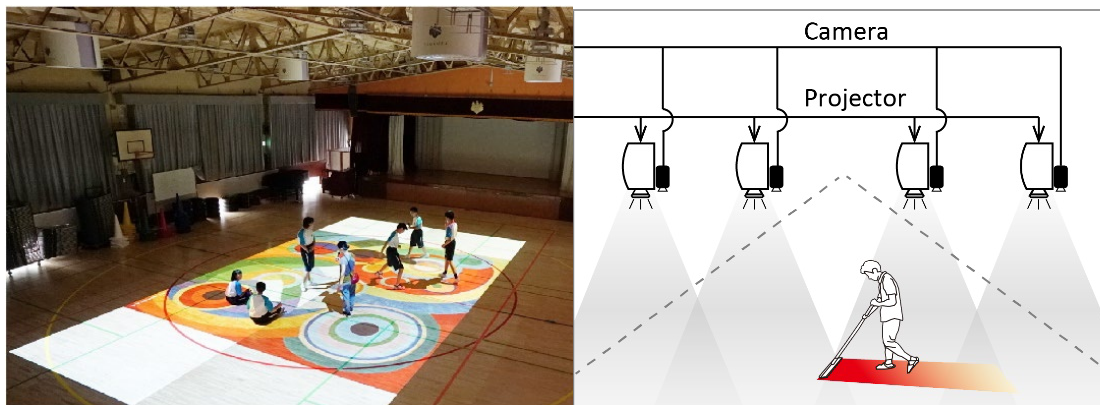


Figure 3-1. A real-word, large-scale SAR application: The floor projection system for learning aid in a school gym.

By analyzing the projection results using the cameras attached to each projector, the lens distortion of the projectors, and the 2D homography between the floor surface and each projector's image plane, can be estimated. The proposed calibration method is based on the following assumptions:

- The floor is approximately planar.
- The projectors are coarsely aligned in advance.
- The coordinate system of the floor is defined and can be identified in the cameras.
- Laser lines projected onto the floor can be captured by the cameras.

A projected image captured by the camera includes the following four kinds of distortion:

- projector lens distortion;
- projective distortion between the projector image plane and the projection plane;
- camera lens distortion;
- projective distortion between the camera image plane and the projection plane.

The lens distortions are caused by the optical design of the lenses of the camera and the projector. The projective distortions are caused by the non-parallel positioning among the projector plane, the camera plane, and the projection plane.

As has been discussed in Chapter 2, it is impossible to apply any calibration method that requires moving a checkerboard to the system. Therefore, it is difficult to estimate the above distortions simultaneously. In the proposed method, these distortions

are estimated separately. The ideal estimation of projective distortion parameters is based on the assumption that the projected and captured images satisfy a perspective transformation, but due to the presence of the lens distortion, they cannot be estimated accurately. On the other hand, the lens distortion parameters can be estimated independently by measuring and correcting the straightness of lines that are supposed to be straight in an image. This method is also known as “straight lines have to be straight” [19]. With this method, it is possible to first correct the lens distortions, and next perform projective distortion correction.

For estimating the lens distortion parameters of the projector, the projection results must be captured by the camera. In that case, the camera lens distortion must have been corrected. Therefore, in the proposed method, it is necessary to correct the camera lens distortion first and then the projector’s lens distortion and finally estimate the projection distortion parameters without suffering any effects from the lens distortions.

3.1 Camera Lens Distortion Correction

In order to apply “straight lines have to be straight”, the peaks (i.e., center positions) of a set of straight lines in the camera images need to be extracted. Since there is no guarantee that enough straight lines exist in the camera’s view field to ensure a reliable result, a laser line pointer (Figure 3-2) is used to project several straight lines onto the projection plane and generate a series of sample points for each line with sub-pixel accuracy. The sample points are the peak positions across the width of a line. A laser line

detection method proposed by A. Molder et al. [20] are applied to detect the line peaks. It first upsamples the image by an existing algorithm [21]. Then it uses the Laplacian of Gaussian (LoG) as a kernel to convolve the gray level profile of the image and detects the peaks in pixel accuracy. Finally, a parabola fitting procedure is applied to refine the peaks to subpixel accuracy by finding the extreme value of the parabola curve. In practice, the image can have many noises; thus, not all the detected peaks belong to a laser line. To solve the problem, a 2-D line fitting with the random sample consensus (RANSAC) algorithm [22] is applied in our method. The peaks that not fit the line model are considered outliers and removed. Figure 3-3 shows an example of the combined detection results of multiple laser lines.

A model of the camera lens distortion is expressed by (3-1) where (x, y) is the original distorted coordinate, (\hat{x}, \hat{y}) is the undistorted coordinate, r is the distance between (x, y) and the center of camera distortion model (x_c, y_c) , (x_c, y_c) is usually the center of the image, and κ_1, κ_2 are distortion parameters.

$$\begin{cases} (\hat{x} - x_c) = (1 + \kappa_1 r^2 + \kappa_2 r^4)(x - x_c) \\ (\hat{y} - y_c) = (1 + \kappa_1 r^2 + \kappa_2 r^4)(y - y_c) \end{cases} \quad (3-1)$$

Let $\{(x_{l,i}, y_{l,i})\}$ with $l = 1, \dots, N$ and $i = 1, \dots, N_l$ be the projection of the N sets of the 3D aligned points in the 2D image, where $\{(\hat{x}_{l,i}, \hat{y}_{l,i})\}$ are the corresponding corrected points

using the distortion model (3-1). The covariance matrix for each point set l is given by (3-2), where $\bar{x}_{l,i}$, $\bar{y}_{l,i}$ are the average of the variables taken over i .

$$\hat{S}^l(\kappa_1, \kappa_2, x_c, y_c) = \begin{pmatrix} \hat{S}_{xx}^l & \hat{S}_{xy}^l \\ \hat{S}_{xy}^l & \hat{S}_{yy}^l \end{pmatrix} \quad (3-2)$$

$$= \frac{1}{N_l} \begin{pmatrix} \sum_{i=1}^{N_l} (\hat{x}_{l,i} - \bar{\hat{x}}_{l,l})^2 & \sum_{i=1}^{N_l} (\hat{x}_{l,i} - \bar{\hat{x}}_{l,l})(\hat{y}_{l,i} - \bar{\hat{y}}_{l,l}) \\ \sum_{i=1}^{N_l} (\hat{x}_{l,i} - \bar{\hat{x}}_{l,l})(\hat{y}_{l,i} - \bar{\hat{y}}_{l,l}) & \sum_{i=1}^{N_l} (\hat{y}_{l,i} - \bar{\hat{y}}_{l,l})^2 \end{pmatrix}$$

The energy of the N sets of the aligned points is given by (3-3). According to Cauchy-Schwarz inequality, E decreases as the lines becoming more straight, and $E=0$ if and only if the points $\{(\hat{x}_{l,i}, \hat{y}_{l,i})\}$ are perfectly aligned linearly for every line l . Distortion parameters κ_1 , κ_2 can then be estimated by solving a minimization problem of E using the algebraic approach proposed by L. Alvarez et al. [23].

$$E(\kappa_1, \kappa_2, x_c, y_c) = \frac{1}{N} \sum_{l=1}^N \left(\hat{S}_{xx}^l \hat{S}_{yy}^l - \hat{S}_{xy}^l{}^2 \right) \quad (3-3)$$

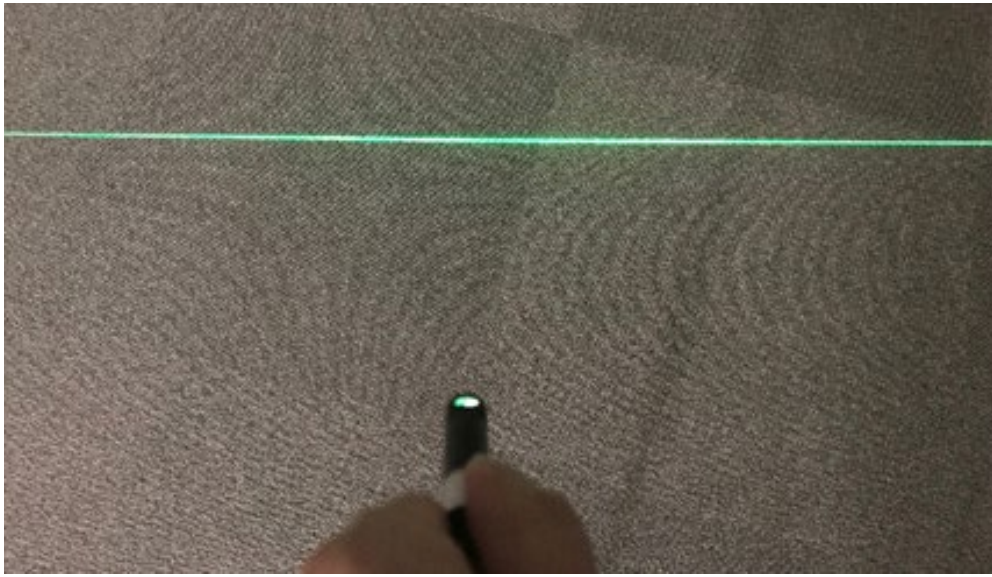


Figure 3-2. Laser line projection on the floor.

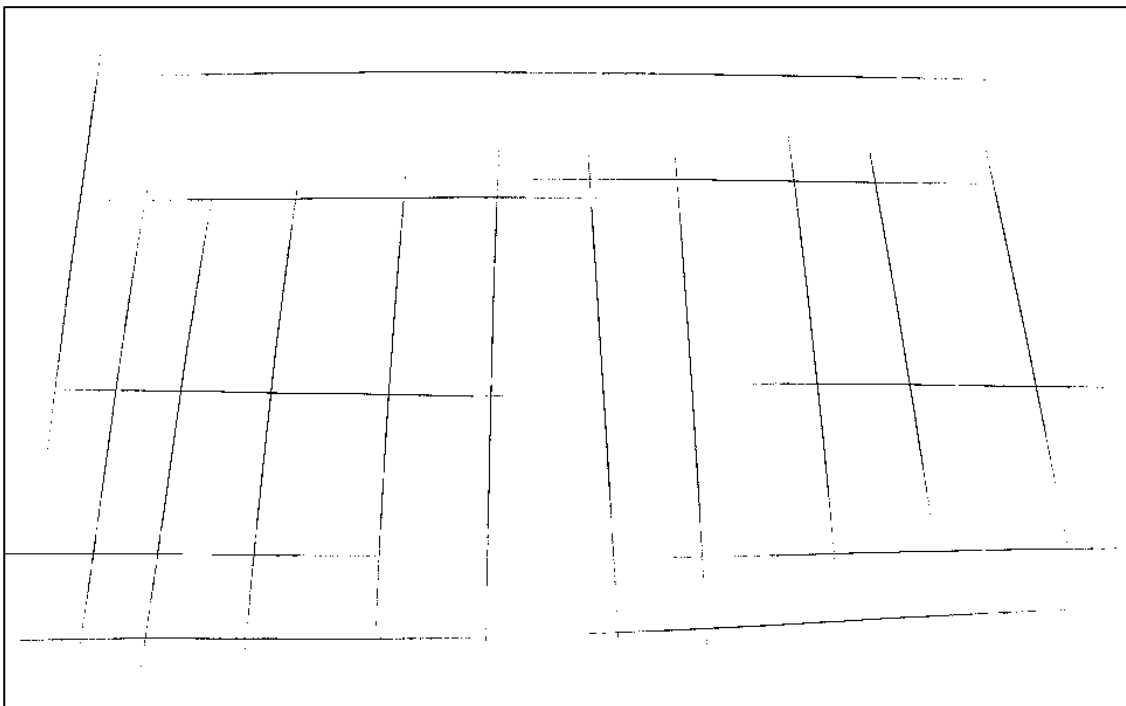


Figure 3-3. An example of the extracted laser lines.

3.2 Projector Lens Distortion Correction

The projector's optical system can be seen as an inverted model of a camera. Thus, the lens distortion model is given by the following expression:

$$\begin{cases} \hat{x} - x_c = (1 + \kappa_1 r^2 + \kappa_2 r^4)(x_0 - x_c) \\ \hat{y} - y_c = (1 + \kappa_1 r^2 + \kappa_2 r^4)(y_0 - y_c) \end{cases} \quad (3-4)$$

Eq.(3-4) is similar to (3-1) but where (x_0, y_0) is the undistorted coordinate in the projector's original image buffer, (\hat{x}, \hat{y}) is the distorted coordinate, and r is the distance between (x_0, y_0) and (x_c, y_c) . (\hat{x}, \hat{y}) can be corrected to undistorted coordinate (x_0, y_0) by pre-warping (x_0, y_0) to (x', y') using (3-5) where κ_1' and κ_2' are the distortion parameters for pre-warping:

$$\begin{cases} x' - x_c = \frac{x_0 - x_c}{1 + \kappa_1' r^2 + \kappa_2' r^4} \\ y' - y_c = \frac{y_0 - y_c}{1 + \kappa_1' r^2 + \kappa_2' r^4} \end{cases} \quad (3-5)$$

The projector lens distortion is estimated by evaluating the straightness of projected straight lines. Since the projector cannot “see” its projection result, instead of the actual (\hat{x}, \hat{y}) , only the coordinate of its projection on the camera plane is known. To efficiently project a projector pixel directly to the camera plane, a projector-to-camera correspondence map $\mathbf{P2C}$, which maps projector pixels to camera coordinates with sub-pixel accuracy, is used. $\mathbf{P2C}$ can be easily generated by SL like line-shifting [24] beforehand. Using $\mathbf{P2C}$, the optimal parameters can be found by first projecting several straight lines and capturing the results virtually, and then evaluating their straightness on the camera plane using (3-2) and (3-3).

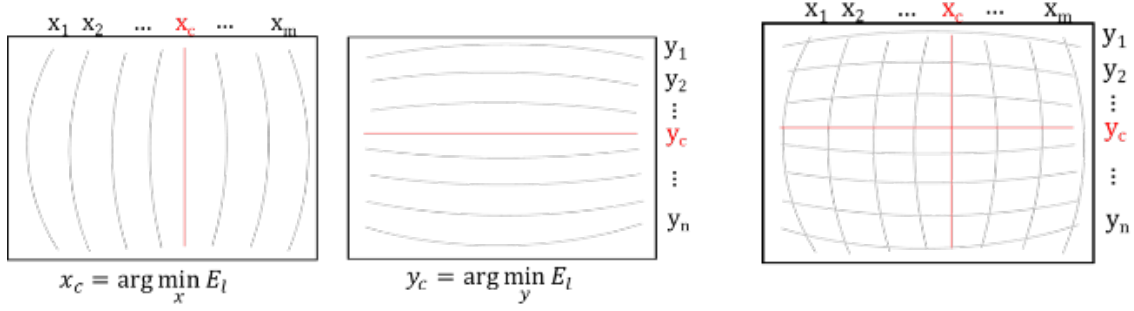


Figure 3-4. The initial estimation of projector principle point.

From the lens distortion model given by (3-4), the image lines passing distortion center (x_c, y_c) are not bent by the distortion regardless of the values of κ_1 , κ_2 and r . Using this property, the distortion center can be estimated by finding the intersection of the straightest row and column. This procedure is demonstrated in Figure 3-4. Although the result provided by this estimation only has pixel precision and may contain errors of several pixels caused by noise, it is nonetheless a good initial estimate for further optimization.

The lens distortion parameters κ_1' , κ_2' , x_c , and y_c are estimated by solving the objective function (3-6) using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. E here represents the average energy (i.e., a straightness evaluator) of the lines extracted from the camera image. These lines can be produced by projecting straight lines from the projector image directly to the camera image using $P2C$.

$$\{\tilde{\kappa}'_1, \tilde{\kappa}'_2, \tilde{x}_c, \tilde{y}_c\} = \operatorname{argmin} E(\kappa'_1, \kappa'_2, x_c, y_c) \quad (3-6)$$

3.3 Projective Distortions Correction

After all the lens distortions are corrected, projective distortion correction is straightforward. First, points on the projector plane are pre-warped by (3-5), then mapped to the corresponding camera coordinates through $P2C$. Then, from the correspondences between the projector image and camera image, the homography matrix between the undistorted projector and camera plane H_{pc} can be estimated.

Similarly, with the coordinate system that is already defined on the floor, correspondences between the floor and the camera can be found and from which the homography matrix H_{cf} is estimated.

Finally, from the two homography H_{pc} and H_{cf} , the third homography H_{pf} , which transforms the projector coordinates to a floor coordinate, can be calculated by (3-7).

$$H_{pf} = H_{pc}H_{cf} \quad (3-7)$$

By pre-warping a projector buffer with H_{pf}^{-1} , the projection distortion can be corrected for each projector. As the coordinate system defined on the floor is global to all the projector cameras, H_{pf} maps the projector pixels to the absolute coordinates of the floor.

Table 3-1. Estimations of projector lens distortion

Variables	κ_1' (10^{-7})	κ_2' (10^{-14})	Distortion center (x_c, y_c)
Ground truth	-1.023	4.387	(512, 384)
Noise = ± 0.1	-1.025	4.428	(512.104, 383.858)
Noise = ± 0.3	-1.024	4.404	(511.214, 383.966)
Noise = ± 0.5	-1.012	4.138	(512.726, 384.681)

3.4 Evaluation

3.4.1 Effectiveness of the Method

A quantitative experiment was conducted by simulation. Three *P2C* maps with white noise $[-0.1, 0.1]$, $[-0.3, 0.3]$ and $[-0.5, 0.5]$ (unit: pixel) were used. Based on these three maps, the ground truth of lens distortion and the calibration results are shown in Table 3-1. It can be seen that the estimation results of κ_1' , κ_2' , and distortion center (x_c, y_c), all converged to the ground truth. The estimations of κ_2' had larger variance as its effect on the lens distortion is much weaker than κ_1' .

3.4.2 Accuracy Evaluation

In order to numerically evaluate the misalignment occurring in the vicinity of the seam of two aligned projections, the max edge misalignment D is defined as being shown in Figure 3-5. D is calculated as twice the maximum of the distance between the corresponding sides of the circumscribed rectangle and the inscribed rectangle of the projection result. Its value represents the maximum error of the most deviated place when aligning two projections having the same degree of distortion with minimum overlapping and no gap.

The desired calibration accuracy was defined and calculated based on human visual acuity (VA), which is a measure of the spatial resolution of the visual processing system. Figure 3-6 demonstrates the concept of an international standard VA test method named Landolt ring [25] [26]: if at m meters, a human eye is able to separate contours

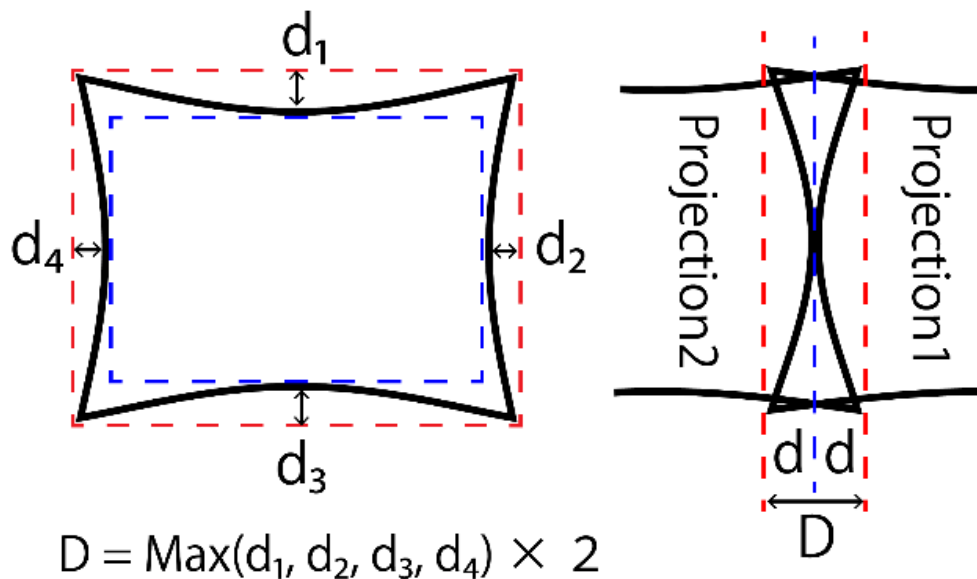


Figure 3-5. The definition of max edge misalignment D .

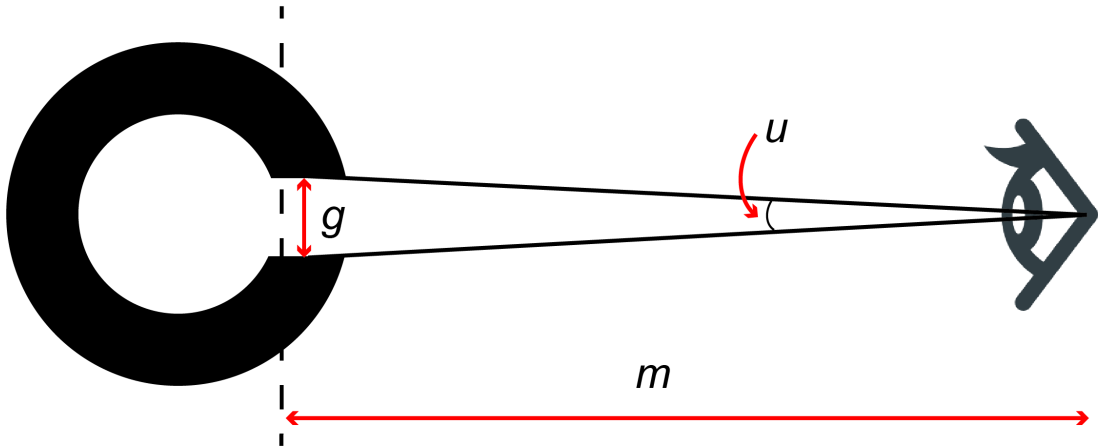


Figure 3-6. The parameters involved in the Landolt ring VA test.

that are g mm apart, then the numerical representation of its performance VA is calculated via (3-8) and (3-9) where u is the visual angle (in arc minute) at the eye under which the gap appears. Conversely, given VA and m , the minimum gap size g_{min} that a human can see can be calculated by (3-10). For example, the minimum gap that a human has VA of 1.0 can see from 5 meters away is approximately 1.45 mm.

$$u = 2 \arctan\left(\frac{g}{2m}\right) \times 60 \quad (3-8)$$

$$VA = \frac{1}{u} \quad (3-9)$$

$$g_{min} = 2m \times \tan \frac{1}{120VA} \quad (3-10)$$

Furthermore, as it is assumed that the sizes of projector pixels in the projection result are almost the same, g_{min} can be convert to pixel unit via (3-10), and then compared to the max edge misalignment D . In (3-11), w and h are the width and height of the

projector image's dimension and W , H are the width and height of the projection area respectively.

$$g_{min_in_pixel} = g_{min} \times \sqrt{\frac{w \times h}{W \times H}} \quad (3-11)$$

In the use case of the floor projection system described at the beginning of this chapter, which is the learning aid activities conducted at a school gym, the users are not staring at some particular locations. Instead, their viewpoints are moving all the time to observe the various parts of the visual information projected on the floor. As a result, the misalignment that happens at the seam between two projections is observed outside a user's central visual field most of the time. It is also known that the VA outside the central visual field decreases to 0.1, even for a person who has a VA of 1.0 [27]. Therefore, the calibration accuracy needs to be achieved is that the misalignment is hardly noticed when it happens outside the user's central visual field. As a general situation, when users (children) with a height of 1.4 meters look at the floor at an oblique angle of 45 degrees, they observe the projected images at a distance m of about 2 meters. In such a situation, the acceptance range of projection misalignment calculated from (3-10) is 5.82 mm or less. Since the dimension of the projector image is 1024×768 pixels, and the projection area of each projector is about 25m², calculated from (3-11), the acceptance range of projection misalignment of this system is 1.03 pixels or less.

The accuracy of the proposed method was evaluated by comparing it with the desired accuracy, and the results of the self-calibration method proposed by Yamazaki et al. [13]. As the prior knowledge of the distortion center of the projector required by

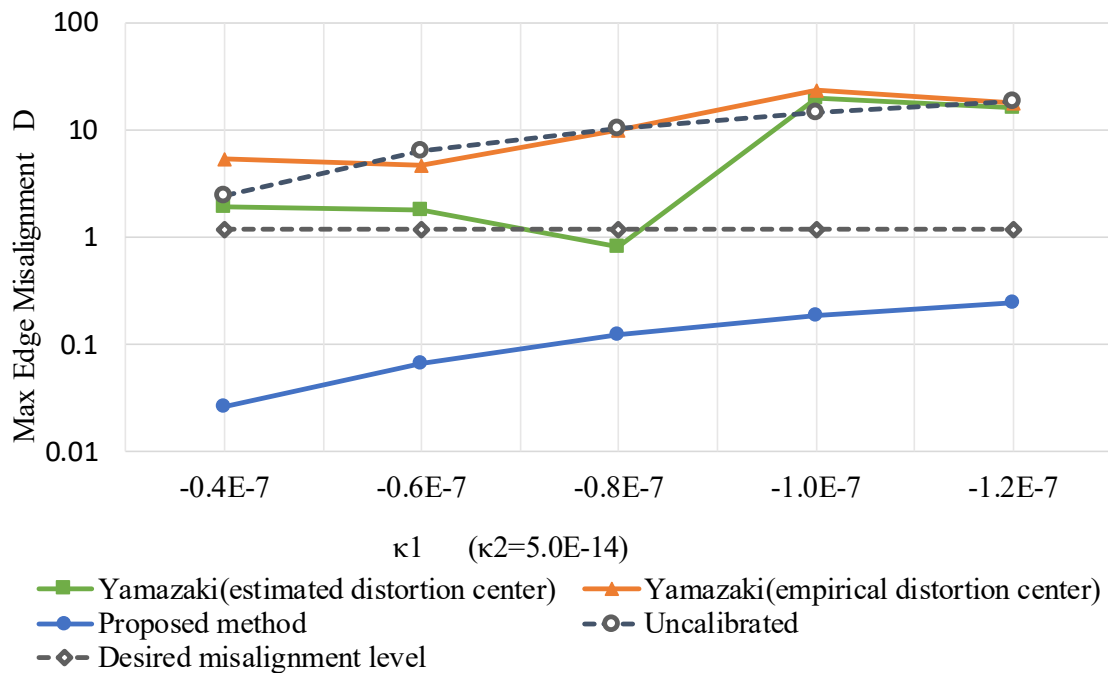


Figure 3-7. A comparison of max edge misalignment after calibration. κ_1 increases along x-axis and κ_2 is constant.

Yamazaki’s method is not available, two initial distortion centers were used in the experiment. One was an empirical position suggested in their work, which is somewhere below the image center; in the experiment, $(projector_width \times 0.5, projector_height \times 0.8)$ was used. The other one was the initial guess of the distortion center estimated by the method described in section 3.2.

The result is shown in Figure 3-7. The dotted line with rhombus marks in the graph is the calibration accuracy (1.16 pixels) desired for this system. The dotted line with circle marks is the misalignment level before calibration. From this result, it can be confirmed that the proposed method achieves the target calibration accuracy. On the other hand, the results of the calibration method proposed by Yamazaki et al. did not reach the target accuracy, and the estimation result was unstable, depending on the degree of distortion.

In higher distortion level, the results were even worse than the uncalibrated case. The reasons are as follows: First, their method uses Bougnoux's formula, which is known to be sensitive to even small errors of the fundamental matrix, while the corresponding points that are used to estimate the fundamental matrix, are unfortunately co-planar, which usually gives bad result. Second, the lens distortion parameter, together with other intrinsic and extrinsic parameters, are estimated by optimizing the reprojection error of the correspondences between projector and camera. The actual appearance of projection on the screen, however, is not taken into consideration. Our method corrects the projection by trying to keep all straight lines on the screen straight, and therefore gives better projection results. Moreover, by correcting the straight lines before estimating the homography, the nonlinear term can be estimated separately. This is considered to be another reason why the proposed method achieved higher calibration accuracy.

3.4.3 Projection Results

In this experiment, four projectors were used to project images on the floor. Their projections were approximately aligned before calibration. Two kinds of images are used to evaluate the appearance of the projection result: a grid pattern that only consists of one-pixel lines and a color block pattern. Figure 3-8(a) shows that for the line grid image, some lines are bent, and noticeable line misalignments occurred around adjacent edges between the projections. Figure 3-8 (b) shows that for the color block images, the gap between the projections is much more obvious than line bending and misalignment. Note that since adjacent edges are curved, the gaps cannot be solved by linear transformations. Figure 3-8 (c) gives a corrected result of the line grid generated by our method. In this

result, distorted lines are straightened, and misalignment and gaps around adjacent edges can barely be observed. Figure 3-8 (d)'s result is similar to that of Figure 3-8 (c). Although it is not seamless since there is a very narrow overlapping between two adjacent projections where the projection intensity is higher than the surrounding areas, the problem can be solved by additional intensity normalization.

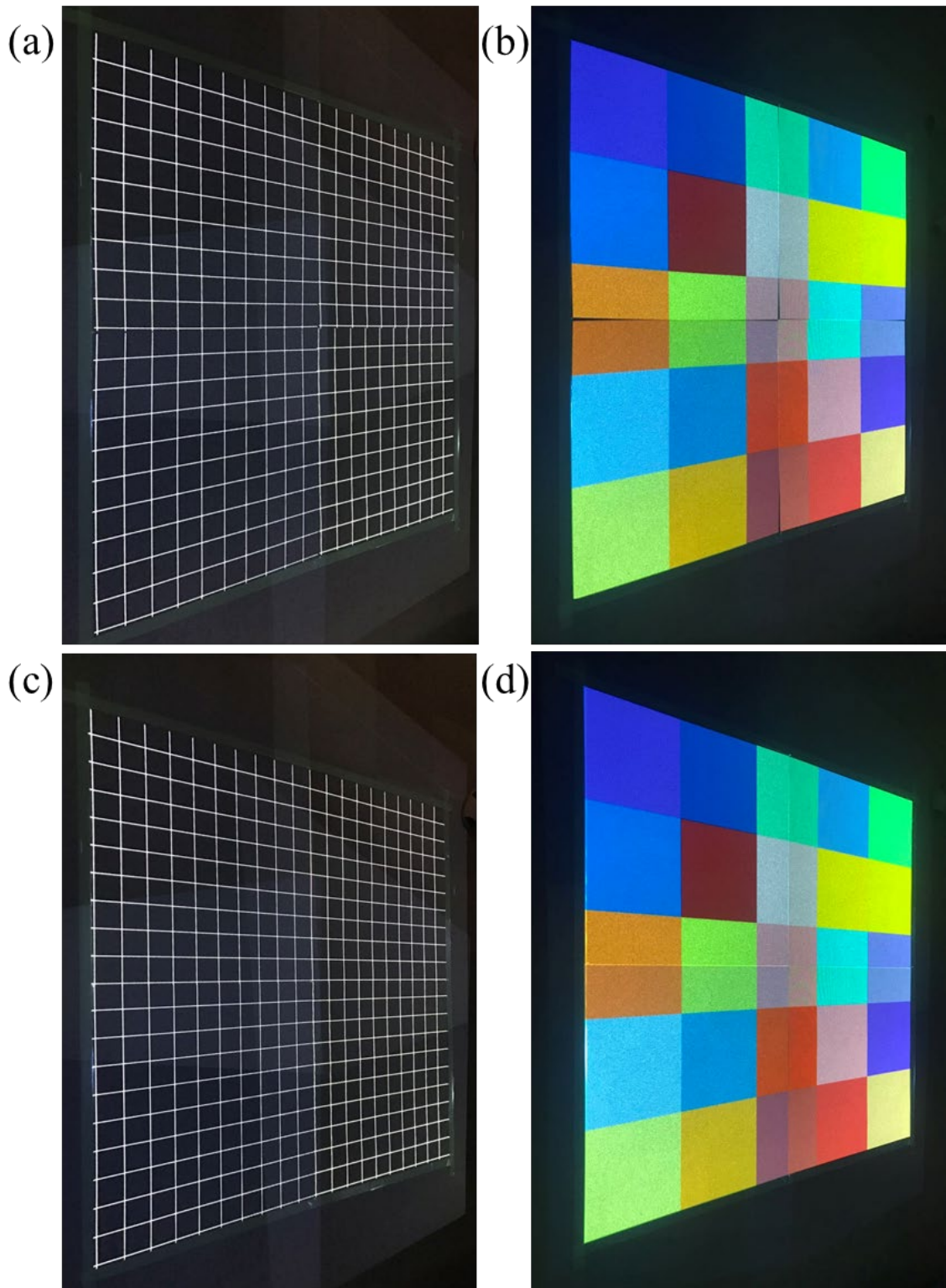


Figure 3-8. The projection results before(a,c) and after(b,d) calibration.

Chapter 4

A Projector Calibration Method for Arbitrary Shaped Surface Using a Mobile Camera

Building a PROCAMS requires the camera being installed and configured such that the projected patterns can be observed, and the pose of the camera and the projector must remain unchanged during the calibration process. Fulfilling these requirements can be difficult in some scenarios, as described in Section 2.3.3. In some real-world cases, installing cameras just for system deployment is considered costly and inefficient as they are usually not used in the actual SAR application.

In this chapter, a new projector calibration method is proposed to solve those problems. The significant difference between the proposed method and conventional methods is that a mobile camera instead of one or several static cameras is used. The characteristics of the proposed method are summarized as follows:

- It allows the users to use any mobile camera such as a built-in camera of a smartphone to carry out projector calibration.
- It uses multiview images to improve triangulation angles while keeps scene completeness.

- Two sets of features - the local image features and the pattern features are used complementarily to achieve a robust calibration result.
- The method uses a single-shot SL pattern. The coding and decoding scheme of the SL is improved to achieve robust decoding result without knowing the calibration data in advance.

4.1 Structured Light Patterns

According to the coding domain, SL techniques can be classified into two categories: temporal coding SL and spatial coding SL. Temporal coding SL [12] [13] [24] [28] - [32] can achieve dense projector-camera correspondences and dense 3D reconstruction since the depth map is generated pixel by pixel. However, due to the strict static requirement of the time-domain coding, it requires that the PROCAMS must be stationary. Therefore, it is not viable for a PROCAMS whose camera pose is dynamic. In contrast, spatial coding SL [33] - [40] allows recovery of depth map with only one pattern coded in the space domain, for which reason it is also referred as single-shot SL. Although most of spatial coding SL originally assume a calibrated and static PROCAMS to achieve accurate and robust acquisition of correspondences, they potentially allow the decoding of SL pattern without knowing the intrinsic and extrinsic parameters of the camera and the projector, as long as the rotation and distortion are reasonably small. However, as a group of pixels needs to be used as a codeword, the resolution of spatial coding SL is









usually lower than that of temporal coding SL, and the accuracy is limited, due to the errors caused by depth discontinuities, distortions, intensity changes, and noises.

A spatial coding SL which is based on the pattern proposed by H. Lin et al. [41] is used in the implementation to generate the projector-camera correspondences. It is a variant of the rhombus pattern proposed by Z. Song et al. [42]. To improve the robustness of the SL for an uncalibrated PROCAMS, the coding scheme and decoding algorithm are modified, and fail-safe features are added so that a feature has higher chance to be decoded correctly. A codeword correction step is also added to recover the wrongly decoded pattern features.

The rhombus pattern is generated from an eight-element pseudorandom array. The pseudorandom array has the unique window property which assigns a unique code word to each pattern feature. It is constructed by folding a pseudorandom sequence which is generated by a Galois Field with eight basics or $GF(8) = \{0, 1, a, a^2, \dots, a^6\}$ with $a^3 + a + 1 = 0, a^7 = 1$ [43]. The size of the array is 65×63 , and the window size of its unique window property is 2×2 , which means for any 2×2 subarrays, the permutation of the four elements in it is unique along the whole array. Replacing the basics of $GF(8)$ with the corresponding geometrical elements, as shown in Table 4-1, the projection pattern can be generated. An example of the pattern image is shown in Figure 4-1. The pattern feature points are defined as the intersection points between adjacent pattern elements. In the proposed method, it is assumed that the four colors R, G, B, and K used in the pattern can be correctly distinguished in general.

The significant advantages of the rhombus pattern over other spatial coding patterns, such as [38], [39], and [40], are that it has a smaller unique window size and greater coding capacity. Additionally, the colors, as well as the arrangement and geometry of rhombus elements, provide richer and more discriminable local image features. It is considered an advantage because not only pattern features but also local image features are used in the proposed method.

Table 4-1. Elements of the pseudorandom array

Basic	0	1	a	a^2	a^3	a^4	a^5	a^6
Pattern	R	G	B	K	R'	G'	B'	K'
Element								

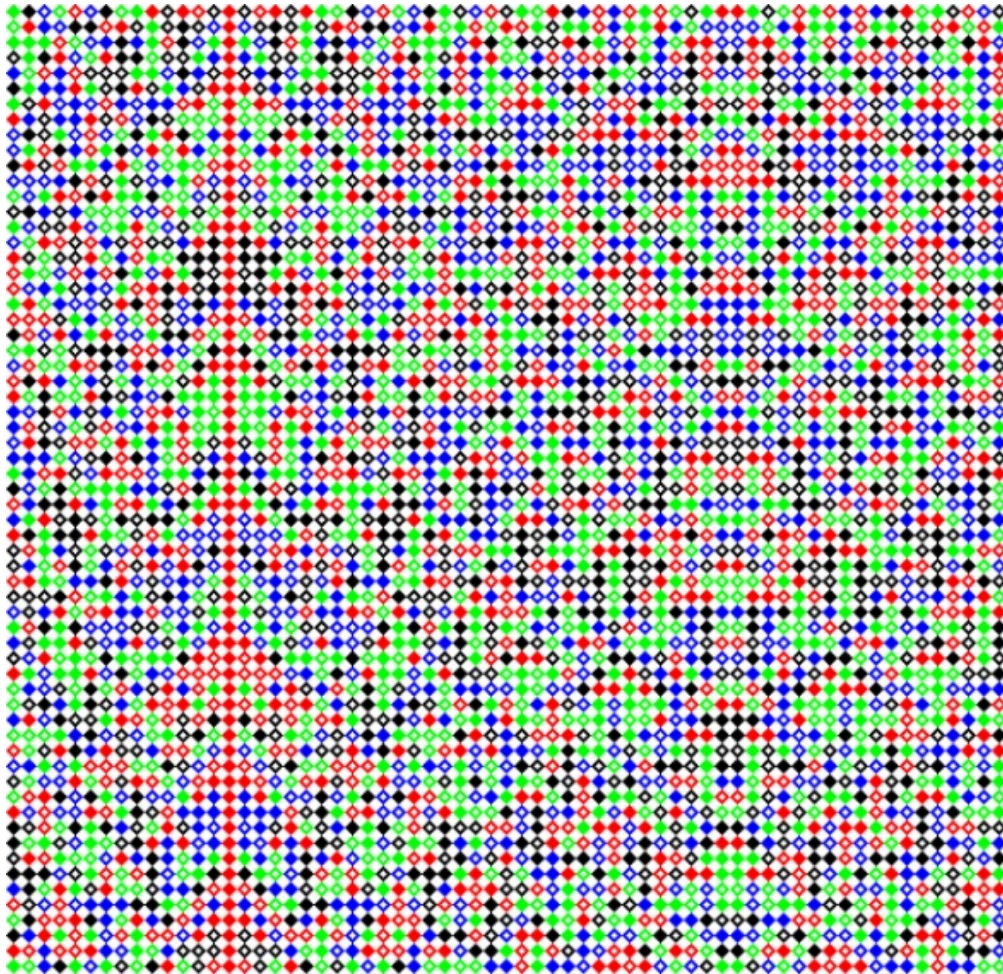


Figure 4-1. An example of the spatial coding SL pattern used in the proposed calibration method.

4.2 Fail-Safe Design of Rhombus Pattern

Introducing fail-safe features to SL can improve the robustness of decoding results as the correct information has a higher chance of being included in the codewords. This is important in real-world cases where the error rate of decoding is high. Unlike conventional methods that assume calibrated PROCAMS, the calibration information is not available to filter the incorrect features. However, they can be removed later by geometry guided feature matching among multiview camera images or revised by a codewords correction step.

4.2.1 Fail-Safe Coding Scheme

As shown in Figure 4-2, there are two types of pattern feature points called P1 and P2. P1 type points have two elements on the left and right, while P2 type points have two elements on the top and bottom. Originally, the codeword of both P1 and P2 are defined by four elements as $c1-c2-c3-c4$, which means a feature point is only encoded by its two neighbor elements and the two elements below for P1 or the two elements at the right side for P2. Take P1 as an example; if the decoding of either $c3$ or $c4$ failed, P1 would also fail. However, due to the unique window property, $c5-c6-c1-c2$ can also determine P1. In order to improve the robustness, the fail-safe feature is implemented by modifying the coding scheme so that the codeword also includes the two elements above P1. The modified codeword has six elements: $c5-c6-c1-c2-c3-c4$. A pair of pattern features are considered as matching if the type, as well as the first or last four digits of their codewords,

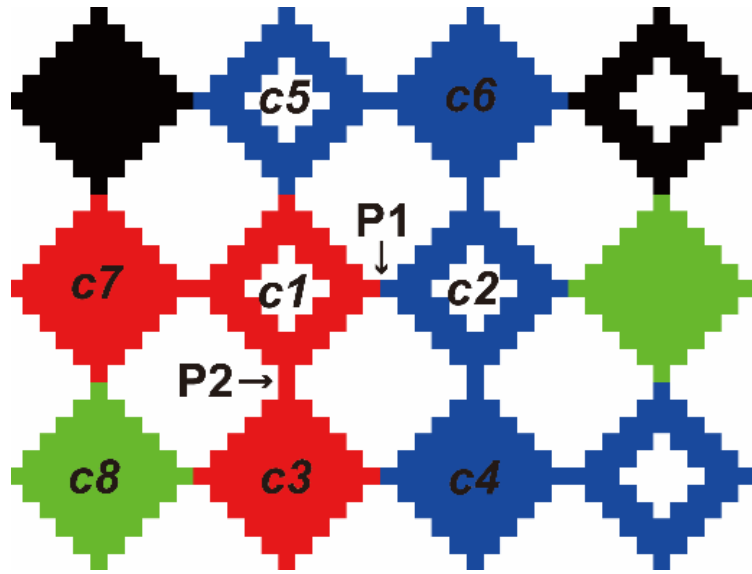


Figure 4-2. The two feature types and their codewords.

are the same. For P2 type features, the left two elements $c7$ and $c8$ are included, and the new codeword of P2 is $c7-c8-c1-c3-c2-c4$.

4.2.2 Fail-Safe Decoding Algorithm

Since the unique window size is 2×2 , and each feature point is associated with two rhombus elements, two adjacent feature points of the same type, referred to as pair-point, can determine a unique window and the corresponding codeword. The original rhombus pattern uses four-element-codeword, so it only searches the bottom pair-point for P1 and the right pair-point for P2. In the improved six-element-codeword design, P1's top pair-point and P2's left pair-point are also searched for, because the top or left codewords can also be used to identify a feature point. Ideally, the first and the last four digits of the codeword should refer to the same feature point. In practice, as the extraction of some feature points may fail, pair point searching can give wrong results. Moreover, element recognition may also fail due to some factors like distortions or intensity changes.

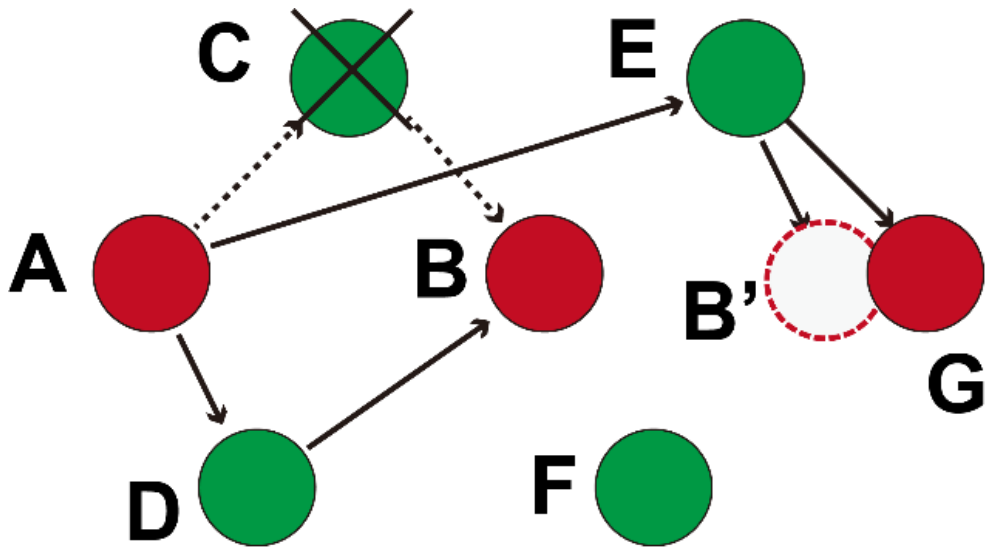


Figure 4-3. The pair-point searching strategy and a failure example.

Consequently, the first and the last four digits may refer to different feature points. However, thanks to the fail-safe coding, a pattern feature can be preserved as long as one of them is correct.

Figure 4-3 demonstrates the searching strategy of pair-point. The two feature types are shown in red and green dots, respectively. In [41], in order to make the searching more robust to local distortions, the position of the right pair-point of **A** is not searched by finding the nearest red neighbor. Instead, it is determined by firstly finding the upper-right and lower-right green feature points, which are point **C** and **D**, and then finding the lower-right red dot of **C** and the upper right red dot of **D**. Ideally, the two searching results refer to the same dot **B**. In case they refer to different dots, the position of **B** is estimated by $\hat{P}_B = P_C + P_D - P_A$, where P_C , P_D and P_A are the coordinates of point **C**, **D** and **A**, and \hat{P}_B is the expected position of **B**. Then the actual position of **B** is determined by finding the nearest red point to \hat{P}_B .

Obviously, this searching strategy only works in an ideal situation where both **C** and **D** have been correctly extracted. In practice, if the extraction of **C** failed, the upper-right green point of **A** will be **E**. In this case, \hat{P}_B will be the position of **B'** whose nearest red dot is **G** instead of **B**. Following this strategy, the searching of **A**'s pair-point fails although it could have succeeded by finding the upper-right green point of **D**.

To solve the problem, both **B** and **G** are kept as the pair-point of **A** and two codewords are found and decoded accordingly. It is another fail-safe mechanism in the modified decoding algorithm. A pair-point can be found as long as one of the searching routes gives the correct result.

4.3 Robust Projector Calibration Using a Mobile Camera

Unlike the conventional PROCAMS, where the pose of both the projector and the camera are constant during calibration, the camera pose changes from frame to frame when using a mobile camera. Also, it is assumed that the users have no prior knowledge about the geometry of the projection surface. In the proposed method, a reliable point cloud of the projection surface is first generated using a set of multiview images captured by the mobile camera. Then a 3D-2D matching is performed between the point cloud and the projector image, from which the intrinsics and extrinsics of the projector were estimated.

4.3.1 Point Cloud Generation

The reconstruction of 3D objects from multiview images is known as Structure from Motion (SfM). Many software and libraries like VisualSfM [44], COLMAP [45], and openMVG [46] provided robust SfM pipelines that generate reliable reconstruction results. Usually, SfM relies on the detection and matching of local image features such as SIFT [47], SURF [48] and AKAZE [49] to build correspondences among different views. However, the object may lack textures, which makes it difficult to extract sufficient features for 3D reconstruction. This problem can be solved by projecting the rhombus SL pattern introduced in section 4.2 onto the object to add rich textures to its surface.

From the projected pattern, two sets of features are extracted for reconstruction: the local image features and the pattern features. The local image features are used to build correspondences between the camera images, while the pattern features are used to build correspondences between projector and camera images. Although the rhombus pattern has its pattern features encoded, feature extraction and decoding accuracy can be significantly affected by factors like distortion, depth discontinuity, and illumination, and therefore the correspondences usually contain many outliers. Conventionally, an epipolar-geometry-based method can be used to remove the outliers. However, it requires that at least the fundamental matrix between the camera and the projector is known in advance. It is not the case in the proposed method at this point. Therefore, more robust and reliable local image features, rather than the pattern features, are used when generating the initial point cloud. In the implementation, SIFT is used as local image feature. In contrast, pattern features produce more reliable correspondences than local

image features between the projector and camera because the appearance of the pattern in their images usually differs significantly due to the influence of many factors like brightness, contrast, blur, and white balance.

An incremental Structure from Motion pipeline [50] is used to generate the initial point cloud as well as the intrinsic and extrinsic parameters of camera views. Although the rhombus pattern projection can provide relatively dense and discriminable SIFT features thanks to the four-color-element and a large number of corners, the similarity of features is not completely ignorable. In order to reduce the number of false-positive matchings, the Nearest Neighbor Distance Ratio (NNDR) test is performed, where a rather restrictive threshold value (0.6) is used. Furthermore, it can be assumed that images captured by the same mobile camera with the same zoom ratio share the intrinsic parameters. It helps the SfM pipeline to produce more stable results but is not essential.

Then, the point cloud is regenerated using a structure estimation from known poses algorithm [46]. As the intrinsics and extrinsics of the camera views have been estimated in the previous step, geometry guided feature matching can be applied along with NNDR test. As a result, many features that were removed due to false-positive matching during the initial point cloud generation can be restored in the new point cloud. A feature filtering step is also applied to remove unreliable features by checking their visibility (i.e., the number of views that observed the feature), the angles among the views containing the feature, as well as the reprojection error and chirality of the triangulated 3D points. The newly generated model is further optimized in a bundle adjustment (BA) step [51].

Next, the optimized camera intrinsics and extrinsics are used to generate the point cloud for pattern features. The algorithm is similar to before. The difference is that the NNDR test is not needed as the feature similarity is not a problem here; instead, a pair of feature points is considered as matching if their fail-safe codewords match. The codewords are also bound to the corresponding 3D points. Using the camera intrinsics and extrinsics estimated previously, most of the outliers of pattern features caused by decoding error can be removed during the geometry guided feature matching. The point cloud is further refined by applying BA. This time the calibration data of all camera views are set to constant, and only the 3D points are refined.

4.3.2 Codeword Correction

Distortions and intensity changes can cause decoding error of pattern feature; therefore, a feature may be decoded to the same but wrong codeword in several views that are close to each other. In such a case, these feature points will build valid correspondences between camera images and be triangulated to a 3D point with an approximately correct coordinate. Discard these points will lose information and cause biased view selection for 3D reconstruction. In order to solve this problem, a simple method is introduced below to correct their codewords.

The codewords bound to the 3D points of pattern features can help to generate the 3D-2D correspondences between the point cloud and the projector image. From these correspondences, a 3×4 projection matrix \mathbf{P} can be estimated robustly using A-Contrario RANSAC (AC-RANSAC) [52] and direct linear transform (DLT) [53]. The correspondences involving 3D points that hold wrong codewords are not supported by \mathbf{P}

and can be detected as outliers. To correct the codewords, the projection coordinate of the 3D point on the projector image plane is calculated with the following equation:

$$\lambda x = PX \quad (4-1)$$

where x and X are the homogenous coordinates of the 2D and 3D points, respectively. Next, the codeword can be found by searching for the pattern feature that is closest to x and has the same feature type as the 3D point on the projector image plane. After that, the codewords of 2D points on the corresponding camera views are corrected. Finally, the point cloud of pattern features is regenerated, and the projection matrix P is estimated again.

4.3.3 Projector Calibration

As shown in (4-2), the projection matrix P is a composition of the projector's intrinsic matrix K , the rotation matrix R , and the translation vector t . According to [54], P can be decomposed to K , R , and t by RQ decomposition. Assuming that the projector pixels are square and there is no axis skew, K can be revised to K' by (4-3). Then K' , R and t are refined, and a lens distortion parameter is estimated through a non-linear optimization step which minimizes the reprojection error from the 3D point cloud to the projector image.

$$P = K[R|t], K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4-2)$$

$$K' := \begin{bmatrix} (f_x + f_y)/2 & 0 & c_x \\ 0 & (f_x + f_y)/2 & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4-3)$$

Adding the pattern features and the optimized calibration data of the projector image, the point cloud of pattern features is regenerated in the same fashion as described in the previous section. Finally, the point clouds of pattern features and local image features are combined, and a full BA optimization is applied to optimize all 3D points and views followed by an outlier removal step similar to [17]. The overall process flow is shown in Figure 4-4.

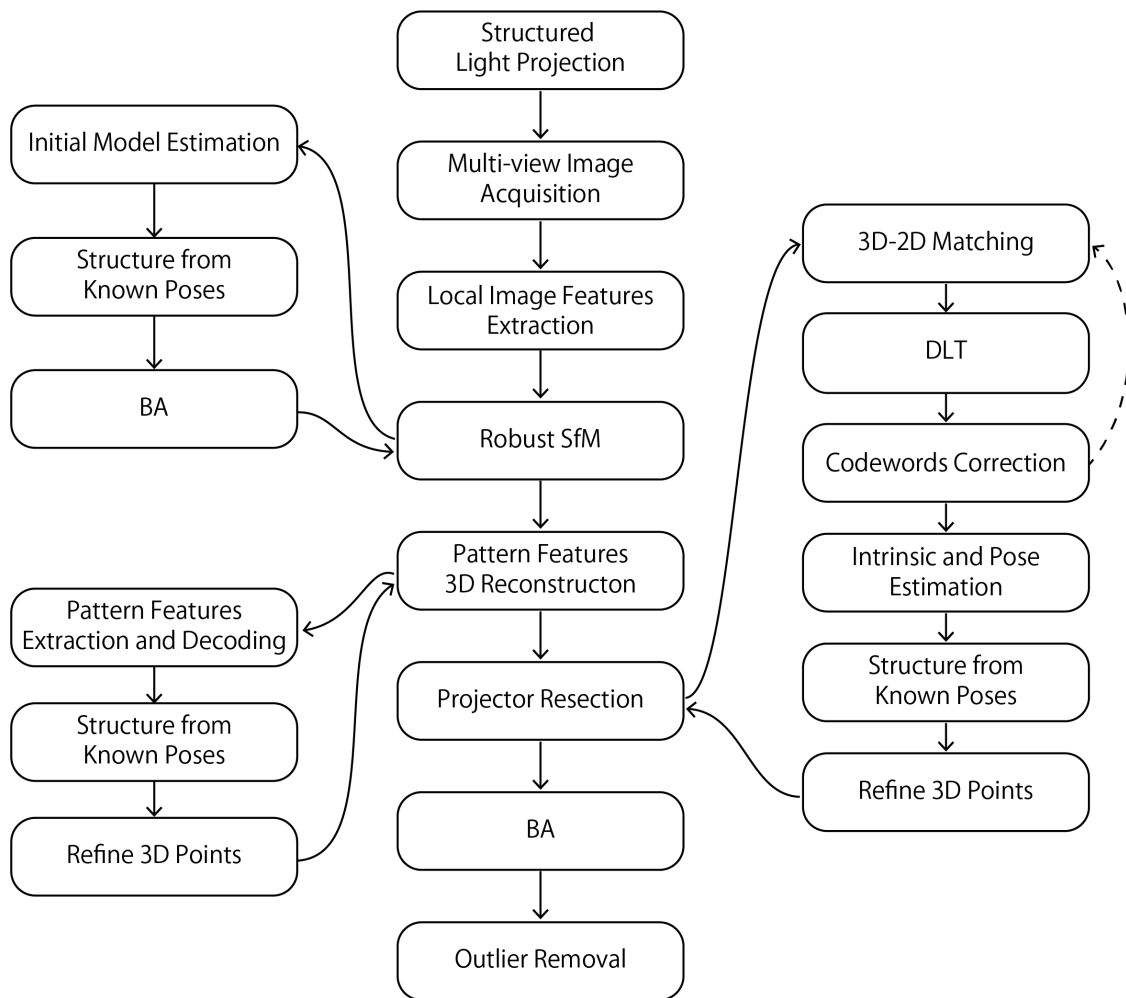


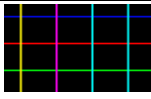
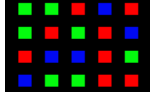


Figure 4-4. The flow diagram of the proposed calibration algorithm

4.4 Evaluation of Local Image Features

This experiment evaluated the quality of the image local features extracted from different single-shot SL projections in terms of quantity and discriminability. Four representative patterns were compared: the four-color rhombus pattern [41] that is used in the proposed calibration method, a monochrome geometrical primitives pattern proposed by C. Albitar et al. [38], a six-color line pattern proposed by J. Salvi et al. [39], and a three-color square pattern proposed by D. Desjardins and P. Payeur [40]. The patterns were projected on to a white paper, and a pair of images were captured from two constant views for each pattern. The captured images of one view are shown in Figure 4-5. A standard SIFT feature extraction and matching algorithm with NNDR test was then applied on each pair. After that, the number of matches, and the average NNDR were compared. The results are shown in Table 4-2. The 4-color rhombus pattern gave the most total matches, the highest percentage of valid matches after NNDR tests. It also gave the smallest average NNDR, indicating that its SIFT features are more discriminable than those of the other three patterns.

Table 4-2. A comparison of the quality of SIFT features extracted from four spatial coding structured light patterns in terms of quantity and discriminability

	#Matches total	#Matches (%) NNDR < 0.8	#Matches (%) NNDR < 0.6	Avg. NNDR
 <p>4-color rhombus</p>	41494	18345 (44.2)	10004 (24.1)	0.764
 <p>Monochrome primitive</p>	11751	4154 (35.4)	1763 (15.0)	0.809
 <p>6-color line</p>	7468	1081 (14.5)	240 (3.2)	0.895
 <p>3-color square</p>	26619	4419 (16.6)	1335 (5.0)	0.886

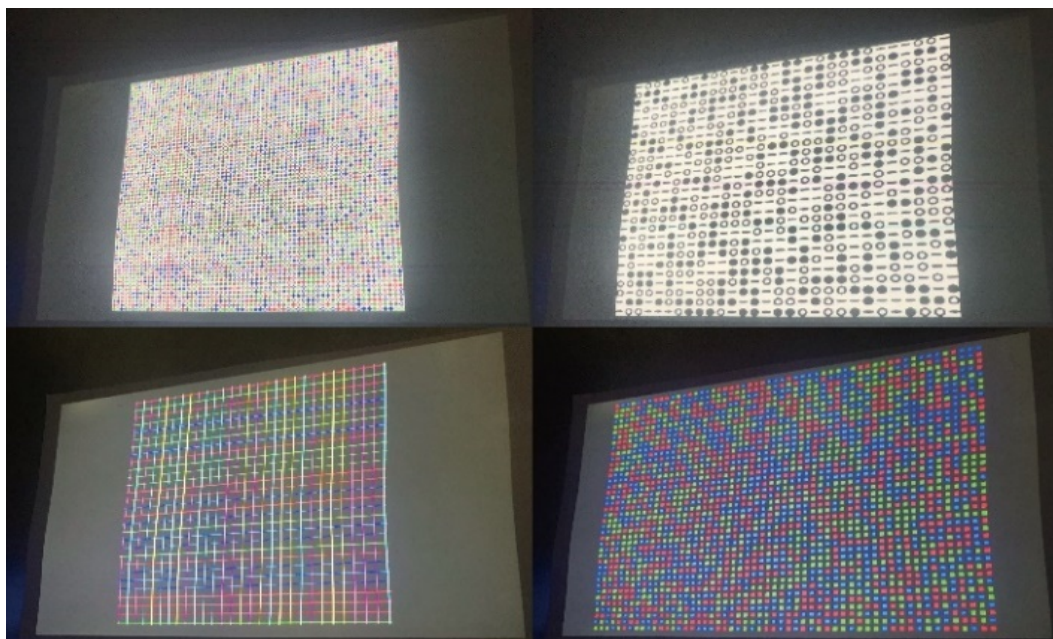


Figure 4-5. The images that are used to evaluate the local image features.

4.5 Evaluation in Close-Range Scenes

4.5.1 Datasets

The proposed method was evaluated on three close-range datasets. Dataset #1 is a plaster sphere. The projection suffers from perspective distortion at areas near to the projection edge, and it makes the pattern feature decoding very difficult. Dataset #2 is a plaster prism. Although the perspective distortion does not affect the pattern feature significantly, decoding failures occur on the surfaces orienting upward due to the reflection. Dataset #3 is a foam face. The rough surface introduces many noises which significantly affect pattern feature detection and decoding. Figure 4-6 shows the experiment setup of dataset #1, #2, and #3. Some images from these datasets are shown in Figure 4-7. A summary of the datasets and their calibration results is shown in Table 4-3.

Table 4-3. Summary of the close-range datasets and calibration results

Dataset	#Cam views	Surface shape	Surface material	Reproj. err. (all views)	Reproj. err. (projector)
#1	18	sphere	plaster	0.51309	0.07966
#2	24	prism	plaster	0.71076	0.12370
#3	25	face	foam	0.72443	0.16634

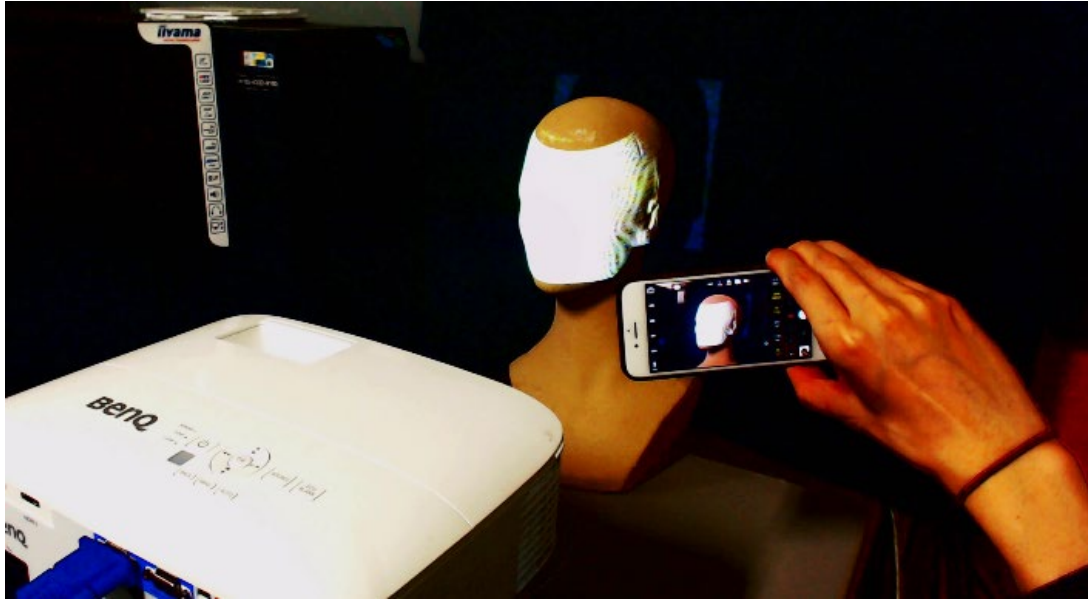


Figure 4-6. Experiment setup of dataset #1, #2 and #3.

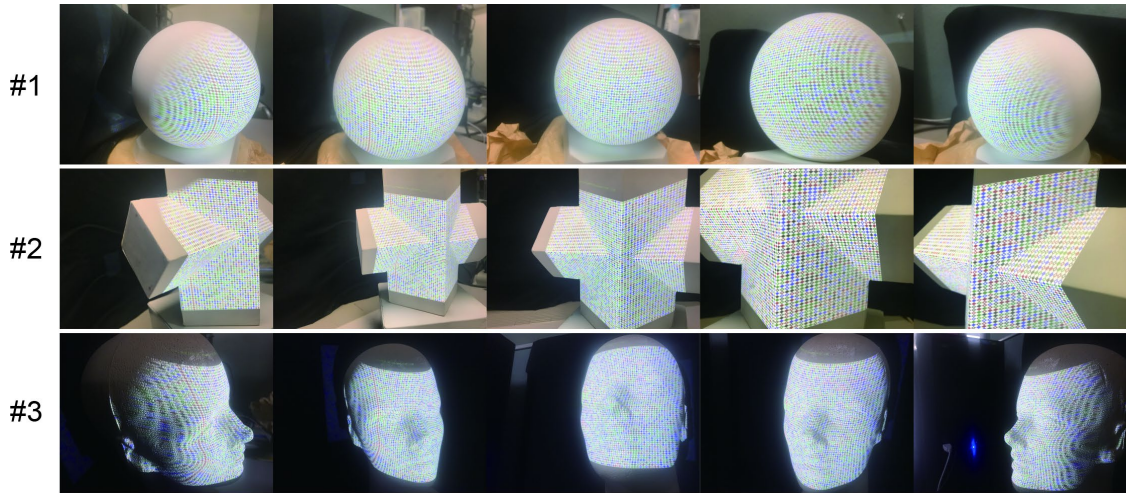


Figure 4-7. Sample images of the close-range experiment datasets.

Table 4-4. Comparison of the estimations of focal length and principle point of the projector made by the proposed method (on datasets #1, #2 and #3) and a checkerboard (CB) method.

Dataset	Focal length	Principle point	Reprojection error of projector
#1	2064.052	654.654, 718.195	0.07966
#2	2113.417	650.388, 754.812	0.12370
#3	2066.189	650.526, 744.187	0.16634
CB	2097.734	639.026, 736.925	0.38304

4.5.2 Comparison to Checkerboard-based Calibration

Another calibration experiment was conducted on the same projector as being used in dataset #1, #2, and #3 using a checkerboard-based method proposed by D. Moreno and G. Taubin [10]. The reprojection error was calculated from the images that were used for calibration (i.e., Gray code patterns projected onto a checkerboard). A comparison of the results is shown in Table 4-4. The results of focal length and principal point are close to each other, and the differences are reasonable, indicating that both methods comparable in terms of accuracy. The proposed method also produced less reprojection error of projector comparing to [10].

4.5.3 Projection Surface Reconstruction Results

The 3D models generated from the reconstructed point cloud of the datasets are shown in Figure 4-8. It can be seen that almost all the area covered by projection is reconstructed regardless of the occlusions in some of the camera images. In all datasets, the reconstructed models reproduced the objects well, indicating that the calibration results are reliable.

4.5.4 Reconstruction Precision

This experiment evaluated the reconstruction precision of dataset #1 via least-square sphere fitting. Because the result of the proposed method is up-to-scale, the point cloud must be recovered to the absolute scale. The procedure is as follows: first, a least square sphere fitting was first applied to the point cloud of dataset #1; then the point cloud was scaled together with the sphere such that the sphere has the same size as the real object, which is 130mm; finally, the root mean square error (RMSE) of sphere fitting was calculated. The result RMSE was 0.109 mm in the absolute scale.

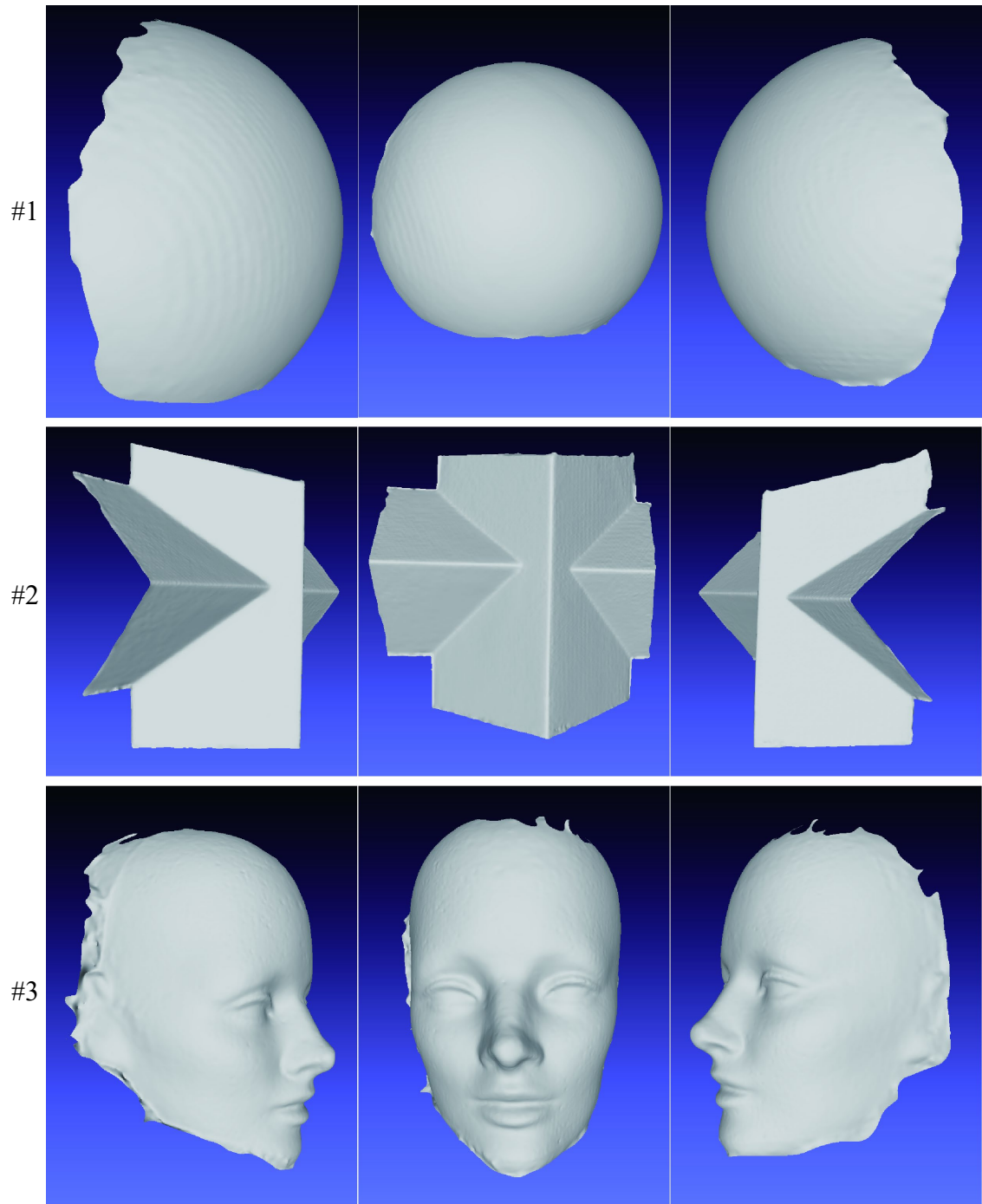


Figure 4-8. Reconstruction results of dataset #1, #2 and #3.

4.6 Evaluation in Large-Scale Scenes

4.6.1 Datasets

The scenes of dataset #4 and #5 are in a large-scale indoor space designed for immersive SAR [4] where the height of projectors is 7.7m away from the ground. Dataset #4 is a rounded wall corner with a height of 7.7m. Dataset #5 contains four objects of different shapes placed on the ground. The projection area is about $8\text{m} \times 8\text{m}$. Figure 4-9 illustrates the scenes of dataset #5 and #6. Some images from these datasets are shown in Figure 4-10. A summary of the datasets and their calibration results is shown in Table 4-5.

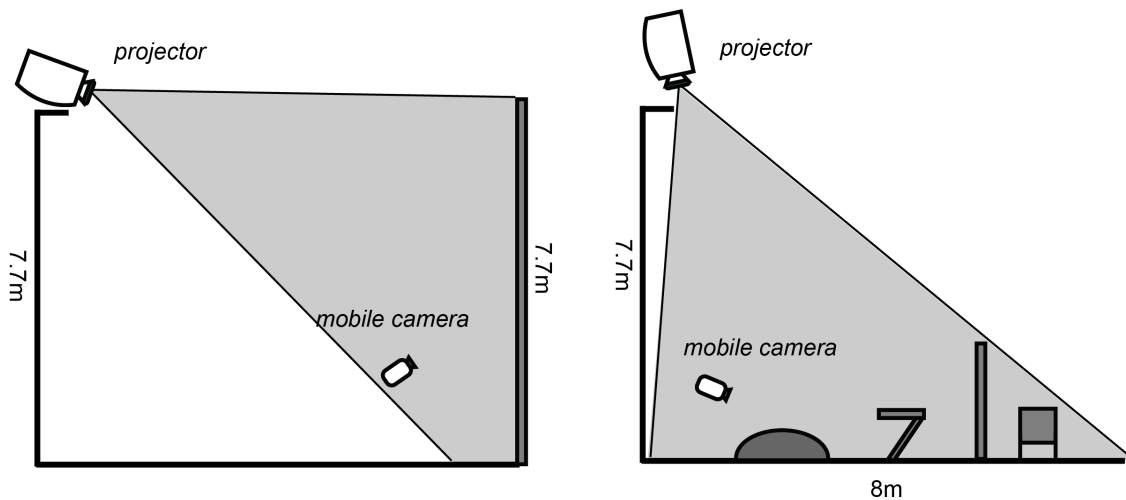


Figure 4-9. Side views of the scene in dataset #4 (left) and #5 (right). The position of the camera is an example to show its approximate height and relative position to the projection target(s).

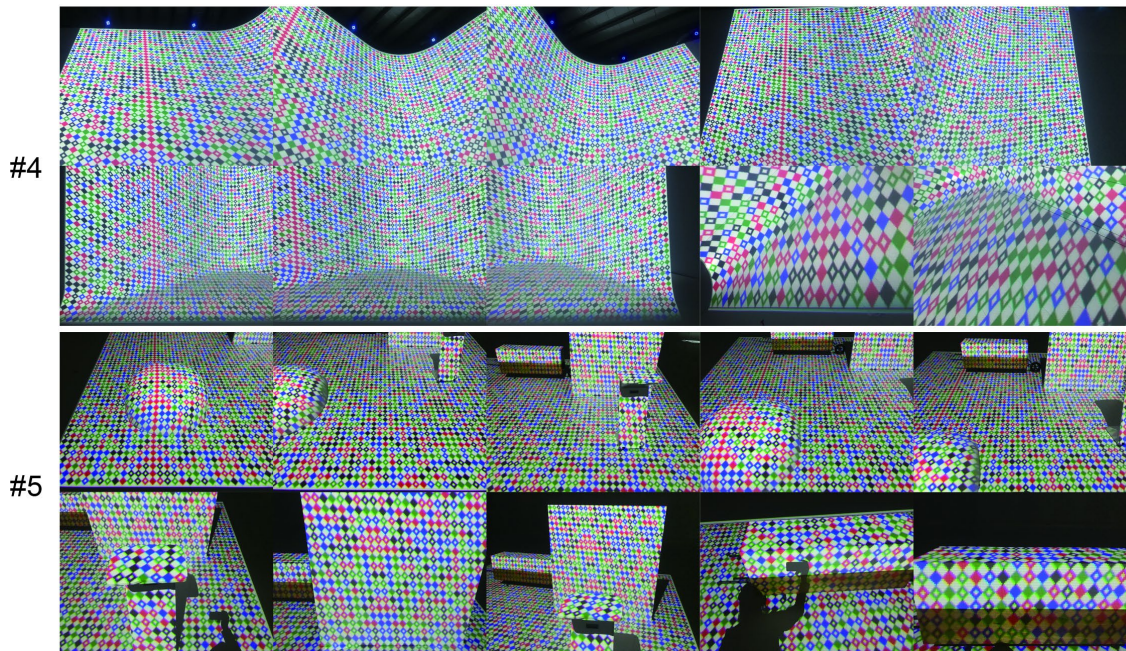


Figure 4-10. Sample images of the large-scale experiment datasets.

Table 4-5. Summary of the large-scale datasets and calibration results

Dataset	#Cam views	Surface shape	Surface material	Reproj. err. (all views)	Reproj. err. (projector)
#4	43	curved wall corner (large scale)	mixed	1.09439	0.23287
#5	146	mixed objects (large scale)	mixed	1.13155	0.22849

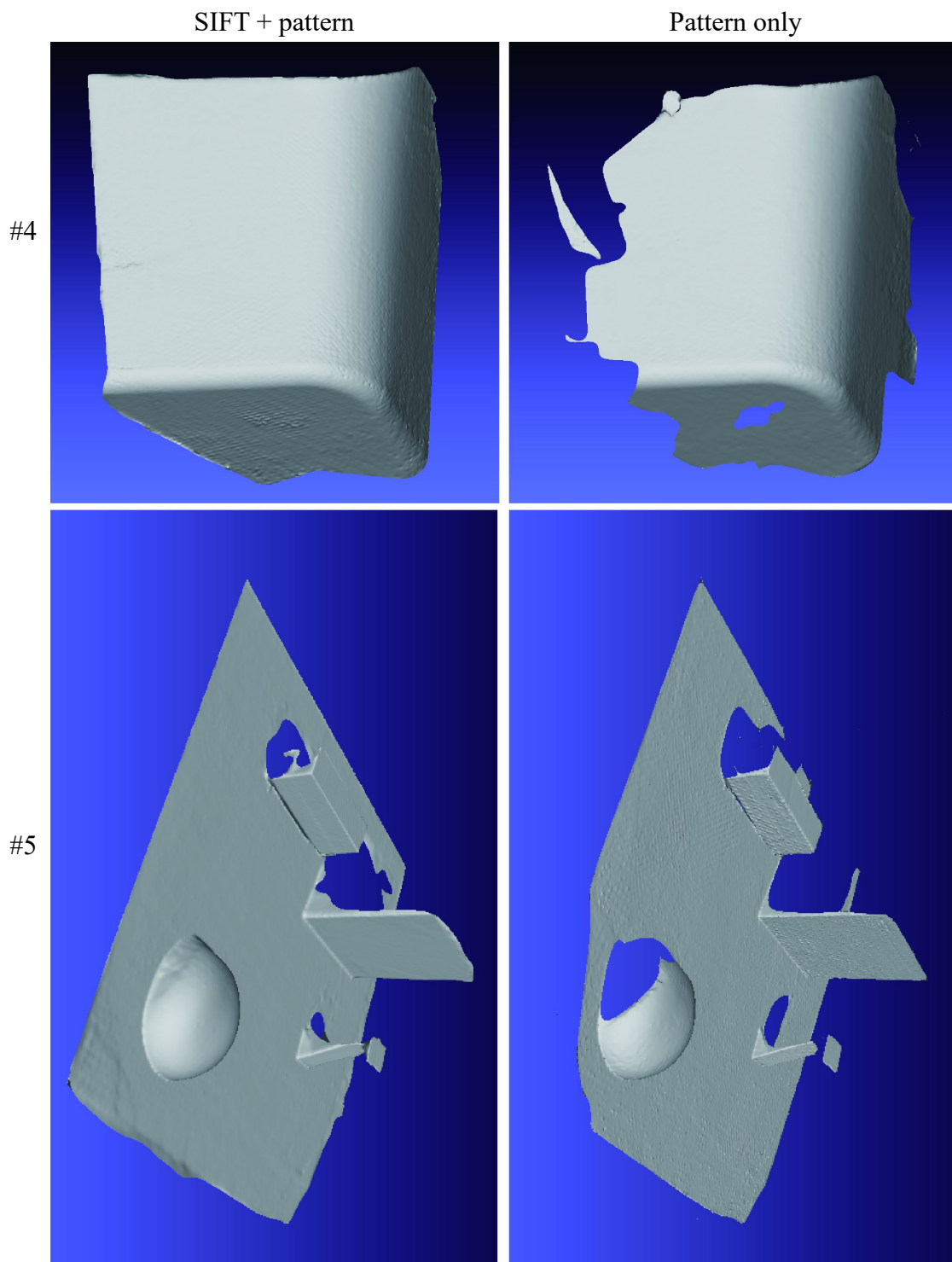


Figure 4-11. A comparison of reconstruction results using both SIFT and pattern features, and only pattern features for dataset #4 and #5.

4.6.2 Surface Reconstruction Result

Figure 4-11 shows a comparison between the reconstruction results of using both features (SIFT and pattern features), and only pattern features for reconstruction. As can be seen, using only pattern features lower the scene completeness as some areas in the scene are difficult to decode.

4.6.3 Comparison to Self-Calibration using Stationary Camera

This section compares the calibration results between the proposed method and the self-calibration method [13], which uses a stationary camera as typical, at the same scene as dataset #5. The datasets stationary cameras were mounted at two different locations. The first one is on the ground (lower view) and the second one is next to the projector (higher view). A sample image captured from each of the locations is shown in Figure 4-12. As it is known that the method is sensitive to the initial value of projector's principal point, two initial values on each dataset of the two camera locations were tried: one is an empirical value ($projector_width \times 0.5$, $projector_height \times 0.8$), and the other one is the value estimated by the proposed method.

It was expected that the lower view (Figure 4-12 (a)) would produce occlusion, while the higher view (Figure 4-12 (b)) would produce larger depth errors in the reconstruction result. However, their method failed at estimating the focal length using Bougnoux's formula regardless of the camera location and the initial value of principle point. In contrast, the proposed method calibrated the projector and reconstructed the scene on dataset #5 successfully without any occlusion.

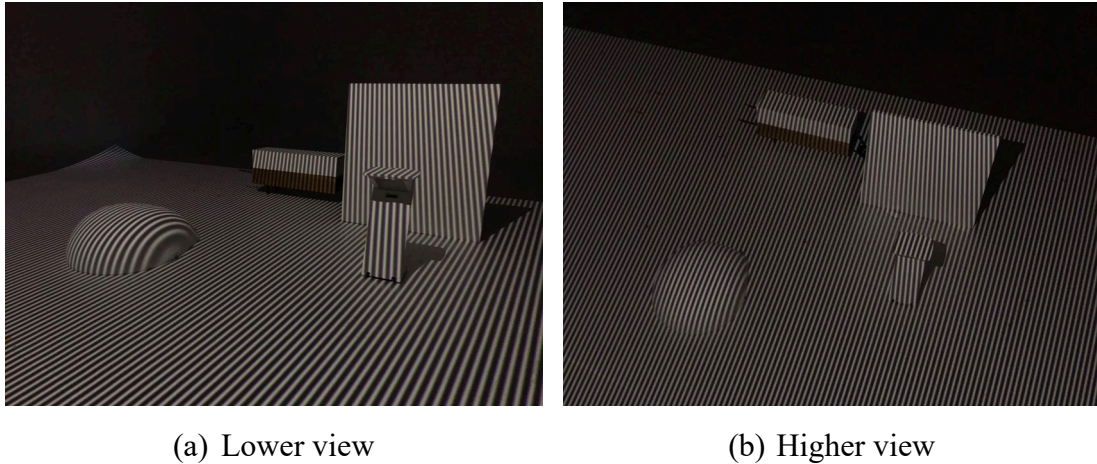


Figure 4-12. The camera views of the stationary camera used for self-calibration [13].

The self-calibration method failed because Bougnoux's formula requires the position of the principal point to be known. Furthermore, the computation of the focal lengths can be very sensitive to the assumed positions of the principal points. Unlike cameras whose principal point can be assumed as the image center in most cases, a projector's principal point can be affected by its offset and lens shift setup, which makes it very difficult to have a pixel-accuracy assumption. Additionally, the self-calibration method uses a conventional PROCAMS that only has two views (one projector and one camera). Therefore, the calibration fails completely once the focal length estimation of the projector-camera pair fails. In contrast, the proposed method uses multiple camera views. If the EXIF data or the camera spec is available, we can have an approximate calibration on camera intrinsic parameters, and then estimate the essential matrix from camera correspondences. The camera poses can then be estimated from the essential matrix, and after that, the scene can be reconstructed. Finally, the calibration data of the

projector can be estimated from the 3D-2D correspondences between the reconstructed points and the projector image, as described in previous sections. This approach has been proven effective by R. Hartley and C. Silpa-Anan [16]. In case approximate camera calibration is not possible, one can still assume the principal point of the camera as the image center can choose two camera views that give the most reliable result of Bougnoux's formula as the initial pair of reconstruction. This strategy is similar to [17].

Chapter 5

Conclusion

In this work, new methods were proposed to solve the problems of using PROCAMS to do projector geometric calibration for large-scale SAR systems.

It was first pointed out that most of the existing PROCAMS calibration approaches are difficult to be applied to large-scale space setup due to the large working distance, significant effect of lens distortion as well as the tradeoff between the baseline length and the quality of reconstruction. Two solutions were then proposed to address these problems in two different SAR scenarios.

For a planar projection surface, a solution was proposed to estimate the camera and projector lens distortion as well as projective distortion separately online. It uses laser line projection and “straight lines have to be straight” to achieve geometrically correct projection results. The calibration accuracy was evaluated under a real-world condition, which is a floor projection system in a school gym. An evaluation criterion was designed based on the max edge misalignment, the user's visual acuity, the viewing distance, the projector resolution, and the size of the projection area. The experiment results showed that the proposed method is effective. It produced less projection misalignment than a typical self-calibration method [13] and achieved the accuracy required in a real-world scenario. The proposed method registers each projector directly to the coordinate system

of the projection surface via the corresponding camera mounted on it. This approach requires the reliable acquisition of the correspondences between each camera and the projection surface, which may be difficult in some cases. An alternate approach is to register the projectors to one root projector, through a projector-camera and camera-camera homography tree, and then register the root projector to the projection surface. It requires the acquisition of correspondences between the projection surface and only the camera corresponding to the root projector. As a cost, it needs to obtain the correspondences between each camera and the neighboring projectors.

To further deal with more complex scenes where the geometry of the projection targets is unknown, an innovative method that allows a user to calibrate a projector via a mobile camera instead of a stationary one was proposed. It uses spatial coding SL projection to obtain the correspondences between the projector and the frames of a moving camera. Using the multiview images captured by the mobile camera, it achieved robust 3D reconstruction of the projection surface as well as a reliable estimation of the projector's intrinsic and extrinsic parameters. Comparing to conventional methods that use one or more stationary cameras, this method successfully solved the problem of the tradeoff between baseline length and reconstruction accuracy, and between baseline length and scene completeness. In the experiments, the proposed method produced comparable results with a checkerboard-based approach [10] in close-range scenes. For large-scale scenes, it successfully reconstructed the geometry of the projection surface and the estimated intrinsic and extrinsic parameters of the projector. At the same time, the self-calibration method [13] failed at estimating the focal length and thereby did not able to reconstruct the projection surface at all. A limitation of the proposed method is that it

works only when the color-coded SL pattern can be correctly recognized to some extent. If the projection surface is highly textured or saturated, the feature extraction and decoding will be difficult. As future work, the color-coded SL can be replaced by a shape-coded SL that can be accurately and robustly decoded without pre-calibration to deal with textured and saturated surface. Moreover, view planning method such as [55] can be applied in the future to assist a non-experienced user to move the camera.

The comparison of the self-calibration method and the two proposed methods are summarized as follows. Self-calibration method failed to produce satisfying result in large-scale SAR scenes because the camera position in the PROCAMS significantly affects the output of Bougnoux's formula (2-3) and the following estimation of intrinsic and extrinsic parameters. As has been shown in section 2.3, the depth uncertainty increase as the distance between camera and projector decrease, therefore results in an unstable estimation of the fundamental matrix F . On the contrary, larger projector-camera distance introduce larger chance of occlusion which lowers the scene completeness and biases the distribution of observable 3D points, thus affects the stability of F . Moreover, the principle point of projector p_p is usually unknown thus an empirical guess has to be made to use Bougnoux's formula which make its output more unreliable. The method introduced in Chapter 4 first performs a robust 3D reconstruction of the scene using multi-view images and SfM, then matches the 3D points to 2D points on the projector image, and finally estimates the projector's intrinsic and extrinsic parameter through RQ decomposition of the projection matrix P . As has been discussed in section 4.6.3, this method produces more robust calibration result because it requires no prior value of p_p , and the Bougnoux's formula is only used to estimate the focal length for a pair of camera

views (if EXIF is not available), instead of the projector-camera pair in the self-calibration method. Additionally, the result of this method relies on the accuracy of SfM, therefore, although the reprojection error is minimized, the lens distortions are not always been correctly estimated. The method introduced in Chapter 3 estimates the lens distortions correctly when the projection surface is planar. The straight laser lines projected on the plane provide direct ground truth for optimization of lens distortion parameters. Using “straight lines have to be straight” method, it is guaranteed that the projector lens distortion is corrected accurately so that the projected straight lines appear straight on the screen.

References

- [1] Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay, "Shader Lamps: Animating Real Objects With Image-Based Illumination," in *Rendering Techniques 2001*, pp. 89-102.
- [2] Taiki Fukiage, Takahiro Kawabe, Masataka Sawayama, and Shinya Nishida, "Animating Static Objects by Illusion-Based Projection Mapping," *Journal of the Society for Information Display*, vol. 25, no. 7, pp. 434-443, August 2017.
- [3] Toshiyuki Amano, "Projection based real-time material appearance manipulation," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 918-923.
- [4] Hikaru Takatori, Yuki Enzaki, Hiroaki Yano, and Hiroo Iwata, "Development of A Large-Immersive Display “LargeSpace”," *日本バーチャルリアリティ学会論文誌(VRSJ)*, vol. 21, no. 3, pp. 493-502, 2016.
- [5] R. Raskar, J. Baar, and JX. Chai, "A Low-Cost Projector Mosaic with Fast Registration," in *Asian Conference on Computer Vision(ACCV)*, 2002, pp. 114-119.
- [6] I. Martynov, JK. Kamarainen, and L. Lensu, "Projector calibration by ‘inverse camera calibration’," in *Image Analysis 2011*, 2011, pp. 536-544.
- [7] X. Fu, Z. Wang, H. Kawasaki, and R. Sagawa, "Calibration of Projector with Fixed Pattern and Large Distortion Lens in a Structured Light System," in *MVA 2013*, 2013, pp. 222-225.
- [8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330-1334, 2000.

- [9] A. Richardson, J. Strom, and E. Olson, "AprilCal: Assisted and repeatable camera calibration," in *IEEE International Conference on Intelligent Robots and Systems*, 2013, pp. 1814–1821.
- [10] Daniel Moreno and Gabriel Taubin, "Simple, Accurate, and Robust Projector-Camera Calibration," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 2012, pp. 464-471.
- [11] I. Din, H. Anwar, I. Syed, H. Zafar, and L. Hasan, "Projector calibration for pattern projection systems," *Journal of Applied Research and*, vol. 12, no. 1, pp. 80-86, 2014.
- [12] S. Inokuchi, K. Sato, and F. Matsuda, "Range imaging system for 3-d object recognition," in *International Conference on Pattern Recognition*, 1984.
- [13] S. Yamazaki, M. Mochimaru, and T. Kanade, "Simultaneous self calibration of a projector and a camera using structured light," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2011, pp. 60-67.
- [14] Jason Deglint et al., "Auto-calibration of a projector-camera stereo system for projection mapping," *Journal of the Society for Information Display*, vol. 24, no. 8, pp. 510-520, 2016.
- [15] Sylvain Bougnoux, "From projective to Euclidean space under any practical situation, a criticism of self-calibration," in *Sixth International Conference on Computer Vision*, Bombay, India, 1998, pp. 790-796.
- [16] Richard Hartley and Chanop Silpa-anan, "Reconstruction from two views using approximate calibration," in *Proc. Asian Conference on Computer Vision*, 2002, pp. 338-343.
- [17] Simon Will and Anselm Grundhofer, "Robust Geometric Self-Calibration of Generic Multi-Projector Camera Systems," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2017, pp. 42-51.

- [18] Berthold Klaus Paul Horn, *Robot Vision.*: The MIT Press, 1986.
- [19] Frédéric Devernay and Olivier Faugeras, "Straight lines have to be straight," *Machine Vision and Applications*, vol. 13, pp. 14-24, 2001.
- [20] Ago Molder, Olev Martens, Tonis Saar, and Rual Land, "Laser Line Detection with Sub-Pixel Accuracy," *Elektronika ir Elektrotechnika*, vol. 20, no. 5, pp. 132-135, 2014.
- [21] Tinku Acharya and Ping-Sing Tsai, "Computational foundations of image interpolation algorithm," *ACM Ubiquity*, vol. 8, 2007. [Online]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.2248&rep=rep1&type=pdf>
- [22] Martin A. Fischler and Robert C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, 1981.
- [23] Luis Alvarez, Luis Gómez Deniz, and Juan Rafael Sendra, "An Algebraic Approach to Lens Distortion by Line Rectification," *Journal of Mathematical Imaging and Vision*, vol. 35, pp. 36-50, September 2009.
- [24] Jens Guehring, "Dense 3-D surface acquisition by structured light using off-the-shelf components," *VProceedings of SPIE - The International Society for Optical Engineering*, vol. 4309, pp. 220-231, 2001.
- [25] Edmund Landolt, "Méthode optométrique simple," *Bulletins et Memoires de la Societe Francaise d'Ophtalmologie*, vol. 6, pp. 213-214, 1988.
- [26] Grimm W, Rassow B, Wesemann W, Saur K, and R Hiltz, "Correlation of optotypes with the Landolt Ring – a fresh look at the comparability of optotypes," *Optometry and Vision Science*, vol. 71, no. 1, pp. 6-13, 1994.

- [27] Joseph Mandelbaum and Louise L. Sloan, "Peripheral visual acuity with special reference to scotopic illumination," *American Journal of Ophthalmology*, vol. 30, no. 5, pp. 581-588, 1947.
- [28] Jeffrey L. Posdamer and Martin D Altschuler, "Surface measurement by space-encoded projected beam systems," *Computer Graphics and Image Processing*, vol. 18, no. 1, pp. 1-17, 1982.
- [29] Eli Horn and Nahum Kiryati, "Toward optimal structured light patterns," vol. 17, no. 2, pp. 87-97, 1999.
- [30] Dalit Caspi, Nahum Kiryati Kiryati, and Shamir Joseph, "Range imaging with adaptive color structured light," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 470-480, 1998.
- [31] Mitsuo Takeda and Mutoh Kazuhiro, "Fourier transform profilometry for the automatic measurement of 3-D object shape," vol. 22, no. 24, pp. 3977-3982, 1983.
- [32] Daniel Scharstein and Richard Szeliski, "High-accuracy stereo depth maps using structured light," in *Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition*, 2003, pp. 195-202.
- [33] Vuylsteke P. and A. Oosterlinck, "Range Image Acquisition with a Single Binary-Encoded Light Pattern," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 148 - 164, 1990.
- [34] Raymond A. Morano et al., *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 322-327, March 1998.
- [35] Thomas P. Koninckx and Luc Van Gool, "Real-time range acquisition by adaptive structured light," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 432-445, March 2006.

- [36] Hiroshi Kawasaki, Ryo Furukawa, Ryusuke Sagawa, and Yasushi Yagi, "Dynamic scene shape reconstruction using a single structured light pattern," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
- [37] X. Maurice, P. Graebling, and C. Doignon, "A pattern framework driven by the Hamming distance for structured light-based reconstruction with a single image," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2497-2504.
- [38] Chadi Albitar, Pierre Graebling, and Christophe Doignon, "Robust Structured Light Coding for 3D Reconstruction," in *2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007.
- [39] J. Salvia, Joan Batlle, and El Mustapha Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement," *Pattern Recognition Letters*, vol. 19, no. 11, pp. 1055-1065, 1998.
- [40] Danick Desjardins and Pierre Payeur, "Dense Stereo Range Sensing with Marching Pseudo-Random Patterns," in *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, Montreal, Que., Canada, 2007, pp. 216–226.
- [41] Haibo Lin, Lei Nie, and Zhan Song, "A single-shot structured light means by encoding both color and geometrical features," *Pattern Recognition*, vol. 54, pp. 178-189, 2016.
- [42] Zhan Song and Ronald Chung, "Grid point extraction and coding for structured light system," *Optical Engineering*, vol. 50, no. 9, 2011.
- [43] F.Jessie MacWilliams and Neil J.A. Sloane, "Pseudo-random sequences and arrays," *Proceedings of the IEEE*, vol. 64, no. 12, pp. 1715-1729, 1976.
- [44] Changchang Wu, "Towards Linear-time Incremental Structure From Motion," in *2013 International Conference on 3D Vision-3DV*, 2013, pp. 127-134.

- [45] Johannes L. Schonberger and Jan-Michael Frahm, "Structure-From-Motion Revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4104-4113.
- [46] Pierre Moulon, Pascal Monasse, and Renaud Marlet, OpenMVG. An Open Multiple View Geometry library.
- [47] David Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [48] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "SURF: Speeded Up Robust Features," in *European Conference on Computer Vision (ECCV)*, 2006, pp. 404-417.
- [49] Pablo F. Alcantarilla, Jesús Nuevo, and Adrien Bartoli, "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces," in *British Machine Vision Conference (BMVC)*, Bristol, UK, 2013, pp. 1281-1298.
- [50] Pierre Moulon, Pascal Monasse, and Renaud Marlet, "Adaptive structure from motion with a contrario model estimation," in *Asian Conference on Computer Vision*, 2012, pp. 257-270.
- [51] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon, "Bundle Adjustment — A Modern Synthesis," in *International Workshop on Vision Algorithms*, 1999, pp. 298-372.
- [52] Moisan Lionel, Pierre Moulon, and Pascal Monasse, "Automatic homographic registration of a pair of images, with a contrario elimination of outliers," *Image Processing On Line*, vol. 10, pp. 56-73, 2012.
- [53] Ivan E. Sutherland, "Three-dimensional data input by tablet," *Proceedings of IEEE*, vol. 62, no. 4, 1974.
- [54] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed.: Cambridge University Press, 2003.

- [55] Okan Erat, Markus Hoell, Karl Haubenwallner, Christian Pirchheim, and Dieter Schmalstieg, "Real-Time View Planning for Unstructured Lumigraph Modeling," in *IEEE Transactions on Visualization and Computer Graphics*, Beijing, China, 2019.
- [56] Chun Xie et al., "A Calibration Method of Floor Projection System for Learning Aids at School Gym," in *Third IEEE International Conference on Image Processing, Applications and Systems*, 2018, p. 6 pages.
- [57] Bimber Oliver and Ramsh Raskar, *Spatial augmented reality: merging real and virtual worlds*, AK Peters, Ed.: CRC press, 2005.
- [58] Mark R. Mine, Jeroen van Baar, Anselm Grundhofer, David Rose, and Bei Yang, "Projection-Based Augmented Reality in Disney Theme Parks," vol. 45, no. 7, pp. 32-40, 2012.
- [59] Kai Li and et al., "Building and Using a Scalable Display Wall System," *Computer Graphics and Applications*, vol. 20, no. 4, pp. 29-37, 2000.
- [60] V. Srinivasan, H.C. Liu, and M. Halioua, "Automated phase-measuring profilometry of 3-D diffuse objects," *Applied Optics*, vol. 23, no. 18, pp. 3105-3108, 1984.