Projector-Camera System Calibration and Non-planar Scene Estimation

by

Katherine Arnold

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2022

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of Contributions

A set of papers are captured within this thesis. I am the first author for each of these papers, contributing to problem definition, solution design, experimentation, data capture and results analysis, and paper writing.

1. [3] Arnold, K., Naiel, M. A., Lamm, M., & Fieguth, P. (2021). Evaluation of Solving Methods for the Fundamental Matrix Computation. *Journal of Computational Vision and Imaging Systems*, 6(1), 1–1. https://doi.org/10.15353/jcvis.v6i1.3563

   Excerpts from this paper appear in Chapter 1 Introduction, and Chapter 3 Problem Formulation.

Contributor	Statement of Contribution	
Arnold, K. (Candidate)	Conceptual Design 75%	Writing and Editing 83%
Naiel, M.	Conceptual Design 10%	Writing and Editing 10%
Lamm, M..	Conceptual Design 5%	Writing and Editing 2%
Fieguth, P.	Conceptual Design 10%	Writing and Editing 5%

2. [4] Arnold, K., Fieguth, P., & Lamm, M. (2022). Formulating a Moving Camera Solution for Non-Planar Scene Estimation and Projector Calibration. *Journal of Computational Vision and Imaging Systems*, 7(1), 10–12. https://doi.org/10.15353/jcvis.v7i1.4890

   Excerpts from this paper appear in Chapter 1 Introduction, Chapter 2 Background, and Chapter 3 Problem Formulation.

Contributor	Statement of Contribution	
Arnold, K. (Candidate)	Conceptual Design 85%	Writing and Editing 88%
Lamm, M..	Conceptual Design 5%	Writing and Editing 2%
Fieguth, P.	Conceptual Design 10%	Writing and Editing 10%

3. [5] Arnold, K., Fieguth, P., & Lamm, M. (2022). 30.3: A Moving Camera and Synthetic Calibration Target Solution for Non-Planar Scene Estimation and Projector Calibration. In press. SID Symposium Digest of Technical Papers.

   Excerpts from this paper appear in Chapter 1 Introduction, Chapter 2 Background, and Chapter 3 Problem Formulation.

Contributor	Statement of Contribution	
Arnold, K. (Candidate)	Conceptual Design 85%	Writing and Editing 93%
Lamm, M..	Conceptual Design 5%	Writing and Editing 2%
Fieguth, P.	Conceptual Design 10%	Writing and Editing 5%

Abstract

A projection mapping display system creates impressive 3D displays with light by mapping a 2D image from a calibrated projector onto a display surface. Projection mapping systems require that geometric information must be known about the projector, its spatial relationship to the display surface, and the surface itself. These relationships are constructed through observation of the projector and the display environment by a camera. The calibration process can be burdensome on the user, and different strategies will rely on prior information about the devices or upon enforcing display environment constraints. High capital costs are associated with generating a prior knowledge of cameras. Display environment constraints limit the range of possible display environments, in some cases requiring a 2D display surface, preventing non-planar 3D display environments. A self-calibration projector-camera(s) process that does not rely on known or fixed cameras, nor calibration targets, is highly desirable to increase both the ease of use and the range of possible environments for existing projection mapping systems.

This thesis develops a method for producing a geometric calibration estimate and 3D display surface estimate for non-planar projection mapping display environments. This approach assumes no prior information on the moving camera or fixed projector. Pixel correspondences relate observations across the camera and projector views, and are used to construct geometric relationships to produce a weak calibration estimate. Many applications of projection mapping technology involve artistic renderings that must be precisely mapped from 2D image projection to a 3D non-planar surface. The drafting of these artistic renderings often necessitates the existence of some prior virtual scene understanding. Limited scene understanding provides the basis for constructing virtual calibration targets to perform a geometric recovery of the weak calibration estimate recovery through bundle adjustment.

Experimental results show that the geometric calibration estimate observed no error in the estimated projector intrinsic parameters, and less than 2 degrees of average angular error in the estimated projector and camera poses when considering 2500 pixel correspondences with $\sigma = 1$ px additive Gaussian noise. The performance accuracy decreases with increasing noise in the pixel coordinates.

Acknowledgements

A deepest graditude must be extended to Dr. Paul Fieguth. A great deal of academic burnout haunted my undergraduate studies. Beyond just the support and expertise he had available for me throughout the challenge of remote mentorship, Dr. Fieguth has succeeded in nurturing a spark of excitement for exploration and understanding that I thought permanently lost.

I would like to thank Dr. Jonathan Kofman and Dr. Javad Shafiee for their time and effort serving as readers for my thesis and their insightful feedback that improved the final product.

This work would not have been possible without the support and skill offered by Mark Lamm and the engineering team Christie Digital Systems. Thank you for your time and knowledge, as well as facilitating the necessary equipment and space for my work.

A strong community supported me through a pandemic, graduate studies, and an entirely new field of study. I want to thank Sepehr Ghavam, Mohamed A Naiel, Saed Moradi, Benjamin Beelen, and Alexander Glover.

Dedication

This thesis is dedicated to my grandmother Leona and my mother Dina, two stubborn and determined women who each insist that I get these qualities from the other.

To Jim - sorry, you'll have to wait for the next one.

Table of Contents

List of Figures **ix**

List of Tables **x**

1 Introduction **1**

1.1 Motivation and Overview 3

1.2 Problem Statement and Objectives 5

1.3 Thesis Organization 6

2 Background **7**

2.1 Geometry of Image Projection 7

2.1.1 Reconstruction 9

2.2 Camera Projection 10

2.2.1 Two Views 12

2.2.2 Three-View Geometry 15

2.3 Self-calibration 16

2.3.1 Structured Light 17

2.3.2 Refining A Geometric Estimate 18

3 Problem Formulation **22**

3.1 Geometric Calibration Estimate 24

3.1.1 Moving Camera 24
3.1.2 Camera Estimation 25
3.1.3 Scene Triangulation and Projector Incorporation 29
3.2 Bundle Adjustment 29
3.2.1 Virtual Calibration Target 30
3.2.2 Triangulated Coordinates 31
3.2.3 Parameter Adjustment Flexibility 32

4 Geometric Camera Calibration **33**
4.1 Fundamental Matrix 34
4.2 Trifocal Tensor 35

5 Bundle Adjustment **37**
5.1 Short Bundle Adjustment: Camera and Virtual Calibration Target Coordinates 38
5.2 Full Bundle Adjustment: All Views and Estimated Scene Coordinates . . . 39

6 Experiments **42**
6.1 Experimental Procedure 42
6.2 Results 45
6.3 Discussion 47

7 Conclusion **55**
7.1 Future Directions 56

References **58**

List of Figures

1.1 Projection Mapping Diagram 2

1.2 Example of Projection Mapping on Historic Buildings 2

1.3 Diagram of a simple Projector-Camera System 4

2.1 Pinhole Camera Geometry Model 11

2.2 Epipolar Geometry Corresponding to Two Viewpoints 13

2.3 Epipolar Geometry Corresponding to Three Viewpoints 15

2.4 Illustration of Different Structured Light Methods 19

2.5 Example of Bundle Adjustment Minimization 20

3.1 Method Information Flow Diagram 23

3.2 Scene Diagram for Geometric Calibration of Projector-Camera System . . 28

3.3 Virtual Calibration Target 31

6.1 Non-Planar Display Scene Models Used in Synthetic Data 44

6.2 Non-planar Display Surface Estimates and Average Reprojection Error . . 50

6.3 Percent Error in Estimated Camera and Projector Intrinsic Parameters for Synthetic Data 52

6.4 Average Angular Error in Estimated Camera and Projector Pose Parameters for Synthetic Data 53

6.5 Reprojection Error in Display Surface Estimate for Synthetic Data 54

List of Tables

2.1 Considered geometries, their transformations, and their invariants. 10

3.1 Competing characteristics comparison of geometry- and calibration target-based approaches. 24

5.1 Comparison between the number of pixel correspondences (N) and performance of Structured Light Methods. 41

6.1 Percent error in estimated camera and projector intrinsic parameters for synthetic data 46

6.2 Estimated camera and projector angular rotation and translation error for synthetic data. 47

6.3 Comparison of Method Best Performance to Existing Methods 48

Chapter 1

Introduction

Projection mapping is a method that is used to turn common everyday objects into display surfaces. This technology is found in applications including 3D surface reconstruction; artistic displays used for education, entertainment, or cultural celebration; and augmented reality. A projection mapping system creates impressive visual displays with light by mapping a 2D image from a calibrated projector onto a display surface, Figure 1.1. These display surfaces may be a flat surface such as a projector screen, or they might be complex 3D surfaces such as historic buildings, Figure 1.2.

To have a projection mapping system a projector, some projection image(s) and a surface upon which to project are needed. Such displays are sensitive to the properties of the projector, and the display will appear very differently depending on where the projector is in relation to the scene. In order to produce highly accurate displays, the projector images are rendered in a 2D representation of the 3D display that will fit the projected light to the display surface. Generating this 2D rendering relies upon knowing the projection model of the system, and requires that geometric information must be known about the projector, its spatial relationship to the projection surface, and the surface itself.

Projection mapping methods require three primary components for geometric calibration: the projector(s) to be calibrated, the desired display environment including the display surface, and some way of measuring the scene. Scene measurement is often accomplished with cameras. The degree of knowledge of the measuring camera and the display surface can dictate the process of geometric calibration. Calibration methods may rely upon highly accurate cameras of known calibration. These cameras can be quite expensive, and raise the lifetime cost of implementing these projection mapping systems significantly while only being needed during calibration. Some calibration methods require

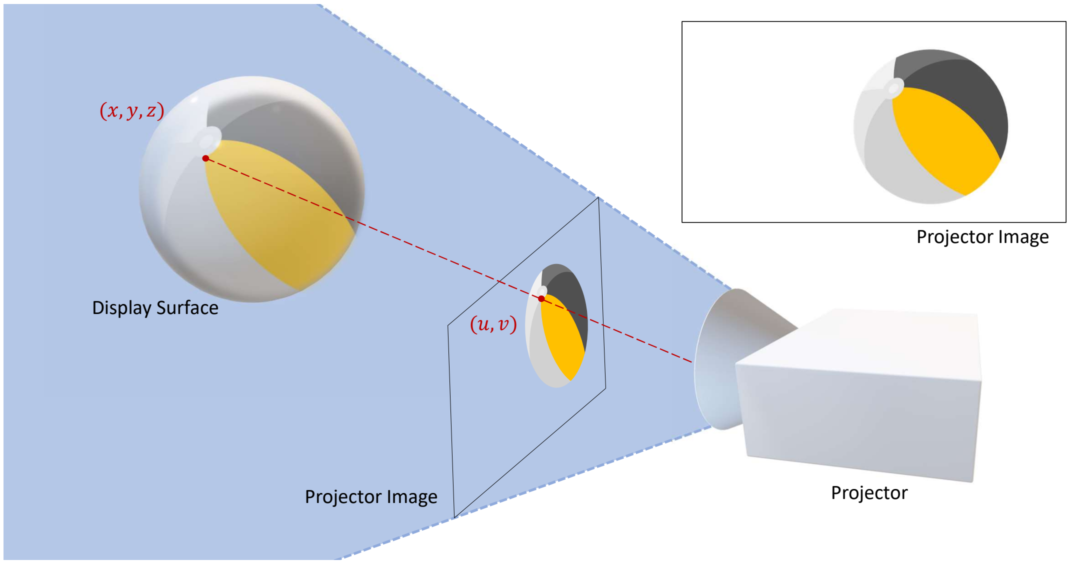


Figure 1.1: A visual display is generated from the projected light of a 2D rendering onto a 3D surface. A pixel coordinate (u, v) in the 2D projector image is mapped to a 3D world coordinate on the display surface (x, y, z).

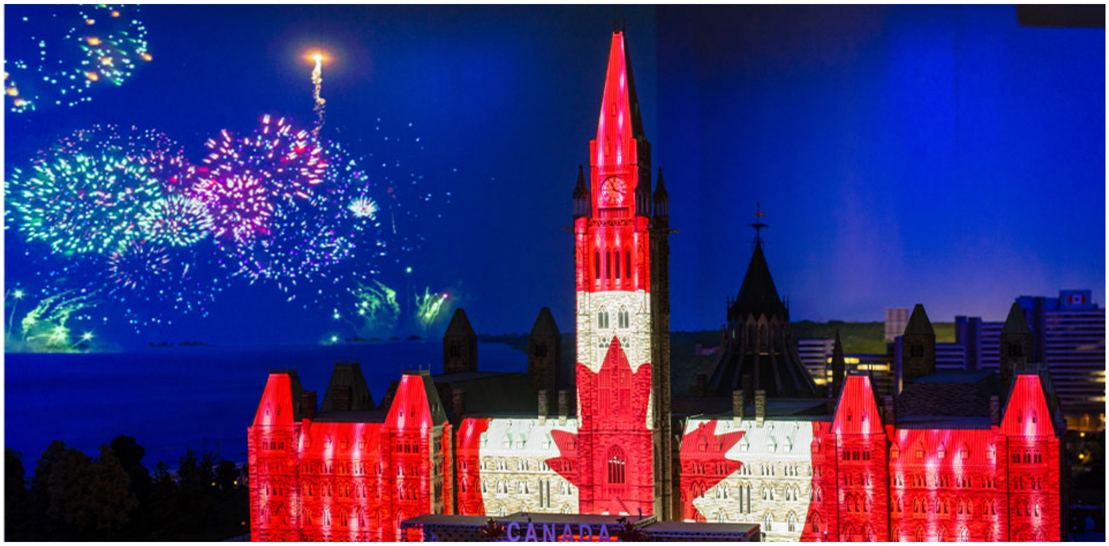


Figure 1.2: A miniaturized Parliament Building is projection mapped at Toronto's "Little Canada" attraction. Photo obtained from Christie Digital.

scene constraints such as calibration targets. Targets might be challenging or disruptive to put in place, limiting the calibration methods available to a user and could prohibit certain display environments. This thesis aims to improve the accessibility and flexibility of projection mapping by presenting a method of projector calibration and scene estimation that eliminates reliance on both environment limiting parameters and on high-capital requirements.

1.1 Motivation and Overview

Geometric calibration of a projection mapping system will rely on the same principles of calibration and estimation whether it is a simple projector purchased for home use or a set of many performative projectors working together for an entertainment industry application. The camera and the scene associated with each specific projection mapping system provide the bulk of the information from which we build a geometric understanding of the system and, consequently, the projector. This thesis develops a method of "self-calibration" for projection mapping applications, where self-calibration indicates little to no input or manipulation required from a user to generate a geometric calibration. This self-calibration will assume that a projection mapping system is also a projector-camera system, in that there is scene measurement with some observing camera in a setup similar to that seen in Figure 1.3.

The prior scene and camera knowledge required by a self-calibration method can establish limits on the range of environments for which the solving method can be used. The assumptions regarding prior knowledge of the display environment are one of the fundamental starting points for calibration. The range of possible scene definitions begins with the most constrained: a known planar calibration target [24, 45, 47] where explicitly known planarity simplifies the calibration. The next level of prior knowledge is non-planar calibration targets, where explicit 3D locations are known but planarity simplifications cannot be used. There are unknown planar scenes where points are not explicitly known but the planarity simplifications can be used. Finally, with the least provided knowledge are unknown scenes where no information can be assumed. Unknown scenes also provide no assurances regarding other scene challenges such as occlusions, or surface warping of structured light patterns. An unknown scene must also be estimated in the projector-camera system calibration.

The camera is primarily a measurement tool, and to generate a scene understanding the first step in this process must be to obtain the set of camera parameters through either prior knowledge or estimation. These parameters include intrinsic values that describe the

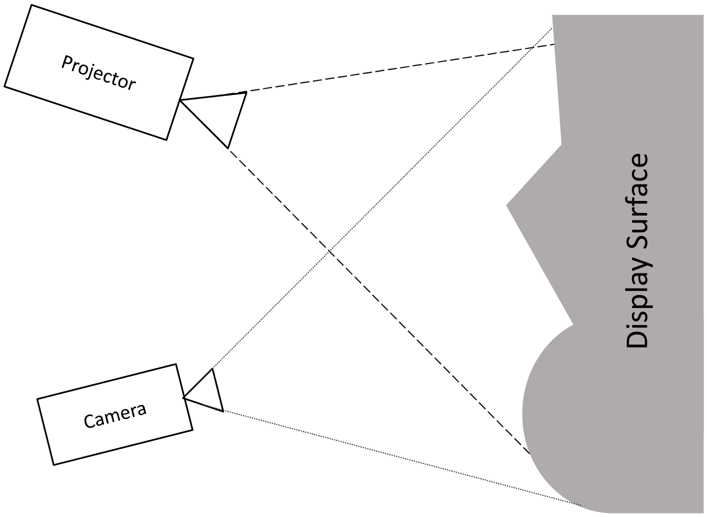


Figure 1.3: Diagram of a simple Projector-Camera System.

camera itself such as focal length and principle point, and extrinsic values that describe the relative position of the camera within an environment. Single- and many-camera strategies have been effective, [27, 38], and the prior knowledge of the camera calibration serves as another limitation-describing characteristic of projector calibration strategies. The most prior knowledge is provided by a pre-calibrated camera of known pose, where no information needs to be estimated. A pre-calibrated camera of unknown pose provides some prior knowledge but requries that the camera position must be estimated. Finally, an unknown camera provides the least definition, where all camera parameters must be estimated in addition to the unknown projector parameters.

When the projector calibration is required, the phase of a projection display project has typically progressed to a place where the intended display artwork has already been drafted for the final display. This process of creating these art pieces can necessitate the existence of some prior virtual scene understanding. This method explores the challenge of unknown scene by introducing a virtual sparse non-planar calibration target. This allows some prior scene information to be introduced without incurring any scene limitations typically associated with calibration targets.

Previous estimation strategies that rely on single-camera and projector pairs [41, 27] treat the projector image plane as a camera image plane of measurement signals. However, as this plane is projecting and not capturing image information, completing this step with two camera image planes provides more measured information from which to construct scene understanding. A moving camera formulation allows the benefit of multiple camera views while minimizing the set of information that must be estimated by requiring only a single set of camera parameters. Where a 3D estimation of the scene has been constructed from the camera scene understanding, the projector calibration follows as an estimation of parameters that must meet the geometric requirements established by the cameras. This allows scene estimation more rooted in reality as constructed from two sets of measurements rather than a set of measurements and a set of projections, and allows better constraints on the final projector estimation.

The foundation of a moving camera allows the future incorporation of additional camera perspectives, potentially overcoming common size and occlusion challenges faced in scene observation by stationary cameras with augmented scene understanding. A moving camera calibration also provides a framework for a handheld camera calibration that, assuming common use of cell phones, has the potential to drastically reduce the cost of obtaining the required cameras.

1.2 Problem Statement and Objectives

A self-calibration projector-camera process that does not rely on known or fixed cameras nor calibration targets is highly desirable to improve the accessibility of projector-camera systems, as well as to increase both the ease of use and the range of possible environments for existing systems. In order to accomplish a moving-camera self-calibration a set of objectives must be met:

- Estimation of the unknown geometric surface.
- Estimation of the camera intrinsic and extrinsic parameters.
- Estimation of the projector intrinsic and extrinsic parameters.

The developed method that accomplishes the above must meet the following constraints:

- Eliminate reliance on high-capital requirements such as fixed-pose and precalibrated cameras
- Eliminate reliance on display environment-limiting parameters such as in-scene calibration targets.

The problem will be further developed in Chapter 3.

1.3 Thesis Organization

This thesis consists of seven chapters, with the remaining chapters structured as follows.

Chapter 2 describes the relevant background material for projector and camera calibration, including geometry of image projection, including the progression from 1-, 2- to n- view calibration methods, prominent methods in obtaining point correspondences for calibration, and some material on optimization and evaluation of projector calibration.

Chapter 3 details the problem upon which this thesis focuses.

Chapter 4 introduces variations on two prominent initialization methods for camera and projector calibration, the Fundamental Matrix and the Trifocal Tensor.

Chapter 5 explores bundle adjustment as a recovery from weak camera calibration and introduces a synthetic calibration target based on known information in display system applications.

Chapter 6 describes the dataset, experiments and results of the discussed initialization and bundle adjustment techniques.

Chapter 7 discusses the final results and future work, and concludes the thesis.

Chapter 2

Background

It is an understatement to say that there is substantial background on the formulation of camera calibration and scene reconstruction. This chapter focuses on introducing the knowledge required to understand the proposed content in this work. First, Section 2.1 introduces the geometric relationships that define and constrain the system. Section 2.2 discusses the application of this geometry to camera projections in single and multi-view environments. Section 2.3 discusses existing strategies for projector-camera system calibration and describes the methods of relating observed 2D points across a set of image planes and their real 3D coordinates. Finally, Section 2.3.2 describes the general uses for optimization strategies within camera calibration and scene reconstruction, and details the formulation of Bundle Adjustment for later use.

2.1 Geometry of Image Projection

Geometry, one of the oldest branches of mathematics, concerns itself with the properties of space, and relations in distance, shape, size, and relative position [1]. It is important to first establish the 'geometry' of our problem as the sets of governing rules that constrain these relations. There are multiple geometries that might be encountered in image projection and scene reconstruction. This section considers the 2D planar Euclidean geometry $\mathbb{R}^2$, the 3D Euclidean geometry $\mathbb{R}^3$, the 2D projective geometry $\mathbb{P}^2$, and the 3D projective geometry $\mathbb{P}^3$.

A captured or projected image plane follows the familiar rules of planar geometry that describe the Euclidean representation of a 2D plane [23, 17, 14]. We define a 2D coordinate in an image plane by the pair of coordinates in $\mathbb{R}^2$:

$$x = \begin{bmatrix} u \\ v \end{bmatrix} \tag{2.1}$$

A line in this 2D plane can be represented by an equation such as $Au + Bv + C = 0$, where changes of A, B and C will give rise to different lines. Not all of these lines will be unique, as $(kA)u + (kB)v + (kC) = 0$ and $Au + Bv + C = 0$ will be the same for any non-zero constant k. These lines and their representing vectors $[A, B, C]^T$ and $[kA, kB, kC]^\top$ are considered equivalent, or *homogeneous vectors* [23].

A point x lies on the line $[A, B, C]^\top$ if it satisfies $Au+Bv+C = 0$. By representing the point as a 3×1 vector $[u, v, 1]^\top$, a point on the line may be written algebraically as a product of vectors as $[u, v, 1][A, B, C]^\top = 0$. This 2D coordinate $[u, v]^\top$ in $\mathbb{R}^2$ is represented as a 3×1 vector by adding this final coordinate of 1. It follows that $[ku, kv, k]^\top$ for varying values of k is a set of homogeneous vectors that represent the same 2D point x, [6, 23, 14, 17]. Given a coordinate triple $[ku, kv, k]$, we can get the original coordinates back by dividing by k. An intuitive geometric interpretation of homogeneous coordinates allows the embedding of the 2D plane $\mathbb{R}^2$ into the 3D Euclidean space $\mathbb{R}^3$ [17]. The camera image plane $x = [u, v]^\top$ is considered the 2D plane in $\mathbb{R}^3$ where the u-axis and v-axis are parallel to the x-axis and y-axis respectively, with x, y coordinate bounds defined by the camera resolution (U, V), and $z = 1$. This allows us to define the coordinate on the image plane:

$$x = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{2.2}$$

separately from the 3D world coordinate:

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{2.3}$$

within the same frame of reference.

Projective Geometry is used to describe how perspective and position within a coordinate system will alter distance and space [14, 23]. Projective geometry preserves straightness and we may define a projective transformation of a plane as any mapping of points on the plane that preserves straight lines [23]. The set of homogeneous vectors in $\{\mathbb{R}^3 - [0, 0, 0]^\top\}$ forms the projective space $\mathbb{P}^2$ [23]. In other words, $\mathbb{P}^2$ represents a 2D coordinate $[u, v]^\top$ as the set of all equivalent $[ku, kv, k]^\top$ homogeneous vectors [23]. Just as 2D coordinates in $\mathbb{R}^2$ correspond to 3D coordinates in $\mathbb{P}^2$, 3D coordinates in $\mathbb{R}^3$ are represented in $\mathbb{P}^3$ projective geometry by the set of equivalence vectors $\{\mathbb{R}^4 - [0, 0, 0, 0]^\top\}$ [23, 14, 17]. Similarly, $\mathbb{P}^3$ represents a 3D coordinate $[x, y, z]^\top$ as the set of all equivalent $[kx, ky, kz, k]^\top$ homogeneous vectors [23].

This projective representation is important because it is projective correlation that provides the foundation for mapping a 3D world coordinate to a 2D image plane [23, 14, 17]. Given the projective space $\mathbb{P}^2$ and the projective space $\mathbb{P}^3$, a mapping $\mathbb{H}$ is an invertible projective correlation that permits the mapping $\mathbb{H} : \mathbb{P}^2 \rightarrow \mathbb{P}^3$ [17]. This interpretation allows us to imagine the camera image plane within the real world geometry, and the projective correlation provides the method for the 2D observation of a 3D world coordinate which represents perspective and position within a coordinate system strongly influence the captured image. The specific formulation of this mapping will be motivated in relation to our camera model in Section 2.2.

2.1.1 Reconstruction

In reconstructing a scene from images, we are attempting to bridge a knowledge gap between the projective geometry definition of camera projection and the Euclidean geometry definition of the real world. This involves learning or estimating governing parameters in the Euclidean geometry that are not well predicted by behaviour in the projective geometry. In projective geometry none of angles, distance, or ratios of distance are preserved [23, 14, 17] (see Table 2.1). When equipped with only image coordinates only a projective scene reconstruction is achievable, which does not allow for meaningful interpretation, [23, 14].

In order to improve the reconstruction, some additional information about the world or system of cameras must be known beyond just the image coordinates, [23, 14]. Information about the projection model can allow for an affine reconstruction, and self-calibration methods which estimate camera and world information can make a metric reconstruction possible, where a metric reconstruction is equal to a Euclidean reconstruction up to a scale factor [14]. Additional information may be provided through a scene definition or calibration target where the projection surface is known, or may be provided by a known camera, which contains the set of true camera parameters. The benefit of a calibration target or known camera is that it means world and camera knowledge are not subject to estimation errors and can resolve ambiguities.

Table 2.1: A comparison of geometries and their particular transformations and invariants; where each geometry is a subset of the next. *[x] indicates that this property exists, where blank indicates that it does not exist for each particular geometry.* [14, 23]

	Geometries:	**Euclidean**	**Metric**	**Affine**	**Projective**
	Degrees of Freedom	3	4	6	8
Transformations	Rotation	x	x	x	x
	Translation	x	x	x	x
	Isotropic Scaling		x	x	x
	Scaling Along Axis, Shear			x	x
	Perspective Projections				x
Invariants	Distance	x			
	Angles	x	x		
	Ratio of Distances	x	x		
	Parallelism	x	x	x	
	Centre of Mass	x	x	x	
	Incidence, Cross Ratio	x	x	x	x

2.2 Camera Projection

We use the pinhole camera model [17, 14, 23] to describe both camera and projector systems. This model follows from our 2D image plane in the 3D world representation developed in Section 2.1. Figure 2.1 provides visualization to illustrate the camera model from the 2D image coordinate to its 3D counterpart in a Euclidean coordinate system. The homogeneous coordinate representation of projective geometry allows us to describe this model as a linear mapping between the 2D homogenous coordinate x to 3D homogenous coordinate X [14, 23]. Projection matrix P describes the transformation from 2D to 3D and can be rewritten in a way that explicitly makes reference to the individual parameters used to define spatial calibration:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = KR \begin{bmatrix} I_3 & C \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2.4}$$

where K is the 3×3 intrinsic matrix, R is the 3×3 rotation matrix, C is the 1×3 position of the camera centre I_3 is the 3×3 identity matrix. These parameters can be sorted into either intrinsic or extrinsic values describing the camera or projector behaviour. Within

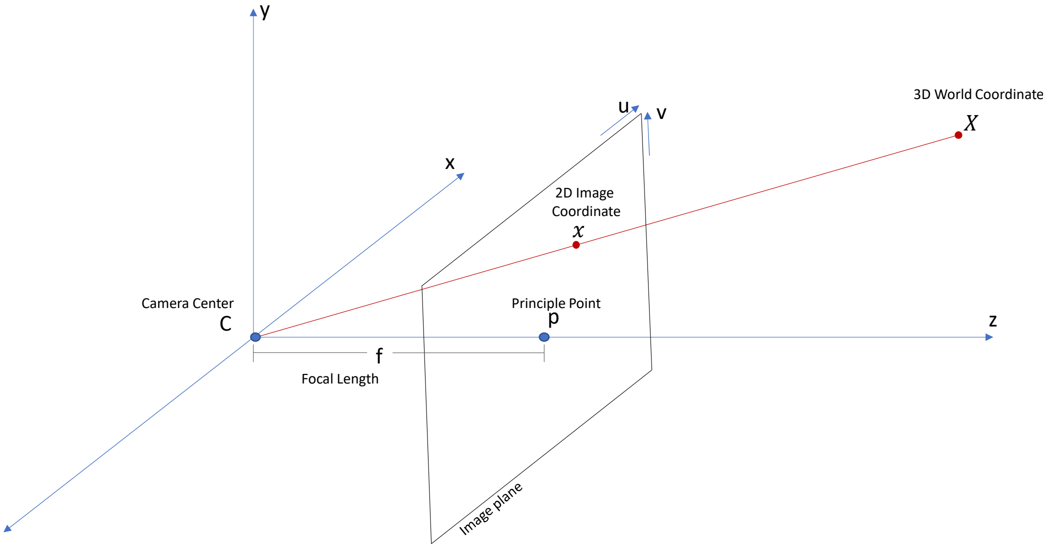


Figure 2.1: The pinhole camera geometry model relates a 2D coordinate x in an image plane to a 3D observed world coordinate X through projective transformations.

the projection matrix P, the intrinsic properties K are defined as:

$$K = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.5}$$

where intrinsic matrix K contains the principle point (p_u, p_v) which is the center of the perspective projection of the image, focal length (f_x, f_y), and skew (s) which describes the angle between x and y. This K matrix and its parameters describe the transformation from the 2D camera pixel coordinate reference frame to the Euclidean world reference frame. The extrinsic parameters are used to define the position of the camera centre (C) and orientation of the camera (R) within the Euclidean world.

$$R \begin{bmatrix} I_3 & -C \end{bmatrix} = \begin{bmatrix} R \mid t \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{bmatrix} \tag{2.6}$$

Instead of explicitly using the camera centre, translation t can be used where $t = -RC$. A direct solution for determining the projection matrix from image and scene points exists as the direct linear transformation. This requires that there be at least six pixel correspondences matching observed points in an image to their 3D world coordinates and that the surface be non-planar. For a set (x_i, X_i) of corresponding 2D x and 3D X coordinates where there are $i = 1 : N$ points as related by Equation 2.4, we can find a direct relationship:

$$\begin{bmatrix} x_i \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_i \\ 1 \end{bmatrix} \tag{2.7}$$

For the 2D image point $[u_i, v_i, 1]^\top$ and 3D coordinate $[x_i, y_i, z_i, 1]^\top$:

$$u_i = \frac{p_{11}x_i + p_{12}y_i + p_{13}z_i + p_{14}}{p_{31}x_i + p_{32}y_i + p_{33}z_i + p_{34}} \quad v_i = \frac{p_{21}x_i + p_{22}y_i + p_{23}z_i + p_{24}}{p_{31}x_i + p_{32}y_i + p_{33}z_i + p_{34}} \tag{2.8}$$

For each 2D to 3D correspondence there are two equations:

$$\begin{array}{lll} -X_i^\top[p_{11}, p_{12}, p_{13}, p_{14}] & & +u_i X_i^\top[p_{31}, p_{32}, p_{33}, p_{34}] = 0 \\ & -X_i^\top[p_{21}, p_{22}, p_{23}, p_{24}] & +v_i X_i^\top[p_{31}, p_{32}, p_{33}, p_{34}] = 0 \end{array} \tag{2.9}$$

This allows us to generate a system of equations over $i = 1 : N$, $N \geq 6$ points that will allow us to solve for the parameters p_{nm} of the projection matrix P from only the relations between one 2D image view and the 3D scene coordinates.

2.2.1 Two Views

This section discusses geometric relations between two views in a scene. We focus on systems of distinct optical centres because our formulation assumes that every image plane is captured from a new position, and there will always be at least two distinct optical centres when calibrating a projector-camera system: one camera and one projector. We neglect systems of single optical centres (pure rotation) and their methods because they cannot be used to describe the projector-camera system.

Epipolar geometry provides the basis from which we can relate two viewpoints [14, 23]. Figure 2.2 describes this epipolar geometry. The line connecting the optical camera centres C_a, C_b intersects with image planes at points e_a and e_b, respectively. The lines through e_a and e_b that connect with their respective observed points x_j^1, x_j^2, where $j = a, b$ are the epipolar lines. Each point x in the first image is viewed as a corresponding point somewhere on the epipolar line in the second image. The 3×3 fundamental matrix F describes the correspondence between a point and its epipolar line. A pair of views for which F is known

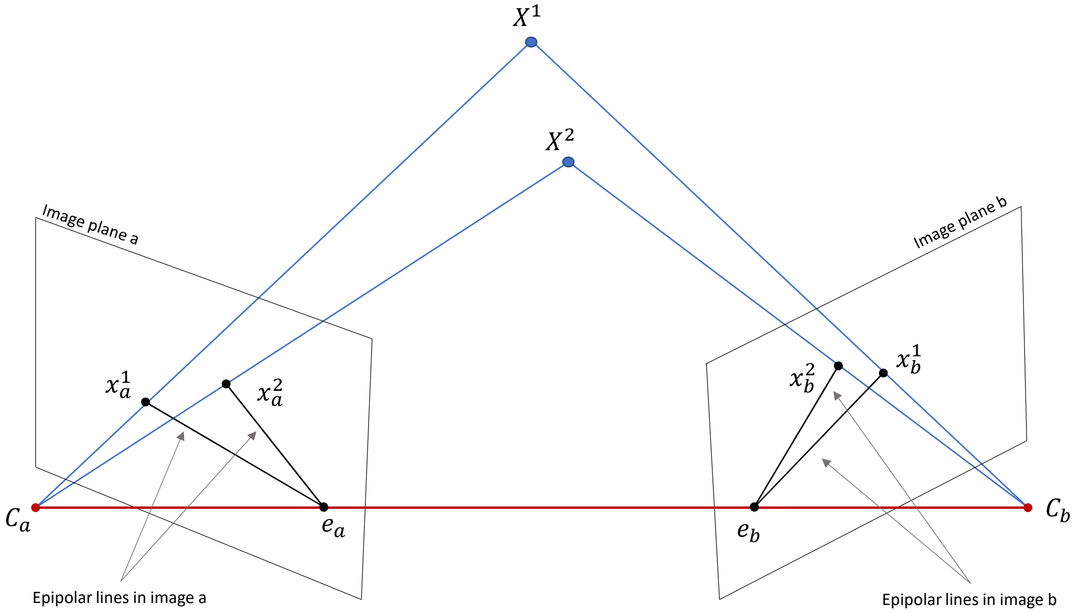


Figure 2.2: Epipolar geometry relates 2D pixel coordinates x_a in image a to the observed 3D coordinate X and the corresponding 2D pixel coordinate x_b in a second image b.

is said to be weakly calibrated. Epipolar line l_b corresponding to the ith observed point in image b can be defined with the fundamental matrix [23]:

$$l_b \approxeq F x_a^i \tag{2.10}$$

Because l_b contains point x_b^i by definition:

$$x_b^{i\top} F x_a^i = 0 \tag{2.11}$$

We define $[a]_\times$ as the skew-symmetric matrix representation of the 1 × 3 vector a:

$$[a]_\times := \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \tag{2.12}$$

The fundamental matrix can be expressed as a function of the perspective projection matrix of the first camera and any inverse perspective projection matrix of the second camera. Defining $\langle ab \rangle_\times$ as the cross product operator of two 3D vectors a and b, we can further express the fundamental matrix as a function of the two perspective projection matrices [23, 17, 14].

$$F = P_b^{-\top} \langle C_a C_b \rangle_\times P_a^{-1} \tag{2.13}$$

Relating the fundamental matrix and definition of the projection matrix $P = K[R|t]$, we can see the following decomposition:

$$F = K_a^{-\top} R_a^{-\top} \langle C_a C_b \rangle_\times R_a^{-1} K_a^{-1} \tag{2.14}$$

We define the Essential Matrix as containing the pose relating camera view a to camera view b and has the form:

$$E = [t_{ab}]_\times R_{ab} \tag{2.15}$$

for the rotation R_{ab} and the translation t_{ab} between the two view planes. The fundamental matrix can be transformed into the essential matrix by removing the intrinsic parameters represented in the intrinisc matrix K:

$$E = K_b^\top F K_a \tag{2.16}$$

which allows us to consider the pose estimated from point correspondences where a camera's intrinics are known.

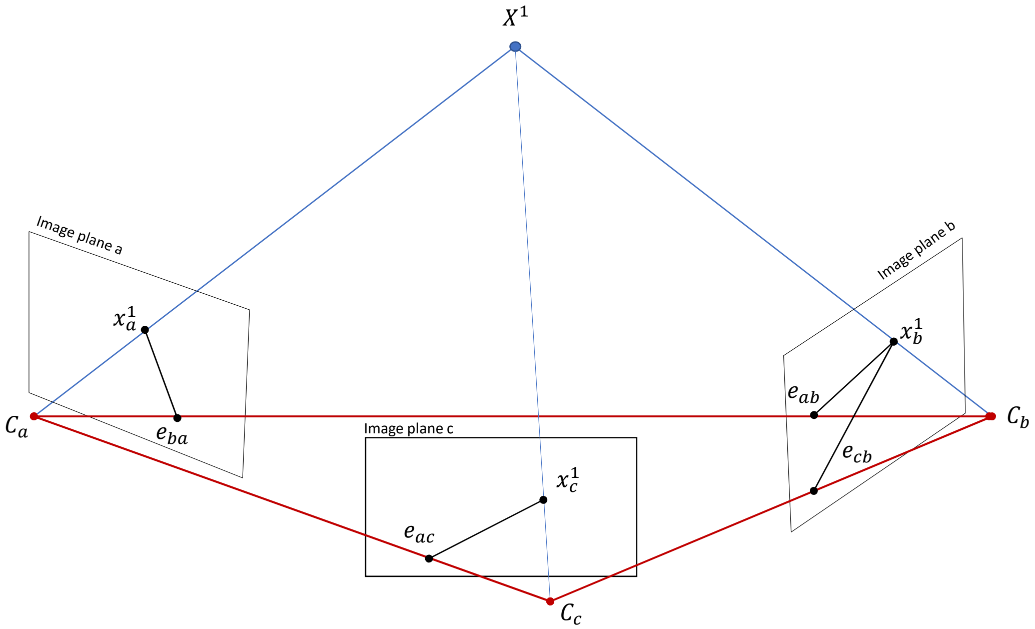


Figure 2.3: Epipolar geometry relates 2D pixel coordinates x_a in image a to the observed 3D coordinate X and the corresponding 2D pixel coordinates x_b, x_c in a second image b and a third image c.

When using calibrated cameras the intrinsic parameters K are provided. When the cameras are uncalibrated this matrix K must be estimated. Of particular note is Bougnoux's estimation [8] employing the fundamental matrix to estimate the focal length, which is often used in projector camera systems [38, 17, 27, 21, 31]:

$$\hat{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad f = \sqrt{-\frac{p_b^\top [e_b]_\times \hat{I} F p_a p_a^\top F^\top p_b}{p_b^\top [e_b]_\times \hat{I} F \hat{I} F^\top p_b}} \tag{2.17}$$

for two image planes a and b related by F, each with a principle point $p = [p_u, p_v]^\top$.

2.2.2 Three-View Geometry

The trifocal tensor is the natural progression into three views of the epipolar geometry relationships defined in two views in the fundamental matrix formulation [23, 14, 17, 25, 2]. The trifocal tensor is a 3×3×3 tensor associated with three camera views which contains 27 parameters. $\mathbf{T} = [T_1, T_2, T_3]$, defined for three projective cameras $P_1 = [I_3|0]$, $P_2 = [A|a_4]$, $P_3 = [B|b_4]$ with each slice T_n the 3×3 matrix:

$$T_n = a_n b_4^\top - a_4 b_n^\top \tag{2.18}$$

where a_i, b_i are the columns of A and B. Similar to the fundamental matrix, the trifocal tensor can be derived from the relation between a 2D pixel correspondence relation between the three images. The following equation for triplets of the i-th corresponding image point across image planes a, b, c defines this relationship between pixel correspondences and the trifocal tensor, for $i, j, k = 1 : 3$ as indices implied by the Einstein convention produce a set of 9 equations [20, 23]:

$$x_a^i (x_b^j x_c^k T_{i33} - x_c^k T_{ij3} - x_b^j T_{i3k} + T_{ijk}) = 0 \tag{2.19}$$

Given an estimate for the trifocal tensor $[T_1, T_2, T_3]$, we can retrieve the epipoles e_{ab} and e_{ac}. The epipole e_{ac} can be computed as the common intersection of the lines represented by the right null-vectors of T_1, T_2, and T_3. Similarly, the epipole e_{ab} can be computed as the common intersection of the lines represented by the left null-vectors of T_1, T_2, and T_3 [23, 14, 17, 25, 2]. With these epipoles we can retrieve an estimate for the essential matrices:

$$E_{ba} \approx [e_{ab}]_\times \left[T_1^\top e_{ac} \quad T_2^\top e_{ac} \quad T_3^\top e_{ac} \right], \quad E_{ca} \approx [e_{ac}]_\times \left[T_1^\top e_{ab} \quad T_2^\top e_{ab} \quad T_3^\top e_{ab} \right] \tag{2.20}$$

The application of Equation 2.16 allows the decomposition of these essential matrices into pose parameters describing the rotation and translation between reference view a and both views b and c.

2.3 Self-calibration

Self-calibration, sometimes referred to as autocalibration, of projector camera systems refers to the ability of a system to be able to estimate the calibration parameters (pose and intrinsic values) of the imaging devices without any form of user interaction [37]. Some self-calibration methods will also estimate the display scene [16, 27]. Many state-of-the-art methods rely on calibration tools. It is common to see checkerboard planar calibration target methods [24, 47, 12, 30], which are based on the initial work of Zhang [45]. Many state-of-the-art methods rely on either calibration tools such as these checkerboards, additional scene information such as calibration objects [44] or require flat planar display surfaces [12, 24, 26, 28, 30].

Calibration relying on only geometric information provided by the Fundamental Matrix [7, 15, 41, 27, 37, 32, 38] or the Trifocal Tensor is also common [16, 32, 2, 25]. These geometric calibration methods will provide their own refinement or optimization strategy meant to improve the initial estimate afforded by the weak calibration of the geometric methods [37]. These optimization strategies may be through a unique cost function [37, 7, 16, 41, 27], or rely on existing refinement techniques such as bundle adjustment [24, 43, 32].

2.3.1 Structured Light

In order to estimate the fundamental matrix or trifocal tensor so that geometric calibration can be done, it is necessary to obtain information relating pixel coordinates in the image planes. A prominent component of geometric projector-camera self-calibration is the extraction of information relating pixel coordinates in the image planes. Systems that focus on scene understanding and calibration from images will also rely on some ability to relate the real 3D world to the information extracted from the captured images. Geometric relationships themselves are only useful when provided with some data or information within which to build relations. A significant advantage of a projector-camera system over general camera-only systems is the ability to fully control a layer of information projected onto the 3D scene which will allow for very strong relationships to be generated. Systems employing *structured light* use a projector to illuminate the scene with a particular pattern of images which encode information that will be used to establish corresponding points within the image plane of the projector and image planes of the observing camera(s) [19, 27, 24, 33].

There are many different models of structured light as demonstrated in Figure 2.4, accommodating various scenarios, with two prominent categories of single-shot [39, 40, 33] or multi-shot/continuous patterns [19, 27]. For our purposes, the differentiating characteristic between these categories is that a single-shot method provides one pattern that must be captured by a single camera, while a multi-shot method assumes that cameras are fixed and stable throughout a series of projected patterns, or a continuous pattern that varies over time, where the entire sequence of the pattern must be observed by the camera in order to be decoded. A single-shot method better facilitates a system where one camera moves through multiple poses and observes the unchanging pattern as it moves. Some projector-camera methods relying upon physical calibration targets will project structured light onto the calibration object to generate additional relations between the structured light patterns and known information about the world [24, 47].

These structured light techniques aim to produce sets of pixel correspondences that identify coordinate locations in each image plane that correspond to the same observed 3D feature in the display environment. All of the 2D points x_j^i correspond to the same single 3D point X^i.

$$x_j^i = \begin{bmatrix} u_j^i \\ v_j^i \end{bmatrix} \tag{2.21}$$

where j describes to which image plane (projector view, camera view) the set of points belongs, and $[u, v]^\top$ indicates the 2D pixel location of the ith 3D feature that is captured in the image plane. From these sets of point correspondences, we can use the identified geometric relationships to begin to construct an understanding of our scene. Establishing pixel correspondences using structured light methods allows us the ability to relate the projected pattern in the projector view to the captured camera images and treat the projector as a camera [33]. This enables the geometric camera calibration strategies discussed to be applied to the projector calibration problem.

2.3.2 Refining A Geometric Estimate

Epipolar geometry formulations such as the fundamental matrix and the trifocal tensor are said to provide a weak calibration for camera systems [23, 17, 14]. Many estimation methods require a minimum number of pixel correspondences to get an exact solution, but where long sequences of thousands of pixel correspondences are available from robust structured light solutions, these estimation methods seek a minimal solution instead of an exact solution. Such an optimization might be a uniquely proposed strategy [37, 7, 16, 41, 27], or might rely on tested strategies such as RANSAC estimation [23, 14, 17, 27, 5] or Bundle Adjustment [36, 10, 24, 11, 18, 34, 9, 46].

We focus on the bundle adjustment refinement strategy. Through consideration of the 2D reprojection error, bundle adjustment aims to refine a visual reconstruction to produce jointly optimal 3D scene and camera calibration estimates:

$$\min_{\Theta, X} \frac{1}{M} \frac{1}{N} \sum_{j=1}^{M} \sum_{i=1}^{N} ||x_j^i - \pi(\Theta_j, X^i)||^2 \tag{2.22}$$

over $i = 1 : N$ image points, where π describes the $3D$ to $2D$ projection (Equation 2.4) of 3D scene coordinate X by estimated parameters $\Theta_j = \{K_j, R_j, t_j\}$, for the jth image plane in the set of M image planes. The following equation defines π as a mapping from 3D coordinate X to it's 2D representation in the image plane characterized by camera parameters K, R, t:

$$\pi(\Theta, X) : X \to x, \quad x = K[R|t]X \tag{2.23}$$

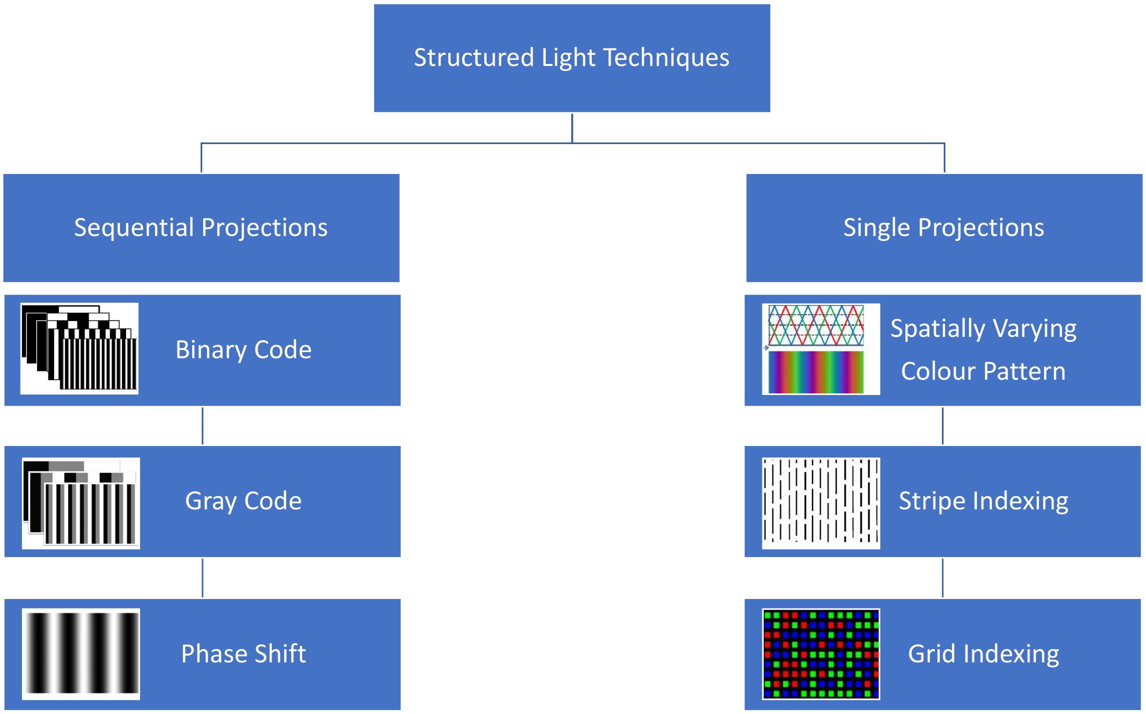


Figure 2.4: Illustration of different structured light methods categorized by sequential images versus single images projected and captured. Example images copied from [19].

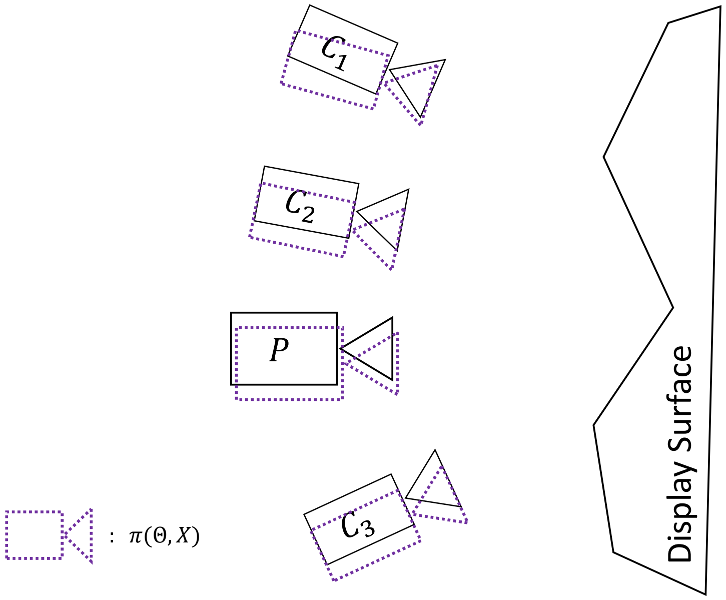


Figure 2.5: Illustration of bundle adjustment refinement. Image planes are estimated based on known or approximate 3D display surface coordinates and camera or projector calibration parameters. Estimated image planes (purple) are compared with the measured image planes (black) and the parameters used to generate the estimated image planes are adjusted to minimize the difference.

Figure 2.5 describes this strategy. The estimation or known 3D coordinates X representing the display surface are projected into an approximate 2D image plane by the estimated camera parameters Θ. These approximate image planes (purple in Figure 2.5) are compared with the measured image planes (black) and the objective of the minimization is to reduce the residual r between them:

$$r_j^i = x_j^i - \pi(\Theta_j, X^i) \tag{2.24}$$

Bundle adjustment has a flexibility which allows it to adapt to various problem formulations, such as fixed or varying intrinsic parameters, and fixed or varying 3D coordinates [36, 23, 17]. Bundle adjustment must contend with being sensitive to the provided initialization [23] and can be an extremely large minimization problem when large sets of pixel correspondences are considered [23, 36]. The use of bundle adjustment will be further formulated in Section 3.2 and Chapter 5.

Chapter 3

Problem Formulation

As motivated in Chapter 1, this thesis aims to develop a process of geometric calibration for projection mapping systems that improves accessibility and affordability of projection mapping technology by reducing the prerequisites of calibration. This thesis aims to focus on eliminating reliance on known cameras or real calibration targets.

As established in Section 2.3, geometric calibration relies on established pixel correspondences between image view planes. This process will assume that some method [19, 27, 24, 33] of obtaining sets of pixel correspondences between image planes and the display environment already has been used to generate the information that will be used as input.

In this thesis an initial estimate of the scene is generated by applying geometric constraints on these pixel correspondence sets. As mentioned in Section 2.2, these methods produce a projective geometric calibration for the scene, which has been found insufficient for complex geometric displays. A bundle adjustment optimization (Section 2.3.2) uses this geometric estimate as an initial point and minimizes reprojection error to recover the scene information.

Discussed in Section 2.3, while the projector is modelled as an inverse camera, it tends to have parameter behaviour that does not follow the same patterns as a camera. This problem formulation establishes a scene understanding from the camera information before integrating the projector, which allows the projector to be estimated separate from any assumptions about patterns within the parameters.

This chapter formulates the methods that take us from an initial set of pixel correspondences across three camera images and a projector image plane to the full set of scene and parameter estimates, as described by Figure 3.1.

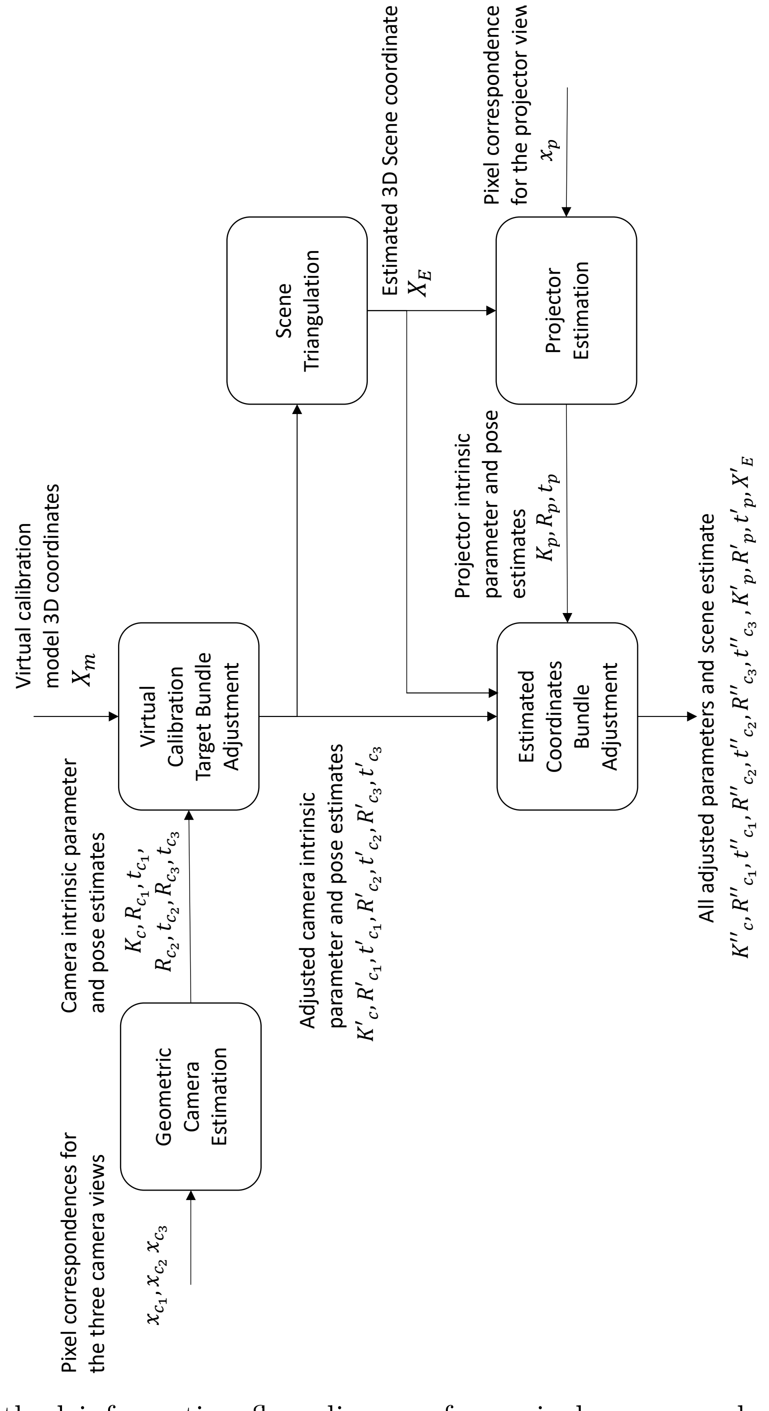

Pixel correspondences for the three camera views
$x_{c_1}, x_{c_2}\, x_{c_3}$
Geometric Camera Estimation
Camera intrinsic parameter and pose estimates
$K_c, R_{c_1}, t_{c_1}, R_{c_2}, t_{c_2}, R_{c_3}, t_{c_3}$
Virtual calibration model 3D coordinates
X_m
Virtual Calibration Target Bundle Adjustment
Adjusted camera intrinsic parameter and pose estimates
$K'_c, R'_{c_1}, t'_{c_1}, R'_{c_2}, t'_{c_2}, R'_{c_3}, t'_{c_3}$
Scene Triangulation
Estimated 3D Scene coordinate
X_E
Pixel correspondence for the projector view
x_p
Projector Estimation
Projector intrinsic parameter and pose estimates
K_p, R_p, t_p
Estimated Coordinates Bundle Adjustment
All adjusted parameters and scene estimate
$K''_c, R''_{c_1}, t''_{c_1}, R''_{c_2}, t''_{c_2}, R''_{c_3}, t''_{c_3}, K'_p, R'_p, t'_p, X'_E$

Figure 3.1: Method information flow diagram from pixel correspondences to final scene and parameter estimates

Table 3.1: Competing characteristics comparison of geometry- and calibration target-based approaches.

Approach	**Competing Characteristics**		
Calibration target based	Higher accuracy	Address depth, scale ambiguity	Scene contraints
Geometry based	Lower accuracy	Does not address depth, scale	Flexibility (no scene constraints)

3.1 Geometric Calibration Estimate

There are two prominent pathways of calibration approaches: calibration target based approaches, and geometric-based approaches. The main properties of these different approaches are compared in Table 3.1. While calibration target based approaches provide more accuracy and scene definition they require the limitations on display environment that we are trying to overcome. Subsequently, we focus on geometry-based approaches for the scene flexibility they afford.

3.1.1 Moving Camera

Having decided upon a geometric based approach, and equipped with the point correspondences as measured information, we are at this point able to construct sets of relations between the projector image plane and the camera image planes that will ultimately result in the final calibration. As described in Section 2.1, the following constraints define the considered geometric approaches for the Fundamental Matrix F relating a 2D coordinate x in two images a and b:

$$x_b{}^{\top} F x_a = 0 \tag{2.11 revisited}$$

and for the Trifocal Tensor $\mathbf{T}_{3\times3\times3}$, relating 2D coordinate x across three images a, b and c in a set of 9 equations indexed by $i, j, k = 1:3$:

$$x_a^i(x_b^j x_c^k T_{i33} - x_c^k T_{ij3} - x_b^j T_{i3k} + T_{ijk}) = 0 \tag{2.19 revisited}$$

In order to proceed, the image planes that correspond to views a, b, and c must now be determined. Previous estimation strategies that rely on single-camera and projector pairs [41, 27, 24] treat the projector image plane the same as a captured camera image plane of measurement signals, while other methods use a set of cameras to establish scene understanding without the projector image planes [38, 39], integrating the projector views into a defined scene.

We consider the information that must be found. To use a camera image plane and a projector image plane for geometric calibration, we have the minimum set of parameters for which to solve. However, these methods rely on decomposition of the estimated Fundamental Matrix (see Section 2.2.1) or Trifocal Tensor (see Section 2.2.2) into the intrinsic and extrinsic components of the different view planes. The decomposition is facilitated by the simplifying assumptions made about the camera intrinsic parameter structure. Common camera assumptions are principle point within the centre of the image, and equal focal length in the x and y directions. These assumptions are not necessarily true for the projector, but are integral to the decomposition with uncalibrated cameras. This makes geometric calibration with the projector image plane challenging.

Each camera or projector is characterized by a set of 11 parameters. From Equation 2.4 the set of camera parameters that define a camera model are:

$$KR\left[I_3 \quad C\right] \tag{2.4 revisited}$$

This intrinsic matrix K contains 5 parameters, focal length f_x, f_y, principle point p_x, p_y, and skew s. The extrinsic parameters include the rotation matrix R has 3 degrees of freedom, which is fully defined by 3 parameters, and the camera centre position C, described by 3 coordinates. Each subsequent imaging device provides an additional 11 parameters that must be found. To avoid this, we formulate with a moving camera which allows us to assume a minimum set of parameters by holding the 5 intrinsic parameters constant across image views. The additional parameters that must be found are reduced to 6 parameters, the 3 for rotation and the 3 for camera centre position, for each additional camera view. A single moving camera allows the benefit of multiple camera views in geometric calibration while reducing the overall set of parameters that must be found. Consequently, we set the image views a, b, c for our Fundamental Matrix and Trifocal Tensor strategies as our moving camera image planes c_1, c_2, c_3, to develop a scene understanding with our camera views for later projector integration:

$$a = c_1, \quad b = c_2, \quad c = c_3 \tag{3.1}$$

3.1.2 Camera Estimation

We aim to eliminate reliance on precalibrated and fixed pose camera limitations as motivated in Chapter 1, which means that we assume all of the camera intrinsic and pose parameters are initally unknown and must be estimated. We make a set of assumptions for the case of a camera-camera calibration problem with a moving camera of equal intrinsic parameters. These assumptions are used throughout the geometric estimation strategies used (Fundamental Matrix and Trifocal Tensor estimation), but as they are themselves approximate assumptions about regular behaviour and not absolutes, they are also adjusted in the later refinement strategies. The following four assumptions are made:

1. $K_{c_1} = K_{c_2} = K_{c_3}$ same intrinsic matrix for all camera views.
2. $p_{c_1} = p_{c_2} = p_{c_3}$ same principle point for all camera views.
3. $p_{c_1} = \frac{1}{2}[UV]$ principle point for the camera is the centre of the camera image plane for a camera resolution $[U, V]$.
4. Assume zero skew $(s = 0)$ and unit aspect ratio $(f_x = f_y)$.

Points 1 and 2 follow from using the same camera to capture all the images, where the only difference is position within the scene. Points 3 and 4 are made based on common camera behaviour of square pixels and centre-image principle point axis. From these assumptions, estimation strategies such as Bougnoux's [23, 14] provide an initial estimate for the focal length f. While points 1,2, and 4 will remain true throughout the calibration process, point 3 and the Bougnoux's focal length method are both used to provide an initial estimate and both initial focal length and principle point will be allowed to adjust in the later refinement. Having obtained reasonable estimations for all the camera intrinsic parameters, the initial single intrinsic matrix for the camera views is constructed from Equation 2.5:

$$K_{c_1} = K_{c_2} = K_{c_2} \rightarrow K_c = \begin{bmatrix} f & 0 & \frac{1}{2}U \\ 0 & f & \frac{1}{2}V \\ 0 & 0 & 1 \end{bmatrix} \tag{3.2}$$

defining intrinsic matrix K, focal length $f = f_x = f_y$, principle point p_x, p_y found from the camera resolution $[U, V]$, and skew $s = 0$. The effort of calibrating the camera is significantly reduced, as these assumptions about typical camera parameter behaviour allow for the intrinsic parameters to be estimated without relying on any prior information other than the image plane (estimating the principle point). This allows us to move to estimating pose quite rapidly, where we aim to later recover any loss of accuracy from the generalization of camera behaviour.

We aim to estimate the position of our cameras within the display environment. We use structured light strategies to provide pixel correspondences between the display surface, the projector image plane, and all the camera image planes. Figure 3.2 describes the system environment.For each coordinate in the set of $i = 1 : N$ observed image points we employ Equation 2.4:

$$x_j^i = P_j X_R^i, \qquad j = p, c_1, c_2, c_3 \tag{3.3}$$

The set of real world 3D coordinates X_R are related by a projection matrix P_j to a set of pixel coordinates in the image view x_j. These relations are made in by the projection matrix P_p to a set of pixel coordinates in the projector view x_p. Corresponding relations are also made by a different projection matrix $P_{c_1}, P_{c_2}, P_{c_3}$ to sets of pixel coordinates in the camera views $x_{c_1}, x_{c_2}, x_{c_3}$ as the camera described with intrinsic parameters K_c moves through the different image perspectives denoted by c_1, c_2, c_3.

For pose estimation, and employing the definition of camera pose from Equation 2.6 we position the world coordinate system such that the camera centre for view one aligns with the origin. Then, each subsequent view is measured with respect to the camera centre at the origin. R_{ij}, t_{ij} describes a rotation and translation from the i-th camera view to the j-th camera view:

$$P_{c_1} = \begin{bmatrix} K_c & 0 \end{bmatrix}, \quad P_{c_2} = K_c \begin{bmatrix} R_{12} | t_{12} \end{bmatrix}, \quad P_{c_2} = K_c \begin{bmatrix} R_{13} | t_{13} \end{bmatrix} \tag{3.4}$$

Provided this assumption and having obtained an estimate of the Fundamental Matrix relating the camera views, we use Equation 2.16 to extract the Essential Matrix E_{ji} which describes the pose relating the j-th camera view to the i-th camera view for which the calibration K_c is known:

$$E \approx K_c^\top F K_c \tag{3.5}$$

where this Fundamental Matrix estimation and decomposition to the Essential matrix would be completed once for each additional image view after the first image pair. Alternatively, provided an estimate of the Trifocal Tensor, the essential matrix between views might be computed from the epipoles retrieved from the Trifocal Tensor:

$$E_{21} \approx [e_{12}]_\times \begin{bmatrix} T1 & T2 & T3 \end{bmatrix} e_{13}, \quad E_{31} \approx [e_{13}]_\times \begin{bmatrix} T1 & T2 & T3 \end{bmatrix} e_{12} \tag{2.20 revisited}$$

where $[a]_\times$ is the skew-symmetric matrix representation of vector a from Equation 2.12. From Equation 2.15, we know the essential matrix has the form:

$$E_{j1} \approx [t_{1j}]_\times R_{1j}, \quad j = 2, 3 \tag{3.6}$$

This provides us the information needed to complete the initial estimate of calibration for the camera views.

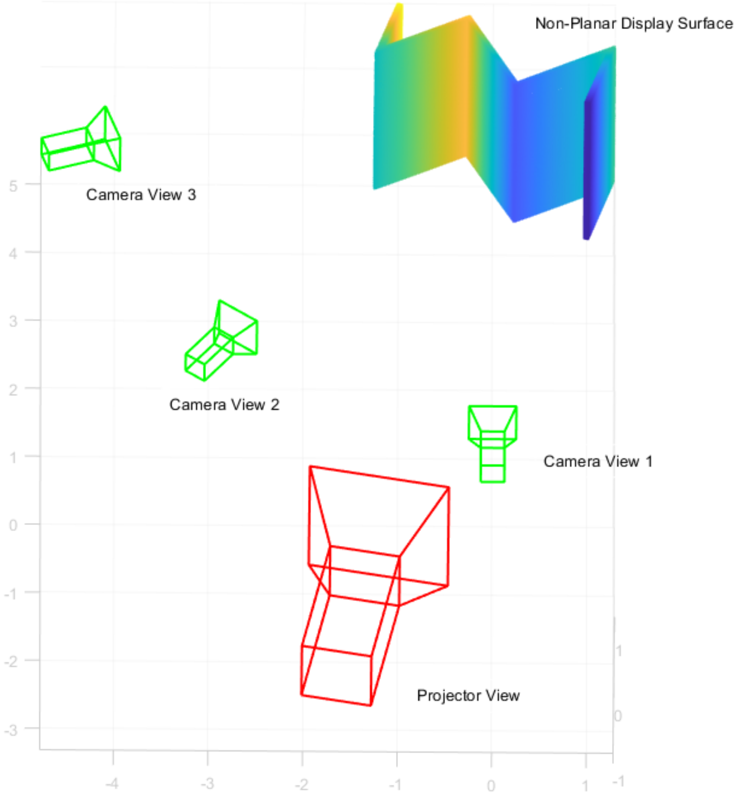


Figure 3.2: Scene diagram illustrates the environment for geometric calibration of the projector-camera system. A single camera (green) moves through a set of camera poses, observing the display surface. The projector (red) is projecting upon the display surface, and remains stationary.

3.1.3 Scene Triangulation and Projector Incorporation

Without a calibration target or real world 3D coordinates, we do not know the set of real world 3D coordinates X_R. Part of the geometric calibration of the display environment is estimating the 3D surface. The calibration estimate of the camera can be estimated from only 2D pixel coordinate relationships between image planes. This calibration estimate can then be used to construct a 3D scene estimate X_E. X_E is an estimate of the set of real world 3D coordinates X_R:

$$X_E \approx X_R + \epsilon \tag{3.7}$$

for some unknown estimation error ϵ.

Provided a set of adjusted camera intrinsic parameters and pose estimates a triangulation method τ [42, 22, 17] uses the set of captured pixel correspondences $\{x_{c_1}, x_{c_2}, x_{c_3}\}$ to construct the 3D scene estimate X_E:

$$X_E = \tau(\Theta_{c_1}, \Theta_{c_2}, \Theta_{c_3}, x_{c_1}, x_{c_2}, x_{c_3}), \tag{3.8}$$

where $\Theta_j = \{K_j, R_j, t_j\}$, for the jth image plane.

These estimated coordinates are then used to estimate the calibration of the projector through the direct linear transformation estimation described by Equation 2.8. As our projector has provided the location encoding pattern for obtaining our camera point correspondences, the locations of corresponding points between x_p and X_E are known. This strategy allows projector calibration to be constrained by incorporating the projector calibration into a known scene. This will be further detailed in Chapter 4.

3.2 Bundle Adjustment

A Fundamental Matrix or Trifocal Tensor estimate for a camera scene is said to weakly calibrate the system [23, 14]. As discussed in Section 2.3.2, calibration methods often employ some optimization strategy to produce a more accurate system estimate. Bundle Adjustment [14, 17, 23, 36] is a flexible and reliable method for adjusting the found calibration of the cameras as well as the points within the scene estimate. We consider two formulations of bundle adjustment, one where the 3D coordinates X are the estimated coordinates X_E, and one where the 3D coordinates X are the virtual model coordinates described in Section 3.2.1. Through consideration of the 2D reprojection error, bundle adjustment aims to refine a visual reconstruction to produce jointly optimal 3D scene and camera calibration estimates:

$$\min_{\Theta,X} \frac{1}{M}\frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{N}||x_j^i - \pi(\Theta_j, X^i)||^2 \tag{2.22 revisited}$$

over $i = 1 : N$ image points, where

$$\pi : X \to x \tag{2.23 revisited}$$

describes the $3D$ to $2D$ projection (Equation 2.4) of 3D scene coordinate X by estimated parameters $\Theta_j = \{K_j, R_j, t_j\}$, for the jth image plane in the set of M image planes.

This refinement strategy flexibility allows a growing number of M image planes to be used, providing an opportunity to incorporate further views that may aid in overcoming unknown scene challenges such as occlusions. This formulation is also very adaptable to the application, with methods that assume precalibrated cameras holding fixed the intrinsics K over adjustment, or holding fixed the set of 3D scene coordinates X if they are provided as known ground truth. An example of this would be the coordinates of known landmarks in remote sensing images; these coordinates would not need to be adjusted as they are known and are instead their fixed location can be used to improve the camera position estimate.

3.2.1 Virtual Calibration Target

In the case of the projection mapping display system, an image or sequence of images is projected onto a surface in a way that fits the surface. These images are part of an intended artistic display. In order to generate the art to be displayed it is necessary that some sort of virtual understanding of the projection surface is known. At some point in advance of the final calibration and operation of the projection mapping system, some sort of 3D understanding of the surface was needed by the graphic artist to produce the image to be projected, especially in the case where the artist aims to fit the design to certain 3D features of the projection surface. This allows us to assume that some virtual model knowledge exists.

We must also assume that the surface knowledge is approximate and may not be relied upon completely. This is reflected in the optimization by relying only on a sparse set of keypoints instead of any full virtual model that might be available. Many display environments will experience some wear or weathering, or the virtual knowledge might be an approximation of something that is difficult to measure accurately such as a complex statue. We assume that a sparse set of 3D model coordinates $X_m^i = (x^i, y^i, z^i)$ are known, and that we can find the closest point correspondences in the set $x_j^i = (u^i, v^i)$ for each camera view j, as seen in Figure 3.3.

This short fixed keypoint bundle adjustment is completed with the virtual calibration target X_m as our set of 3D coordinates. This is completed after the estimation of the camera intrinsic parameters K_c and the pose parameters $R_{12}, t_{12}, R_{13}, t_{13}$, and used to provide an adjusted set of these camera parameters in advance of using them to generate the estimated coordinates X_E.

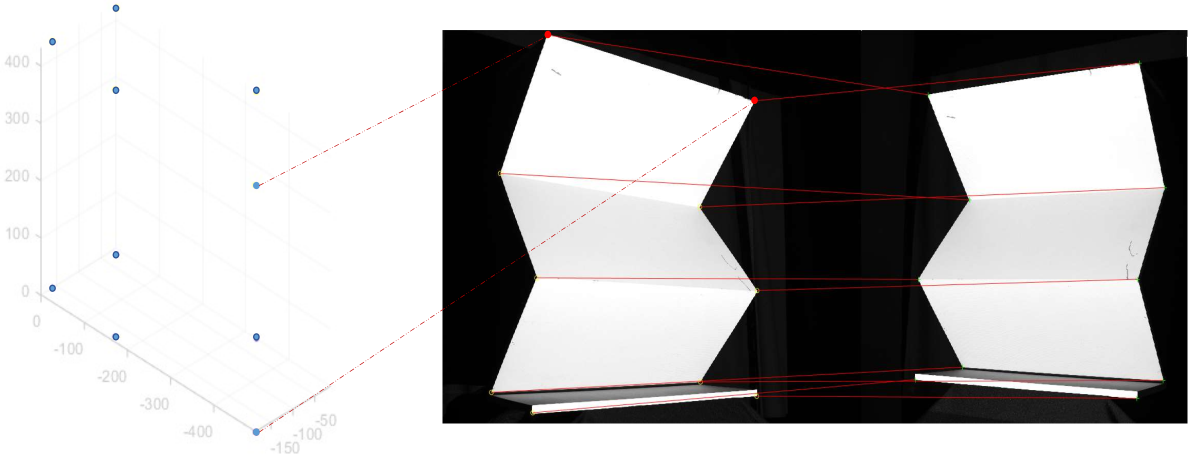


Figure 3.3: Example of a virtual calibration target. Key points of the synthetic model (left) are matched to their locations in the point correspondence sets of two moving camera views. Lines added to visualize the correspondence relationships.

3.2.2 Triangulated Coordinates

As discussed in Section 3.2.1, a sparse set of model coordinates X_m are assumed to be sufficiently sparse that the set is not enough to assure a complete surface understanding, and even if accurate, such a set would not be able to reflect expected changes in a scene's surface, such as a building facade weathering over time. A sparse set of model coordinates would also not be able to make use of the large amount of scene information that would be captured by a structured light system. Following the short bundle adjustment the adjusted camera parameters are used to generate a set of estimated 3D coordinates X_E. This estimation of X_E is completed by triangulation τ as discussed in Section 3.1.3. This set of coordinates will be used to estimate the projector parameters, and a full bundle adjustment will adjust the camera and projector parameters, and the estimated X_E 3D coordinates to generate the final scene understanding.

3.2.3 Parameter Adjustment Flexibility

The flexibility of bundle adjustment allows different formulations of bundle adjustment to be used throughout a calibration method. A formulation may hold particular parameters fixed throughout adjustment, versus allowing them to be adjusted throughout the method. In the short bundle adjustment where the virtual key point set X_m is being used, the 3D coordinates X_m are held fixed as they can be assumed to be a known ground truth. For the set of estimated coordinates X_E, these 3D coordinates X_E are be adjusted with the other parameters as they too are estimates influenced by error in the parameters.

Similarly, any parameters within the estimated camera and projector calibration might be held fixed or otherwise linked. As the camera images are all from a single moving camera, the intrinsic parameters K_c are adjusted jointly across the image planes to reflect that the intrinsic parameters are consistent for the camera across the images. Chapter 5 will detail the different formulations of bundle adjustment employed in this thesis.

Chapter 4

Geometric Camera Calibration

The geometric camera calibration aims to estimate the camera intrinsic and pose parameters which are used to estimate the projector intrinsic and pose parameters and the 3D scene coordinates. As described in Section 3.2, this camera parameter estimate aims to provide a sufficient initialization for a bundle adjustment recovery. As this thesis aims to eliminate reliance on physical calibration targets and prior information describing the cameras, the geometric camera calibration parameter estimate must rely only on information provided by the image planes (captured by the camera or cast by the projector). The parameter estimation strategy was formulated in Section 3.1, and a detailed solution will be proposed here. This chapter details the use of pixel correspondences x_j to produce an estimate of the camera intrinsic parameters K_c and the set of camera poses $R_j, t_j, j = c_1, c_2, c_3$ through two geometric estimation processes, the fundamental matrix estimation described in Section 4.1, and the trifocal tensor estimation discussed in Section 4.2.

The captured images are known information, including the location encoding information from structured light methods (Section 2.3.1 that provides the sets of pixel correspondences x, as well as the resolution of the images. We can say that the resolution of the camera or the projector, which is the size of the captured or projected image, can be easily found if unknown. Methods exist [19, 27, 24, 33, 13] to produce these pixel correspondences with structured light strategies, which are patterns projected on the scene surface which encode location information in camera and projector image planes [28], as described in Section 2.3. These methods allow us to construct our set of pixel correspondences:

$$x_j^i = \begin{bmatrix} u_j^i \\ v_j^i \end{bmatrix} \tag{2.21 revisited}$$

for a set of $i = 1 : N$ 2D pixel coordinates where these points are matched across the set of $j = 1 : M$ observed camera image planes and the known pattern in the $j = p$ projector image plane. From these sets of pixel correspondences, we can begin to construct an understanding of our scene. We use only the set of $j = \{c_1, c_2, c_3\}$ observed camera image planes in order to construct our geometric camera calibration parameter estimate. The projector image plane pixel correspondences will be later used to estimate the projector parameters from the scene understanding constructed from the camera estimate.

Both the fundamental matrix and the trifocal tensor formulations detailed in Section 3.1 can be estimated with just pixel correspondences. Then, both the fundamental matrix and the trifocal tensor rely upon an initial knowledge of the camera intrinsic parameters K_c to extract the pose information. This set of intrinsic parameters includes the camera focal length $f(f_x, f_y)$ and principle point $p = (p_x, p_y)$, both of which must be estimated. We generate an estimate of the intrinsic parameters assuming that the principle point is at the centre of the camera image, $p = \frac{1}{2}[U, V]^\top$ for camera with resolution $[U, V]^\top$. A focal length can be estimated using the principle points and the Fundamental matrix F relating two images a, b through Bougnoux's method [8]:

$$\hat{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad f = \sqrt{-\frac{p_b^\top [e_b]_\times \hat{I} F p_a p_a^\top F^\top p_b}{p_b^\top [e_b]_\times \hat{I} F \hat{I} F^\top p_b}} \tag{2.17 revisited}$$

With this focal length estimate, the intrinsic matrix K_c can be assembled:

$$K_c = \begin{bmatrix} f_x & 0 & \frac{1}{2}U \\ 0 & f_y & \frac{1}{2}V \\ 0 & 0 & 1 \end{bmatrix} \tag{3.2 revisited}$$

This estimate for the camera intrinsic parameters is then used to extract pose information from the fundamental matrix or the trifocal tensor.

4.1 Fundamental Matrix

As described in Section 2.1, the fundamental matrix can be estimated from pixel correspondences. The formulation for a three-view fundamental matrix estimation considers two pairs of image planes; the pair $\{c_1, c_2\}$ and the pair $\{c_1, c_3\}$. The world coordinate system is positioned such that the camera centre for c_1 aligns with the origin with zero rotation. This allows both the other camera view poses $R_{c_2}, t_{c_2}, R_{c_3}, t_{c_3}$, to be measured with respect to the first image. Revisiting Equation 2.11, two fundamental matrices are estimated applying the following constraint to matched point correspondences in images $1, 2, 3$:

$$x_2{}^\top F_{21} x_1 = 0, \quad x_3{}^\top F_{31} x_1 = 0 \tag{4.1}$$

The estimated K_c intrinsic parameters are now used with these estimated fundamental matrices F to extract the essential matrices E by Equation 2.15.

$$E_{21} \approx K_c^\top F_{21} K_c, \quad E_{31} \approx K_c^\top F_{31} K_c \tag{4.2}$$

Then, the pose information contained in the essential matrix is decomposed into rotation and translation information. As the world coordinate system is positioned such that the camera centre for c_1 aligns with the origin, the rotation and translation captured in the essential matrix is entirely in the other considered image plane. Equation 2.15 is used to generate these pose estimates.

$$E_{21} \approx [t_{c_2}]_\times R_{c_2}, \quad E_{31} \approx [t_{c_3}]_\times R_{c_3} \tag{4.3}$$

where R_{c_1} is the 3×3 identity matrix and t_{c_1} is a 3×1 zero vector. At this stage the fundamental matrix has been used to generate an estimate for the camera parameters K_c as well as the three camera view poses $R_j, t_j, j = \{c_1, c_2, c_3\}$. Next, a short bundle adjustment refinement is performed as described in Section 5.1, followed by an estimation of the projector parameters and display surface as described in Section 3.1.3.

4.2 Trifocal Tensor

Like the fundamental matrix, and as detailed in Section 2.1, the trifocal tensor can also be estimated from pixel correspondences. The trifocal tensor approach is employed to explore whether the additional image plane provides such additional geometric constraints as to improve the camera parameter estimate. Where the fundamental matrix can only be used to relate two images and must be found twice for a set of three images, the trifocal tensor can establish relations between the set of three images directly. This approach considers the set of image planes $\{c_1, c_2, c_3\}$. The world coordinate system is still positioned such that the camera centre for c_1 aligns with the origin with zero rotation. This allows both the other camera view poses $R_{c_2}, t_{c_2}, R_{c_3}, t_{c_3}$, to still be measured with respect to the first image. Revisiting Equation 2.19, the $3 \times 3 \times 3$ trifocal tensor $[T_1, T_2, T_2]$ is estimated.

$$x^i_{c_1}(x^j_{c_2} x^k_{c_3} T_{i33} - x^k_{c_3} T_{ij3} - x^j_{c_2} T_{i3k} + T_{ijk}) = 0 \tag{2.19 revisited}$$

where the Trifocal Tensor $\mathbf{T}_{3\times3\times3}$, relates the 2D pixel correspondences x_j across three images a, b and c in a set of 9 equations indexed by $i, j, k = 1 : 3$.

To estimate the focal length with Equation 2.17 we need a fundamental matrix. The trifocal tensor can provide an estimate for the fundamental matrix that we need:

$$F_{21} \approx [e_{12}]_\times \begin{bmatrix} T_1 & T_2 & T_3 \end{bmatrix} e_{13}, \quad F_{31} \approx [e_{13}]_\times \begin{bmatrix} T_1 & T_2 & T_3 \end{bmatrix} e_{12} \qquad \text{(2.20 revisited)}$$

using the epipoles e_{12} and e_{13}. These epipoles are computed from the left and right null-vectors respectively of $[T_1, T_2, T_2]$. With this estimate of F_{21}, F_{31}, the intrinsic parameters K_c can be assembled with a focal length and principle point estimate. The next steps in this approach are the same as the fundamental matrix. Equations 4.2 use this K_c estimate to extract the essential matrices E_{21}, E_{31}, then because the world coordinate system is positioned such that the camera centre for c_1 aligns with the origin, Equations 4.3 can be used to provide an estimate for the pose information $R_j, t_j, j = \{c_1, c_2, c_3\}$. Provided this set of generated camera intrinsic and pose parameters, the short bundle adjustment refinement is performed as described in Section 5.1, followed by an estimation of the projector parameters and display surface as described in Section 3.1.3.

Chapter 5

Bundle Adjustment

The objective of this application of bundle adjustment [36, 10, 24, 11, 18, 34, 9, 46] is to take the geometric calibration estimate of the image plane intrinsic and pose parameters K, R, t as an initial estimate and produce a refined estimate for the geometric calibration of the camera views, projector view, and the 3D scene estimate. Bundle adjustment methods were formulated in Section 3.2, and a detailed solution will be proposed here.

Two different bundle adjustment refinement strategies are proposed, a short bundle adjustment (Section 5.1) that focuses just on the camera views and the virtual calibration target, and a full bundle adjustment (Section 5.2) that includes all the captured image planes and the set of estimated 3D coordinates. These different formulations are meant to focus on different objectives. Both these formulations consider bundle adjustment as the minimization of 2D reprojection error of $i = 1 : N$ 3D points $\{X^i\}$ observed in $j = 1 : M$ image planes characterized by the image plane intrinsic and pose parameters $\Theta_j = \{K_j, R_j, t_j\}$:

$$\min_{\Theta, X} \frac{1}{M} \frac{1}{N} \sum_{j=1}^{M} \sum_{i=1}^{N} ||x_j^i - \pi(\Theta_j, X^i)||^2 \tag{2.22 revisited}$$

The short bundle adjustment formulation considers a single camera with multiple view planes, and does not adjust the 3D coordinates, assuming they are known and fixed, as discussed in Section 3.2.1. The full bundle adjustment considers all of the camera views and the projector view, and adjusts a set of estimated 3D coordinates, as detailed in Section 3.2.2. These details are further elaborated on below. Briefly we will focus on the similarities across the formulations for bundle adjustment employed in this thesis.

A key aspect of the single moving camera formulation (Section 3.1.1) was minimizing the number of parameters that must be estimated. This moving camera allows us to assume the set of intrinsic parameters K_c to be the same across all camera views. This assumption remains consistent throughout bundle adjustment, where the camera intrinsic parameters are adjusted jointly across the image views. The assumption is represented in the set of camera parameters as the following, across the employed bundle adjustment methods:

$$\Theta = \{\Theta_j = \{K_c, R_{c_j}, t_{c_j}\}\} \quad \text{for} \quad j = c_1, c_2, c_3 \tag{5.1}$$

We do not use this assumption for $j = p$, the projector parameters, as the projector will have its own intrinsic parameters K_p. Both the short and full bundle adjustment methods rely on a method of projecting the 3D coordinates to pixel correspondences based on the estimated camera or projector parameters. This projection is defined consistently across the bundle adjustment strategies:

$$\pi(\Theta_j, X) : X \to \hat{x}, \quad \hat{x}^i_j = K_j \begin{bmatrix} R_j & t_j \end{bmatrix} \begin{bmatrix} X^i \\ 1 \end{bmatrix} \tag{2.23 revisited}$$

This results in a homogeneous 3D coordinate $\hat{x} = [wu, wv, w]^\top$, where the final pixel coordinate is found by the following operation:

$$x = \begin{bmatrix} u/w \\ v/w \end{bmatrix} \tag{5.2}$$

where this pixel coordinate falls in the estimated camera plane corresponding to the set of parameters used to project it from 3D to 2D.

Finally, both short and full bundle adjustment strategies share that they are formulated as a nonlinear least squares optimization method [36]. Nonlinear least squares estimation searches for the minimum of an objective function fitting N observations with a non-linear model of k unknown parameters such that $N \geq k$. Here, we apply non-linear least squares estimation to Equation 2.22. Our observations are our pixel correspondences and our parameters are the sets of estimated camera and projector parameters Θ_j and estimated 3D scene coordinates X. Levenberg-Marquardt method [29, 36] is used to solve this nonlinear least squares problem minimizing 2D reprojection error for both short and full bundle adjustment.

5.1 Short Bundle Adjustment: Camera and Virtual Calibration Target Coordinates

The first bundle adjustment strategy employed in the proposed method encountered after an estimated has been generated for the camera parameters. The virtual calibration model is used as a set of known and fixed 3D coordinates to adjust these estimated camera parameters in advance of scene estimation and projector parameter estimation. This fixed keypoint adjustment focuses on recovering the camera estimate with linked intrinsic parameters across the camera views. This strategy assumes that there is a provided set of 3D virtual scene model coordinates X_m that represent the observed scene up to a scale, discussed in Section 3.2.1. It also assumes that there has been some matching of the nearest pixel correspondences in each image plane to the virtual representation of each observed 3D key point. For the purposes of this thesis this matching has been completed manually across a small set of 10 key points. This keypoint bundle adjustment focuses on adjusting the camera intrinsic and pose estimate generated from the camera geometric calibration estimate, and we can rewrite Equation 2.22 as the following:

$$\min_{\Theta} \frac{1}{3}\frac{1}{10}\sum_{j=1}^{3}\sum_{i=1}^{10}||x_j^i - \pi(\Theta_j, X_m^i)||^2 \tag{5.3}$$

over $i = 1 : 10$ 3D scene model coordinates, and the set $j = 1 : 3$ of camera views is c_1, c_2, c_3. The complete set of parameters considered in this minimization are the following:

$$\theta_{c_1} = \{K_c, R_{c_1}, t_{c_1}\}, \theta_{c_2} = \{K_c, R_{c_2}, t_{c_2}\}, \theta_{c_3} = \{K_c, R_{c_3}, t_{c_3}\} \tag{5.4}$$

where the intrinsic parameters K_c are adjusted jointly across the views.

A key characteristic of this strategy is that it does not adjust the set of 3D coordinates, instead holding them fixed. The objective of this formulation is to use the provided virtual scene knowledge to provide an improved estimate of the camera position and intrinsic parameters as this estimate of the camera is relied upon for the subsequent estimation of the projector parameters and the display surface. Error in the camera estimate will accumulate more inaccuracy in the later estimation. The 3D coordinates in the virtual model are a scaled representation of the real observed 3D display surface. Holding these points fixed allows the camera parameters to be bounded by the real relative distances between 3D coordinates. A set of estimated 3D coordinates would have error themselves if projected from an erroneous camera estimate, and would not provide quite as reliable of an adjustment to the camera parameters.

5.2 Full Bundle Adjustment: All Views and Estimated Scene Coordinates

The second bundle adjustment strategy in the proposed method is a full adjustment of all the estimated camera parameters, projector parameters, and estimated 3D scene coordinates. As detailed in Section 3.1.3, this strategy assumes a large set of of estimated 3D coordinates X_E generated from triangulation of the adjusted camera parameters following the first bundle adjustment. The upper and lower bounds for the total number of N scene coordinates employed by the full bundle adjustment can be established based on the performance accuracy of structured light methods. We have assumed a single moving camera that observes a single frame from three positions, so we can consider the performance of Single-Shot Structured Light (SSSL) methods [33, 13], compared in Table 5.1. We say that $1000 \leq N \leq 8000$ for the approximate range of possible pixel coordinates shared across the four image planes and we can rewrite Equation 2.22 as the following:

$$\min_{\Theta, X_E} \frac{1}{4} \frac{1}{N} \sum_{j=1}^{4} \sum_{i=1}^{N} ||x_j^i - \pi(\Theta_j, X_E^i)||^2, \tag{5.5}$$

over $i = 1 : N$ estimated 3D scene coordinates, where the intrinsic parameters K_c are adjusted jointly across the camera views, K_p is used for the projector view, and the set $j = 1 : 4$ of considered image views are c_1, c_2, c_3, p. The complete set of parameters considered in this minimization are the following:

$$\theta_{c_1}, \theta_{c_2}, \theta_{c_3}, \theta_p = \{K_p, R_p, t_p\}, X_E \tag{5.6}$$

where $\theta_{c_1}, \theta_{c_2}, \theta_{c_3}$ are as defined previously in Equation 5.4.

A key characteristic of this strategy is that it does adjust the set of estimated 3D coordinates, where the previous short bundle adjustment does not as it holds the keypoints fixed. The set of estimated scene coordinates are generated from a camera estimate, and will have some associated approximation error. The objective is to generate a final adjusted scene estimate and set of camera and projector parameters. The benefit of this set of points over the virtual model is that these points will fit the surface of the scene and any features between the points in the sparse virtual model may be observed and represented in the scene estimate. The optimization is then able to search for a minimum that accounts for error in both the estimated scene and the image parameters.

Table 5.1: Comparison between the number of pixel correspondences (N), RMSE between the 3D reconstructed shape with the CAD model of the object and detection run time of two SSSL methods [33, 13]. Table data provided by [33].

Shapes	Methods	N	RMSE (pixels)	Run time (seconds)
Curve	Method A [33]	4496	7.12	10.80×10^3
	Method B [13]	1186	11.71	29.67
ZigZag	Method A [33]	8064	5.77	9.82×10^3
	Method B [13]	2348	6.07	47.12

Chapter 6

Experiments

The developed geometric calibration estimation methods for calibrating a projector-camera system are implemented for both the fundamental matrix (Section 4.1) and trifocal tensor (Section 4.2) formulations. These methods are used to estimate the camera intrinsic parameters and the three camera poses. This camera estimate is adjusted with a short bundle adjustment (Section 5.1). The adjusted parameters are used to triangulate a display surface estimate and generate a projector estimate (Section 3.1.3). Then the full set of estimated information, which includes the camera intrinsic parameters, the three camera poses, the projector intrinsic parameters, the projector pose, and the estimated 3D coordinates of the display surface are all jointly adjusted with a full bundle adjustment (Section 5.2). This chapter discusses the experiments conducted to explore the performance of the method on synthetic data. Synthetic data permits the direct comparison of estimated values to the 'ground truth' of each parameter, which can be difficult with real data. Section 6.1 explores the performance metrics considered and the setup of the synthetic experiments. Section 6.2 presents the results of these experiments, and Section 6.3 discusses these results and their implications.

6.1 Experimental Procedure

This thesis considers several metrics for performance. The “look” of a resulting reconstruction is often used to measure performance, which makes a quantitative assessment difficult [37]. A benefit of synthetically generated display environments is that error in the estimated parameters corresponding to each image plane (K, R, t) can be directly compared to the parameters used to generate synthetic data.

Reprojection error is measured in pixels, and is found as a sum across all pixel correspondences, where Q is the total number of pixel correspondences $Q = M \times N$, the number of pixel correspondences times the number of image planes. Average reprojection error ϵ_p is measured in pixels and can be used easily for synthetic and real data tests because it compares the set of estimated values to the known pixel correspondence sets. Each pixel coordinate in each image plane is considered.

$$\epsilon_P = \sqrt{\frac{\sum_i^Q (x^i - \hat{x}^i)^2}{Q}} \tag{6.1}$$

Parameter error can be measured for the synthetic data as a comparison between the 'true' synthetic parameter and the estimated parameter. We measure the angular error ϵ_R (in degrees) between the true rotation R_a and the estimated rotation R_b:

$$\epsilon_R = \cos^{-1}\left(\frac{\mathrm{tr}(R_a^\top R_b) - 1}{2}\right) \tag{6.2}$$

and we measure the angular error in translation ϵ_t using the cosine formula for dot products between two vectors between the true translation t_a and the estimated translation t_b:

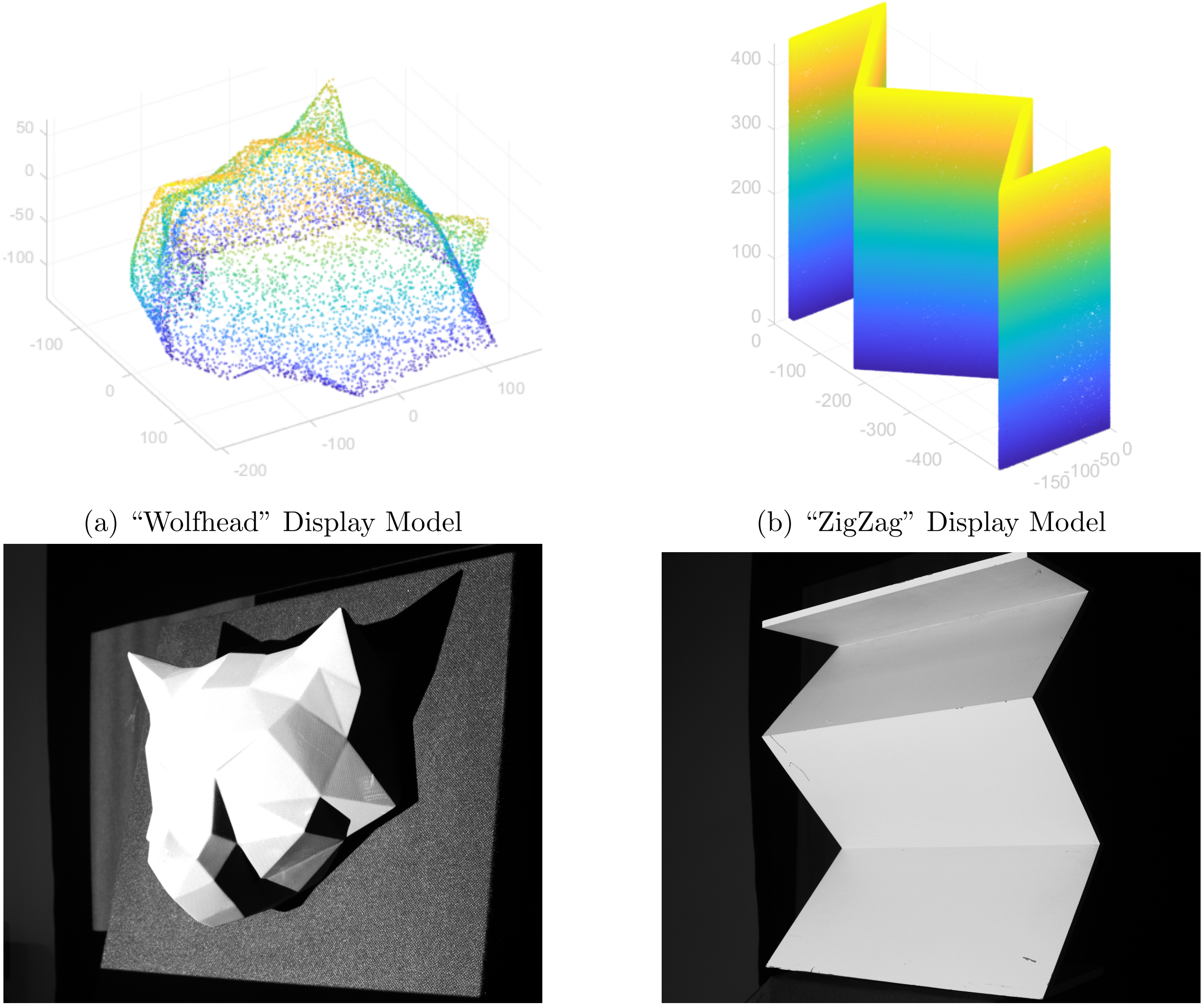
$$\epsilon_t = \cos^{-1}\left(\frac{t_a \cdot t_b}{||t_a|| \cdot ||t_b||}\right) \tag{6.3}$$

The intrinsic parameters are considered seperately as focal length $f = (f_x, f_y)$ and principle point $p = (p_x, p_y)$ for the camera and projector. These parameters are evaluated based on the percent error, measured as the difference between each estimated parameter b and the true parameter value a over the true value:

$$\epsilon_{\%} = \frac{|a_x - b_x| + |a_y - b_y|}{(a_x + a_y)} \times 100 \tag{6.4}$$

The standard display environment for our experiments is as described in Chapter 3 and represented in Figure 3.2. Two display scenes are employed in the experiments, and both are each composed of a set of 3D coordinates representing some non-planar display surface. Display surfaces are provided by Christie Digital, and can be seen in Figure 6.1.

The synthetic data are generated through the projection of a provided set of 3D coordinates representing a known model by a generated set of calibration parameters and pose for each image view. Points are projected onto three camera views with consistent camera intrinsic parameters K_c, and three different poses $R_{c_j}, t_{c_j}, j = c_1, c_2, c_3$. Points are also



(a) “Wolfhead” Display Model

(b) “ZigZag” Display Model

(c) “Wolfhead” Real Display Scene

(d) “ZigZag” Real Display Scene

Figure 6.1: Non-planar display scene model used in synthetic data compared to their real world counterparts. Models and images provided by Christie Digital Systems Inc.

projection onto a projector view, with projector intrinsic parameters K_p and a fourth pose R_p, t_p. The intrinsic parameters of the camera were modeled from available information describing cell phone cameras, and the projector was modeled from the real-world setup of a Christie DWU670-E WUXGA (1920 x 1200) projector. These points are generated with varying additive Gaussian pixel noise.

The intrinsic parameters of the camera are held constant in the synthetic data over the set of three camera views generated. This is aligned with the formulation presented in Section 3.1.1, where a minimum set of parameters is presented for a single moving camera. Often some intrinsic parameters, such as focal length, may be re-adjusted by a device equipped with automatic focusing. It is assumed that this is not the case, however such automatic adjustment can be accommodated in the bundle adjustment formulation, as mentioned later in Section 7.1.

6.2 Results

Two different experiments are conducted on the synthetic data. First, a large sample of $N = 2500$ pixel correspondences are used to explore the performance of the trifocal tensor (TFT) estimation versus the fundamental matrix (FMat) estimation. This first experiment is conducted twice, once for the provided Zigzag model, and once for the Wolfhead model, as seen in Figure 6.1. The objective of this test is to explore the performance of the estimations strategy over a full set of pixel correspondences that might be obtained from a single shot method as described in Table 5.1. Gaussian noise of $\sigma = 1$ px is added to the generated point correspondences. The percent error in the estimated focal length (f_x, f_y) and principle point (p_x, p_y) is compared in both the projector and camera image planes. This error is presented in Table 6.1. The initial geometric estimate for the projector parameters sees quite low error in the projector intrinsic parameters (highest of 8.5%), and the final set of projector parameters are fully recovered (0% error). This is not true for the estimated camera parameters after the final bundle adjustment, which see a high of 26.5% error in these tests. The angular error in the estimated rotations (Equation 6.2) and translation directions (Equation 6.3) is presented in Table 6.2. The estimated display surface and corresponding reprojection error is presented in Figure 6.2.

The second experiment considers a smaller sample of $N = 100$ pixel correspondences under noise in the provided pixel correspondences. The results of the estimation of the camera and projector parameters and the 3D scene coordinates are considered over 40 iterations for each noise level. The change in reprojection error, intrinsic percentage error, and the angular error in the estimated rotations and translation directions is evaluated against varying additive Gaussian noise level $0 \leq \sigma \leq 3$ px added to $N = 100$ pixel correspondences over 40 trials. The zigzag model was used in these tests. The intrinsic percentage error is shown in Figure 6.3. The behaviour of the parameters recovery is very similar to that of the previous two tests over 2500 points, the intrinsic parameters of the projector are well recovered, and the intrinsic parameters of the camera are not. The average angular error across the estimated camera and projector poses is presented in Figure 6.4. The poses of the camera and the projector are well recovered, below half a degree in average angular error across all four views, even as noise increases. The final reprojection error is presented in Figure 6.5.

Table 6.1: Percent error in estimated camera and projector intrinsic parameters for synthetic data with $\sigma = 1$ Gaussian noise added to sets of 2500 the pixel correspondences. Initial corresponds to the result after the geometric estimate (no refinement), and final corresponds to the final result after refinement.

Shape	**Method**	**Percent Error ϵ_p, (%)**			
		Camera Focal Length	**Camera Principle Point**	**Projector Focal Length**	**Projector Principle Point**
ZigZag	TFT Initial	3.5	0	8.5	2
	TFT Final	2.8	24	0	0
	FMat Initial	0.5	0	5	4.7
	FMat Final	2.8	26.5	0	0
Wolfhead	TFT Initial	8.3	0	1.8	0.9
	TFT Final	13.6	18.1	0	0
	FMat Initial	1.3	0	1.4	4.2
	FMat Final	9.4	15.3	0	0

Table 6.2: Estimated camera and projector angular rotation and translation error for synthetic data with $\sigma = 1$ Gaussian noise added to sets of 2500 the pixel correspondences. This angular error is the average of the error calculated from Equations 6.2 and 6.3 across the four sets of pose parameters $R_j, t_j, j = c_1, c_2, c_3, p$. Initial corresponds to the result after the geometric estimate (no refinement), and final corresponds to the final result after refinement.

Shape	**Method**	**Average Angular Error (^{o})**	
		Rotation ϵ_R	**Translation ϵ_t**
ZigZag	TFT Initial	0.76	1.02
	TFT Final	0.11	0.12
	FMat Initial	0.21	0.31
	FMat Final	0.11	0.12
Wolfhead	TFT Initial	3.08	2.08
	TFT Final	0.15	0.22
	FMat Initial	0.23	0.67
	FMat Final	0.15	0.22

6.3 Discussion

The estimation strategy was able to produce an estimate for the intrinsic and pose parameters for a set of three camera views and the projector view, and generate a 3D surface reconstruction, relying only upon the four sets of image pixel correspondences $x_{c_1}, x_{c_2}, x_{c_3}, x_p$ and a sparse virtual scene model X_m. This is accomplished without any prior information about the devices (camera and projector), presenting an advantage above methods that rely on prior parameter information [27]. This method allows the calibration and reconstruction of a non-planar projection surface, presenting also an advantage above methods which are limited to planar surfaces [24, 12, 30].

Two epipolar geometry based approaches are explored, the fundamental matrix relating two views, and the essential matrix relating three views. Across the in the $N = 2500$ experiment the trifocal formulation sees a consistent $2 - 4\%$ higher percent error (and therefore poorer performance) in the intrinsic and extrinsic parameter estimation in both camera and projector estimation. This suggests that the trifocal tensor estimation does not provide any significant estimation advantages despite providing direct relations across three views instead of just two. This is consistent with comparison of fundamental matrix and trifocal tensor estimation performance in other vision applications [25]. The trifocal tensor estimation provides the highest reprojection error observed (16.079 px), compared to the fundamental matrix which provides a reprojection error of 2.195 px under the same

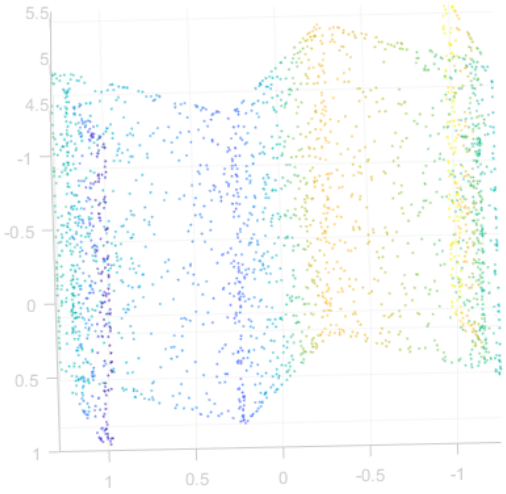
Table 6.3: Comparison of this method's best performance to existing projector-camera calibration strategies. The comparable strategies rely on provided information such as parameter priors and calibration targets. **value not provided by the referenced publication.*

Method	**Camera Focal Length Error**	**Projector Focal Length Error**	**Reprojection Error**
Our Method Best	2.8%	0%	0.5 px
Planar Calibration Target [24]	<2%	<2%	<0.22 px
Parameter Prior [27]	<0.78%	0.99%	*

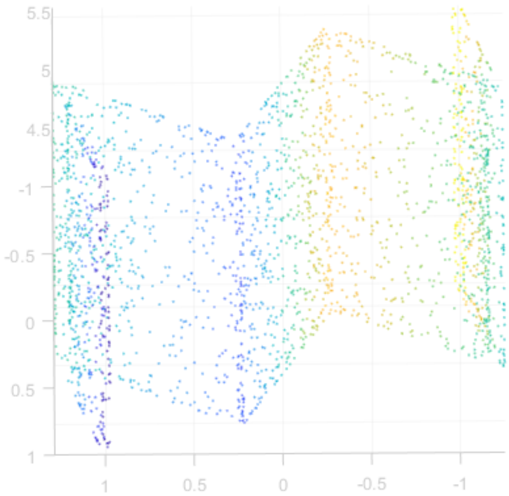
conditions. Conversely, the fundamental matrix estimation sees poorer resilience to noise over the $N = 100$ experiments, but this poor performance is consistent across both fundamental matrix and trifocal tensor, and may be better attributed to a need for a greater number of pixel correspondences for the bundle adjustment refinement.

The refined estimate for the intrinsic values of the camera and projector across both epipolar geometry approaches observed no error in the estimated projector intrinsic parameters, and up to 27% error in the camera intrinsic parameters when considering 2500 pixel correspondences with $\sigma = 1$ px additive Gaussian noise. The projector parameters were equally well recovered when there were only 100 pixel correspondences and varying additive Gaussian noise. For many projection mapping applications, this is sufficient as the cameras are only used for calibrating the projector, and afterwards discarded. For other projector-camera systems where the devices need to operate together continuously, this poorer camera calibration may not be sufficient. Our method is compared to two esiting strategies for projector-camera calibration, with the results illustrated in Table 6.3. The camera parameter estimation in this method is worse than an existing method that relies on parameter prior information [27] an existing method that relies on planar calibration targets [24]. However, the observed very low error in projector parameter estimation (0%) in this method outperforms both. This indicates a similar method performance without the same scene and environment limitations.

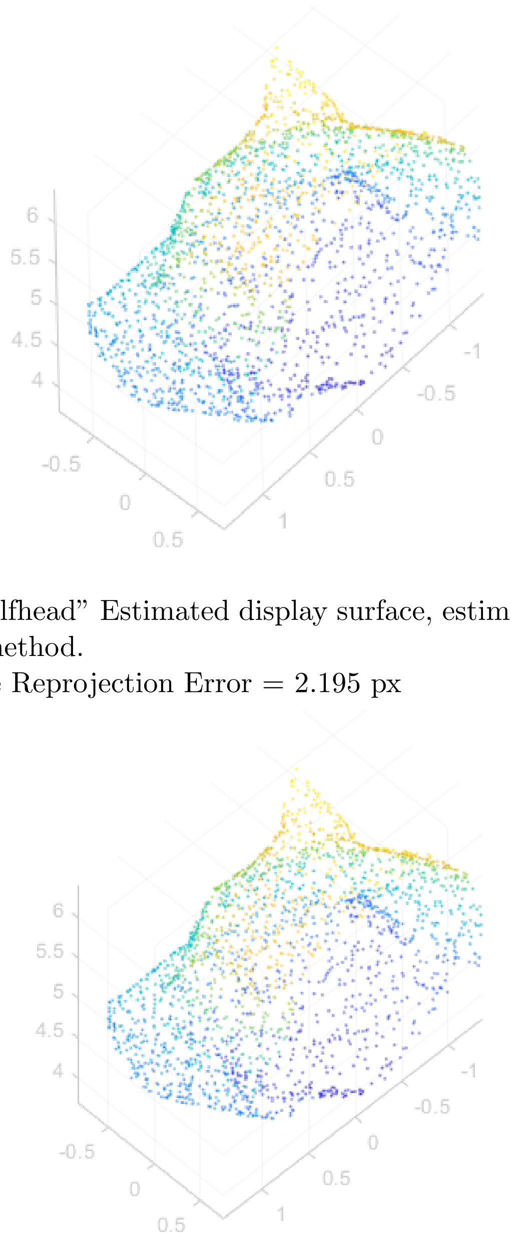
The extrinsic parameter estimates show significant improvement after refinement. The average error in pose estimate drops from 4 degrees to 0.45 degrees at the highest level of pixel noise in the 100 pixel correspondence experiment. Across the 2500 pixel correspondence experiment, the average error in pose estimate is highest at 0.22 degrees across the set of image planes. This consistently good performance suggests that this estimation strategy performs well for extrinsic parameter estimation.



(a) “ZigZag” Estimated display surface, estimated from FMat method.
Average Reprojection Error = 0.5706 px



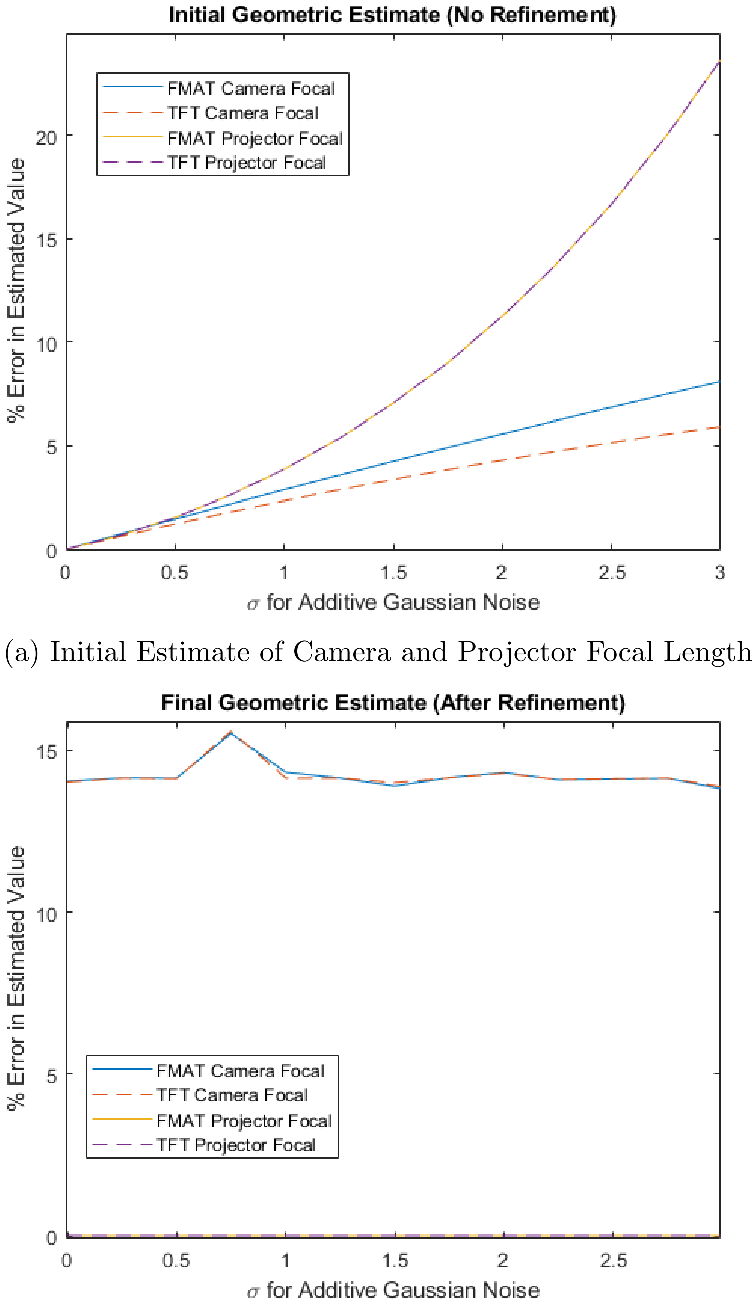
(b) “ZigZag” Estimated display surface, estimated from TFT method.
Average Reprojection Error = 0.7079 px



(c) “Wolfhead” Estimated display surface, estimated from FMat method.
Average Reprojection Error = 2.195 px

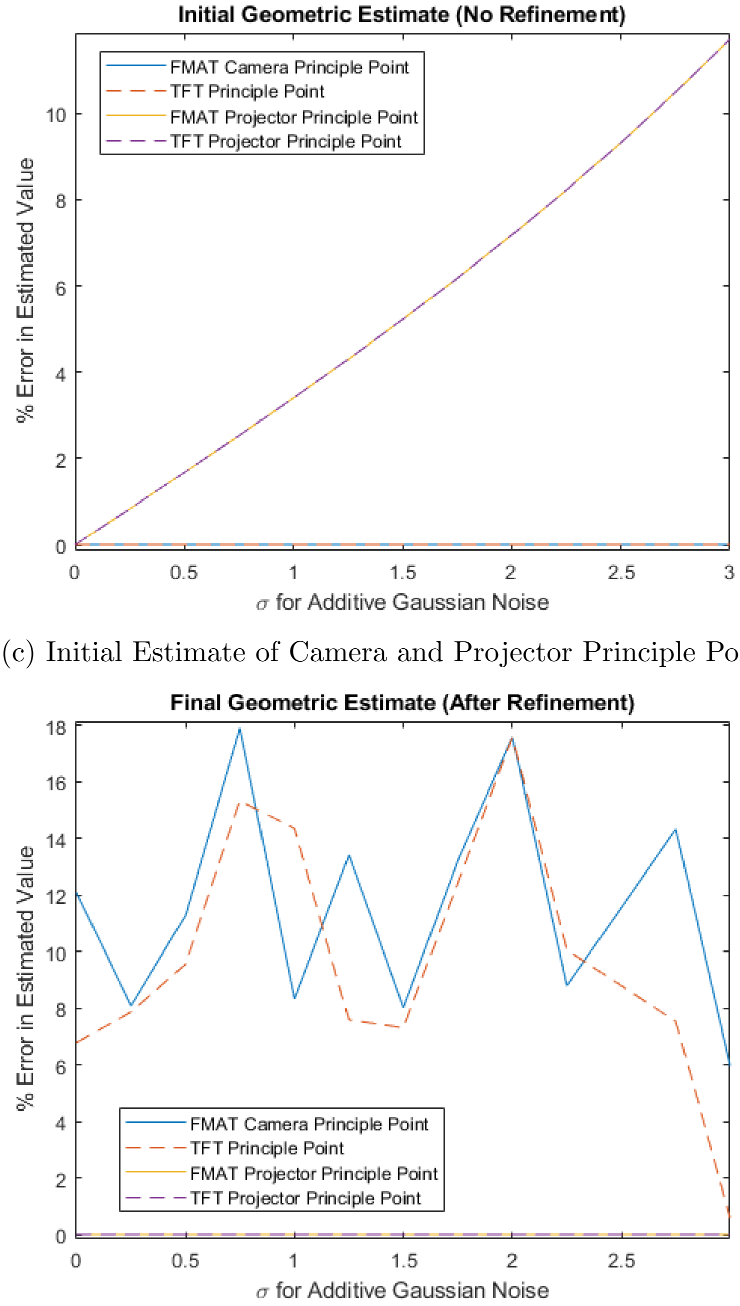
(d) “Wolfhead” Estimated display surface, estimated from TFT method.
Average Reprojection Error = 16.079 px

Figure 6.2: Non-planar display surface estimates and average reprojection error from synthetic data with $\sigma = 1$ Gaussian noise added to sets of 2500 the pixel correspondences.



(a) Initial Estimate of Camera and Projector Focal Length

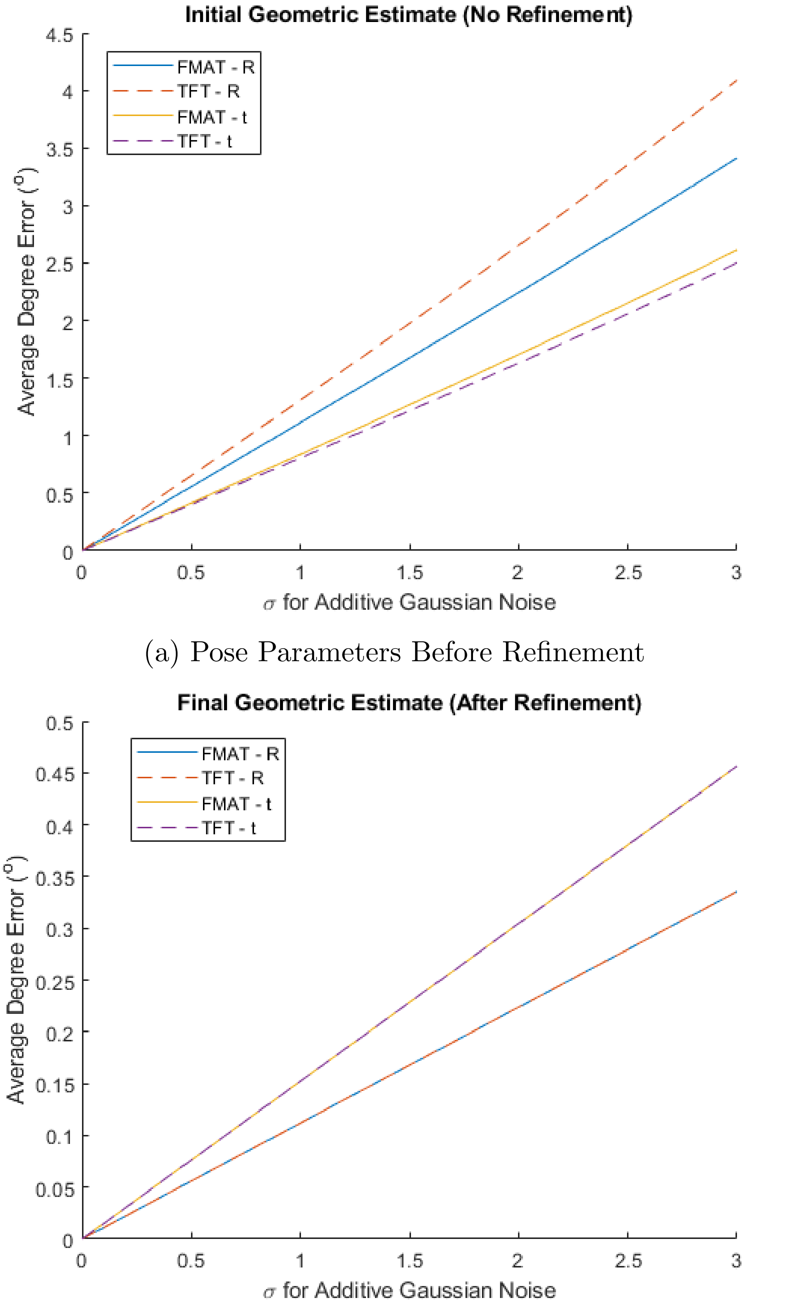
(b) Final Estimate of Camera and Projector Focal Length



(c) Initial Estimate of Camera and Projector Principle Point

(d) Final Estimate of Camera and Projector Principle Point

Figure 6.3: Percent error in estimated camera and projector intrinsic parameters for synthetic data with varying additive Gaussian noise level $0 \leq \sigma \leq 3$ px added to $N = 100$ pixel correspondences, averaged over 40 iterations.



(a) Pose Parameters Before Refinement

(b) Average Degree Error in Final Estimated Pose Parameters

Figure 6.4: Average angular error in estimated camera and projector pose parameters for synthetic data with varying additive Gaussian noise level $0 \leq \sigma \leq 3$ px added to $N = 100$ pixel correspondences, averaged over 40 iterations.

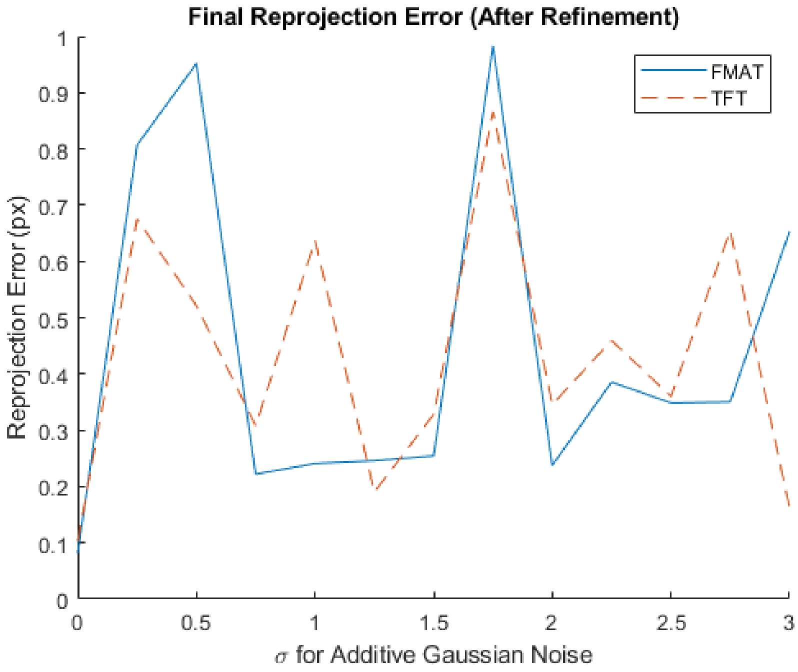


Figure 6.5: Reprojection error in estimated camera and projector parameters and display surface estimate for synthetic data with varying additive Gaussian noise level $0 \leq \sigma \leq 3$ px added to $N = 100$ pixel correspondences, averaged over 40 iterations.

Chapter 7

Conclusion

This thesis proposes an approach that employs a single moving camera, gaining the scene understanding advantages of multiple camera perspectives, the environmental flexibility and adaptability of an unknown camera initialization, and the simplification of requiring only one set of camera intrinsic parameters. A refined estimation for the camera intrinsic and extrinsic parameters is used to generate a set of 3D coordinates representing the display surface reconstruction. The projector calibration is estimated by incorporation of the projector into the display environment estimate described by the surface and camera parameter estimates. The objective of this thesis was to develop a method that was able to calibrate a projector-camera system with unknown cameras and an unknown non-planar display surface. Demonstrated by the comparison in Table 6.3, this is done with similar estimation accuracy as current methods that rely on environment limiting knowledge such as parameter priors or (planar) calibration targets.

Two epipolar geometry based approaches are explored, the fundamental matrix relating two views, and the essential matrix relating three views. The fundamental matrix estimation strategy performed generally better than the trifocal tensor estimation strategy. This result is suggesting that it is sufficient to proceed with the fundamental matrix formulation to generate relations between image pairs, and that no advantage is lost by neglecting the explicit relations available in triplets of images. This result is consistent with results in other vision applications [25]. Both epipolar geometry formulations observed no error in the estimated projector intrinsic parameters. These projector intrinsic parameters had consistently the best estimation performance. The camera intrinsic parameters experienced between $2\% - 27/\%$ error with additive Gaussian noise. The camera intrinsic parameters had the worst performance. The average angular error across all estimated poses is less than 0.5^o at the end of the refinement. The performance of this approach is evaluated over varying additive Gaussian noise in the pixel correspondences, and varying length of pixel correspondences. As noise increases, the accuracy in the estimated scene parameters decreases. Accuracy in the estimated scene parameters is higher with more pixel correspondences.

This moving camera calibration strategy forms the foundation of future handheld camera calibration for non-planar scene estimates and projector calibration. There is also potential for incorporating strategies to rapidly assess scene knowledge and overcome scene occlusions, scene size challenges, or similar camera view challenges.

7.1 Future Directions

There are a number of key places where this method would benefit from future development.

The method relies upon a focal length estimation strategy proposed by Bougnoux [8]. While sufficient in this implementation, this method has been found quite sensitive to noise, and has been unreliable [16, 31]. The method would be improved by replacing this focal length estimation with another way of defining or estimating the focal length.

Parameters within the estimated camera and projector calibration might be held fixed or otherwise linked throughout the bundle adjustment refinement. As the camera images are all from a single moving camera, K_c is adjusted jointly across the image planes to reflect that the intrinsic parameters are consistent for the camera across the images. Alternatively, K_{c_1} might be adjusted separately from K_{c_2} and K_{c_3}, to reflect how a camera moving through a scene might have a varying focal length throughout, or to permit the incorporation of different cameras with different intrinsic parameters.

For the display surface estimate X_E, the 3D coordinates are triangulated once from the current camera parameters. They are then adjusted with the set of camera parameters in a bundle adjustment refinement. Given that the parameters used to estimate X_E are being adjusted themselves, X_E might be reprojected at every iteration rather than held as a separate set of parameters to be adjusted. This might be explored to see if it may produce a better or faster display surface estimate. This method also relies upon simple point triangulation methods, and might benefit from more robust geometric reconstruction strategies [17].

Finally, this method assumes that there has been some matching of the nearest pixel correspondences in each image plane to the virtual representation of each observed 3D key point for the virtual calibration model X_m employed in short bundle adjustment. For the purposes of this thesis this matching has been completed manually across a small set of 10 key points. A fully automatic calibration process would need some feature extraction and matching to produce these relations.

References

[1] *Geometry*. Merriam-Webster, 2022.

[2] A. Alzati and A. Tortora. A geometric approach to the trifocal tensor. *Journal of Mathematical Imaging and Vision*, 38:159–170, 11 2010.

[3] K. Arnold, P. Fieguth, and M. Lamm. 30-3: A moving camera and synthetic calibration target solution for non-planar scene estimation and projector calibration. *SID Symposium Digest of Technical Papers*, 53(1):357–360, 2022.

[4] K. Arnold, P. Fieguth, and M. Lamm. Formulating a moving camera solution for non-planar scene estimation and projector calibration. *Journal of Computational Vision and Imaging Systems*, 7(1):10–12, Apr. 2022.

[5] K. Arnold, M. Naiel, M. Lamm, and P. Fieguth. Evaluation of solving methods for the fundamental matrix computation. *Journal of Computational Vision and Imaging Systems*, 6(1):1–1, Jan. 2021.

[6] J. Bloomenthal and J. Rokne. Homogeneous coordinates. *The Visual Computer*, 11:15–26, 2005.

[7] B. Boudine, S. Kramm, N. El Akkad, A. Bensrhair, A. Saaidi, and K. Satori. A flexible technique based on fundamental matrix for camera self-calibration with variable intrinsic parameters from two views. *Journal of Visual Communication and Image Representation*, 39:40–50, 2016.

[8] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 790–796, 1998.

[9] N. Börlin, A. Murtiyoso, and P. Grussenmeyer. Efficient computation of posterior covariance in bundle adjustment in dbat for projects with large number of object points. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:737–744, 08 2020.

[10] Y. Chen, Y. Chen, and G. Wang. Bundle adjustment revisited, 2019.

[11] C. Engels, H. Stewénius, and D. Nistér. Bundle adjustment rules. *Photogrammetric computer vision*, 2(32), 2006.

[12] G. Falcao, N. Hurtos, and J. Massich. Plane-based calibration of a projector-camera system. *VIBOT Master*, 9, 01 2008.

[13] S. Farsangi, M. Naiel, M. Lamm, and P. Fieguth. Rectification based single-shot structured light for accurate and dense 3d reconstruction. *Journal of Computational Vision and Imaging Systems*, 6(1):1–3, Jan. 2021.

[14] O. Faugeras, Q. Luong, and T. Papadopoulou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, Cambridge, MA, USA, 2001.

[15] T. Fetzer, G. Reis, and D. Stricker. Robust auto-calibration for practical scanning setups from epipolar and trifocal relations. In *2019 16th International Conference on Machine Vision Applications (MVA)*, pages 1–6, 2019.

[16] T. Fetzer, G. Reis, and D. Stricker. Stable intrinsic auto-calibration from fundamental matrices of devices with uncorrelated camera parameters. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.

[17] W. Förstner and B. Wrobel. *Photogrammetric computer vision*. Springer, 2016.

[18] R. Furukawa, G. Nagamatsu, and H. Kawasaki. Simultaneous shape registration and active stereo shape reconstruction using modified bundle adjustment. In *2019 International Conference on 3D Vision (3DV)*, pages 453–462, 2019.

[19] J. Geng. Structured-light 3d surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, Jun 2011.

[20] R. Hartley. A linear method for reconstruction from lines and points. In *Proceedings of IEEE International Conference on Computer Vision*, pages 882–887, 1995.

[21] R. Hartley and C. Silpa-Anan. Reconstruction from two views using approximate calibration. In *Proc. 5th Asian Conf. Comput. Vision*, volume 1, pages 338–343, 2002.

[22] R. Hartley and P. Sturm. Triangulation. *Comput. Vis. Image Underst.*, 68(2):146–157, nov 1997.

[23] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[24] B. Huang, Y. Tang, S. Ozdemir, and H. Ling. A fast and flexible projector-camera calibration system. *IEEE TASE*, 2020.

[25] L. Julia and P. Monasse. A critical review of the trifocal tensor estimation. In *Image and Video Technology: 8th Pacific-Rim Symposium, PSIVT 2017, Wuhan, China, November 20-24, 2017, Revised Selected Papers*, volume 10749, page 337. Springer, 2018.

[26] S. Kim and Y. Choi. *Dynamic Projection Mapping Using Kinect-Based Skeleton Tracking*, pages 60–66. 01 2020.

[27] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, and A. Wong. Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping. *IEEE Transactions on Computational Imaging*, 3(1):74–83, 2017.

[28] J. Lu, J. Zhang, M. Ye, and H. Mi. Review of the calibration of a structured light system. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 4782–4786, 2020.

[29] J. Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[30] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 464–471, 2012.

[31] N. Pellegrino, M. Naiel, M. Lamm, and P. Fieguth. Sensitivity assessment for projector camera geometry reconstruction systems. *Journal of Computational Vision and Imaging Systems*, 5(1):1, Jan. 2020.

[32] T. Petković, S. Gasparini, and T. Pribanić. A note on geometric calibration of multiple cameras and projectors. In *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, pages 1157–1162, 2020.

[33] K. Sadatsharifi. Enhanced detection of point correspondences in single-shot structured light. Master's thesis, University of Waterloo, Department of Systems Design Engineering.

[34] B. Satouri, K. Satori, and A. El Abderrahmani. Genetic algorithms and bundle adjustment for the enhancement of 3d reconstruction. *Multimedia Tools and Applications*, pages 1 – 24, 2020.

[35] Y. Song, Z. Song, and S. Wu. Calibration methods of projector-camera structured light system: A comparative analysis. In *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 39–43, 2018.

[36] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment — a modern synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–372, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

[37] A. Whitehead and G. Roth. Estimating intrinsic camera parameters from the fundamental matrix using an evolutionary approach. *EURASIP Journal on Advances in Signal Processing*, 2004(8):1–12, 2004.

[38] S. Willi and A. Grundhöfer. Robust geometric self-calibration of generic multi-projector camera systems. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 42–51, 2017.

[39] C. Xie, H. Shishido, Y. Kameda, and I. Kitahara. A projector calibration method using a mobile camera for projection mapping system. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 261–262, 2019.

[40] C. Xie, H. Shishido, Y. Kameda, and I Kitahara. Geometric calibration of projector using a mobile camera for spatial augmented reality. *Transactions of the Virtual Reality Society of Japan*, 25(2):138–147, 2020.

[41] S. Yamazaki, M. Mochimaru, and T. Kanade. Simultaneous self-calibration of a projector and a camera using structured light. In *CVPR 2011 WORKSHOPS*, pages 60–67, June 2011.

[42] W. Yang, K.and Fang, Y. Zhao, and N. Deng. Iteratively reweighted midpoint method for fast multiple view triangulation. *IEEE Robotics and Automation Letters*, 4(2):708–715, 2019.

[43] Y. Yin, X. Peng, A. Li, X. Liu, and B. Gao. Calibration of fringe projection profilometry with bundle adjustment strategy. *Opt. Lett.*, 37(4):542–544, Feb 2012.

[44] J. Yu, F. Da, and W. Li. Calibration for camera–projector pairs using spheres. *IEEE Transactions on Image Processing*, 30:783–793, 2021.

[45] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[46] R. Zhu, C. Wang, C. Lin, Z. Wang, and S. Lucey. Object-centric photometric bundle adjustment with deep shape prior. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 894–902, 2018.

[47] S. Zhu, Z.and Wang, H. Zhang, and F. Zhang. Camera–projector system calibration method based on optimal polarization angle. *Optical Engineering*, 59:1, 03 2020.