Projector Self-Calibration using the Dual Absolute Quadric

by

Sepehr Ghavam

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, 2023

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

The applications for projectors have increased dramatically since their origins in cinema. These include augmented reality, information displays, 3D scanning, and even archiving and surgical intervention. One common thread between all of these applications is the necessary step of projector calibration. Projector calibration can be a challenging task, and requires significant effort and preparation to ensure accuracy and fidelity. This is especially true in large scale, multi-projector installations used for projection mapping. Generally, the cameras for projector-camera systems are calibrated off-site, and then used in-field under the assumption that the intrinsics have remained constant. However, the assumption of off-site calibration imposes several hard restrictions. Among these, is that the intrinsics remain invariant between the off-site calibration process and the projector calibration site. This assumption is easily invalidated upon physical impact, or changing of lenses. To address this, camera self-calibration has been proposed for the projector calibration problem. However, current proposed methods suffer from degenerate conditions that are easily encountered in practical projector calibration setups, resulting in undesirable variability and a distinct lack of robustness. In particular, the condition of near-intersecting optical axes of the camera positions used to capture the scene resulted in high variability and significant error in the recovered camera focal lengths. As such, a more robust method was required. To address this issue, an alternative camera self-calibration method is proposed. In this thesis we demonstrate our method of projector calibration with unknown and uncalibrated cameras via autocalibration using the Dual Absolute Quadric (DAQ). This method results in a significantly more robust projector calibration process, especially in the presence of correspondence noise when compared with previous methods. We use the DAQ method to calibrate the cameras using projector-generated correspondences, by upgrading an initial projective calibration to metric, and subsequently calibrating the projector using the recovered metric structure of the scene. Our experiments provide strong evidence of the brittle behaviour of existing methods of projector self-calibration by evaluating them in near-degenerate conditions using both synthetic and real data. Further, they also show that the DAQ can be used successfully to calibrate a projector-camera system and reconstruct the surface used for projection mapping robustly, where previous methods fail.

Acknowledgements

I would like to begin by thanking Prof. Paul Fieguth. His continued patience, expertise and insight at every step of my research helped me to stay focused and inspired. I often joke with friends that Paul is 5 of the 10 smartest people I've ever had the pleasure of working with. I am forever grateful for his mentorship.

Thank you to Prof. John Zelek and Prof. Charbel Azzi for lending their expertise and taking their time to review my thesis.

I would also like to thank Saed Moradi, Mohamed Naiel and Katherine Arnold for their help and continued support during my research.

Finally, my sincere gratitude to the folks at Christie Digital Systems, in particular Mark Lamm, Andrew Dennison, Tony Van Eerd and Matthew Post for helping me develop a deeper appreciation for the systems I have worked with, and for teaching me so much along the way.

Dedication

This is dedicated to my wife Kayli, without who this work would not have been possible. You are not our anchor, but our mast.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis deals with projector calibration, namely, the calibration of projector-camera systems for use in projection mapping. Projector and camera calibration can take two forms, photometric (calibration of color) and geometric (calibration of lens parameters). The purpose of these calibrations is to take visual content, meant to be displayed on a particular surface, and render the content in the domain of each projector so that it is displayed as intended by the artist. This can be as simple as aligning two projectors in a movie theater on a single screen or to increase the brightness of the video.

Projection surfaces may be flat, as in a cinema screen, or they may be curved for a more immersive experience. Alternatively, projection surfaces need not be screens at all; visual content can be projected on buildings or even on statues, depending on the desired effect. All these applications require correct geometric and photometric projector calibration for correctly rendered content. The calibration process ensures that the content rendered for each projector is warped to provide the appearance of originating from a single projector.

Here, we address the problem of geometric projector calibration. In particular, we tackle the problem of geometric projector calibration without priors or calibration artifacts, using a single uncalibrated camera as a sensor.

1.1 Contribution

The major contribution of this thesis is a novel, global approach to projector self-calibration that is robust in the presence of correspondence noise. It provides a reliable and flexible

solution to the significant costs and complexity associated with current projector self-calibration methods.

1.2 Motivation

Beyond their use in displaying movies on the silver screen, projectors serve as a means of turning essentially any surface into a display, without being limited by size. This has made them an excellent instrument in Augmented and Virtual Reality applications (AR/VR), and has garnered significant research interest over the last several decades[32, 31, 25, 11]. They have also been used extensively to build 3D scanning systems[17, 3, 15, 41] for a diverse set of applications, such as archiving[26] and guided surgical intervention[10]. All of these domains rely upon accurate projector calibration, which also has been the subject of extensive academic attention[29, 31, 40, 18, 15].

Projector calibration in an industrial setting has a variety of associated challenges that include technical, economical, and ergonomical. The ability to support a significant variety of system configurations, such as number of projectors, size and scale of scene, and projection surface geometry, is imperative. These installations can have significant associated costs, such as the number of cameras required to support the calibration process (which naturally scales with the number of projectors), or the availability of operators needed to perform these calibration tasks. As a result, the requirement to support a large variety of projector-camera system configurations, using technically sophisticated calibration methods, enabled with sufficiently ergonomical means that operators may perform in reasonable time, while being robust and cost-effective, is a difficult feat to achieve.

The work in this thesis aims to alleviate the complexity and lack of robustness of projector calibration without priors. The setup required for even a single projector is significant in terms of resources, both time and equipment. Notably, the cameras used in such projector calibration tasks require their own calibration, most commonly performed prior to the projector installation, using a calibration artifact[42]. This constrains the projector calibration process, as the recovered camera parameters must stay constant between the initial camera calibration and the subsequent projector calibration, which is not guaranteed. It further limits the operator in the projector calibration process; machine vision cameras often use a lens with a fixed focal length, which, if not suitable to the scene, may need to be changed. These factors support the adoption of camera self-calibration methods for these applications[5, 38, 27, 13]. However, current methods suffer from degenerate conditions that are encountered with regularity, thus preventing consistent and accurate

results, which is examined in great detail within this thesis. As such, a robust and efficient method for projector self-calibration is of great interest to both academia and industry.

1.3 Objectives

The primary objective of this research was to devise a robust method for projector self-calibration that eliminates the need for prior calibration of cameras using a calibration artifact, thus decreasing material costs, setup time, and flexibility. Instead, only the correspondences available in the scene are used to simultaneously recover the calibration parameters of both camera and projector as well as the scene structure.

The secondary objective is to explore the capabilities and limitations of the proposed method alongside previous projector self-calibration methods, examining the relative robustness and tolerance to noise. This research aims not only to advance the understanding of projector calibration methods but also to provide a practical method that can be easily implemented in real-world applications.

1.4 Thesis Outline

This thesis is organized as follows.

- Chapter 2: Background. This chapter introduces the notation and concepts used throughout the thesis. The established techniques and methodologies relevant to camera and projector calibration are discussed, providing a foundation for the rest of the thesis.
- Chapter 3: Problem Formulation. This chapter presents the challenges underlying the primary objective of this thesis. It also outlines the experimental approaches designed to address these challenges and to test the proposed method.
- Chapter 4: Synthetic Data Evaluation. In this chapter, we evaluate the proposed method with synthetic data to understand its behaviour under controlled conditions.
- Chapter 5: Projector Autocalibration. This chapter applies the proposed method to real-world projector calibration problems using real data. The setup, methodology, and results of the experiments are presented and discussed.

- Chapter 6: Conclusion. This final chapter summarizes the main findings, outlines the contributions of the research, and discusses potential directions for future work.

Chapter 2

Background

2.1 The Camera Model

The camera model [12, 21, 35] is the fundamental expression that underpins the entire calibration pipeline: It represents the forward model[21], mapping a view of some scene in the world to an image. Knowing the internal parameters of this forward model permits not just the localization of the camera within the world but also the localization of any detectable point in the real world.

A camera model is simply a projection that transfers points from $\mathbb{R}^3$ to $\mathbb{R}^2$. Models can be split into two broad categories[27]: perspective preserving, and non-perspective preserving (i.e., cameras at infinity). Perspective-preserving camera models map well to physical cameras (e.g., smartphones, single-lens reflex, mirrorless, etc.) where the size of an object in the image is directly related to its distance from the camera. Non-perspective preserving cameras are used often in video games or design software, such as in an isometric view. The generalized equation for a camera model is given by

$$\mathbf{x} = \mathbf{P}\mathbf{X} \tag{2.1}$$

where $\mathbf{X}$ is the homogeneous coordinate of a point in $\mathbb{R}^3$, $\mathbf{x}$ is the corresponding homogeneous coordinate in $\mathbb{R}^2$ in the camera image plane, such that

$$\mathbf{x} = [x, y, w]^\top = w[u, v, 1]^\top \tag{2.2}$$

and $\mathbf{P}$ is a 3×4 projection matrix. Note that while the forward model has a unique solution, in the absence of additional constraints, the inverse model has infinitely many solutions,

due to the dimensionality reduction of $\mathbf{X} \mapsto \mathbf{x}$. For our purposes, we will discuss only perspective-preserving cameras, namely, the pinhole, generalized pinhole, and projective camera models.

In the metric or Euclidean form of the camera matrix, based on the pinhole camera model, the camera matrix $\mathbf{P}$ is a function of the camera pose given by a 3×3 rotation matrix $\mathbf{R}$, the 3×1 camera center $\tilde{\mathbf{C}}$, and a set of internal camera parameters stored in an upper triangular matrix $\mathbf{K}$:

$$\mathbf{P} = \mathbf{K}\big[\,\mathbf{R}\,|\,-\mathbf{R}\tilde{\mathbf{C}}\,\big] \tag{2.3}$$

The general form of the internal camera parameters $\mathbf{K}$, also referred to as the camera intrinsics, is given by

$$\mathbf{K}_{\text{general pinhole}} = \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \tag{2.4}$$

where f_x and f_y represent the focal length in the x and y directions, the skew measure s represents the deviation from orthogonality of the image basis vectors, and the principal point (c_x, c_y) represents the intersection of the central axis of the camera lens with the image plane. With modern, commercial off-the-shelf cameras, we may generally assume square pixels, zero-skew, and an image-centered principal point. In contrast, $c_x = \frac{w}{2}$ and $c_y = \frac{h}{w}$ is *not* a safe assumption for projector calibration, since the projector principal point values are often manipulated to adjust the display, as this is significantly easier than physically moving an often large and cumbersome projector.

2.1.1 The Pinhole Camera Model

The pinhole camera model is the simplest of the camera models, with unit aspect ratio, no skew and the image origin aligned with its principal axis:

$$\mathbf{K}_{\text{pinhole}} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \tag{2.5}$$

The generalized pinhole camera model expands on this, allowing for varying aspect ratio between x and y axes, and image coordinates with the origin at the top-left corner of the image:

For the purposes of our calibration problem, we can simplify Equation (2.4) by making several assumptions. We assume that these cameras are standard, commercially available

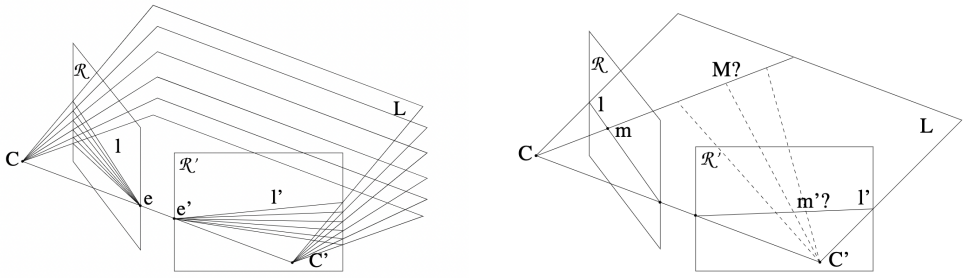


Figure 2.1: A point visible in a single camera view is constrained to a line in $\mathbb{R}^3$. If that point is visible in another view, the intersection of the two rays constrain it to a set of coordinates $\mathbf{X}$. Any triangulated point between two views is coplanar with both the camera centers and the epipoles of the stereo configuration. Figure copied from [27].

cameras with a fixed focal length. We can assume that the pixels are square, i.e. $f_x = f_y$ and with no skew, $s = 0$. We may also assume that the principal point lies approximately in the center of the image, such that $c_x = \frac{\mathrm{w}}{2}$ and $c_y = \frac{\mathrm{h}}{2}$ where w, h represent the camera resolution. Applying a transformation to set the image center to the origin of the image coordinate frame, leaves us with only a single parameter to estimate, f. We can reduce the general model of Equation (2.5).

2.2 Geometry of the Image Pair

A pair of cameras observing the same scene is the simplest configuration with which structure and motion can be estimated. To estimate the location $\mathbf{X} \in \mathbb{R}^3$ of an image point, at least one other view of that same image point is required. A single camera view constrains a detected feature to a single ray originating from the camera center out to infinity. A second camera reduces the solution space to a smaller volume of uncertainty, since the the rays generally will not intersect perfectly (see Figure 2.2). Hence the reconstruction of any point remains probabilistic, and is greatly improved by an increase in the camera baseline for the two-view configuration. Note that a pixel is essentially a frustum in $\mathbb{R}^3$. Assuming a constant baseline, as viewing angles of the two cameras become increasingly parallel, the uncertainty volume of the reconstructed point grows, as shown in Figure Figure 2.2.

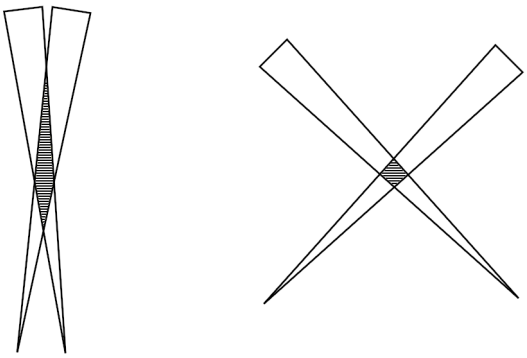


Figure 2.2: The uncertainty volume in two-view geometry. Note that as the distance of the correspondence relative to the stereo baseline increases, the volume of uncertainty - shown in grey - increases dramatically. This tends to produce anistropic errors in triangulation of points in $\mathbb{R}^3$, with errors significantly higher in the z axis of the stereo camera pair.

2.2.1 The Fundamental Matrix

In the two-view configuration with cameras C and C', the fundamental matrix relates two image views by satisfying the following equation:

$$\mathbf{x}^\top \mathbf{F} \mathbf{x}' = 0 \tag{2.6}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are homogeneous image coordinate vectors in the first and second view, respectively. The forward model of $\mathbf{F}$ for two known camera views is:

$$\mathbf{F} = \left(\mathbf{K}'^{-1}\right)^\top \mathbf{R}[\mathbf{t}]_\times \left(\mathbf{K}^{-1}\right) \tag{2.7}$$

where $\mathbf{K}$ and $\mathbf{K}'$ are the intrinsic matrices of the first and second views, respectively, $\mathbf{R}$ is the rotation matrix that transforms C to C', and $[t]_\times$ is the translation from C to C' in skew-symmetric form. Hence, given known intrinsics for both cameras, the relative pose can be trivially estimated. [1]

2.3 Camera Matrix Estimation

We have a set of n points for the structure of our observed scene $\mathcal{X} = \{\mathbf{X}\}$, where $\mathbf{X} = [X, Y, Z, 1]^\top$. There are corresponding homogeneous image coordinates $\mathcal{U} = \{\mathbf{x}\}$, where

[1] The product of $\mathbf{R}[\mathbf{t}_\times]$ is also known as the Essential Matrix [12].

$\mathbf{x} = [x, y, w]^\top = w[u, v, 1]^\top$. The points are captured by an unknown camera C with projection matrix $\mathbf{P}_C$, where

$$\mathbf{P}_C = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{bmatrix} \tag{2.8}$$

The mapping of $\mathbf{X} \mapsto \mathbf{x}$ follows the forward model of Equation (2.1). We may use the Direct Linear Transform (DLT) algorithm [12] to produce a set linear equations of the form $\mathbf{A}z = \mathbf{0}$, such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & w_i\mathbf{X}_i & v_i\mathbf{X}_i \\ w_i\mathbf{X}_i & \mathbf{0} & u_i\mathbf{X}_i \end{pmatrix}_{2n\times 12} \tag{2.9}$$

where z is a 12×1 vector reordering of the elements of $\mathbf{P}_C$:

$$z = [p_1\, p_2\, \ldots\, p_{12}]^\top \tag{2.10}$$

The linear system is solved by identifying the nullspace of $\mathbf{A}$ [12].

A camera matrix has 12 elements and 11 degrees of freedom (the one degree lost because of scale ambiguity). Each set of $\mathbb{R}^2 \to \mathbb{R}^3$ correspondences gives two equations, hence we need five complete $\mathbf{X} \mapsto \mathbf{x}$ correspondences, and either the x or y coordinate of a sixth to fully constrain the system. In practical applications of structured-light with $\gg 6$ such correspondences, leading to an overdefined $\mathbf{A}$, we solve for $\mathbf{P}_C$ by taking the unit eigenvector corresponding to the smallest eigenvalue of $A^\top A$.

2.4 Structured Light

Multiple view geometry, for any number of cameras, requires correspondences to be detected between each image. Classical Structure-from-Motion methods detect these correspondences using passive features, with descriptors such as HOG (Histogram of Oriented Gradients)[8], SIFT (Scale-Invariant Feature Transform)[20], ORB (Oriented FAST and Rotated BRIEF)[30] or AKAZE (Accelerated KAZE)[2]. For active-stereo configurations such as projector-camera systems, the features are created using the projector, which displays spatially- or temporally-encoded Structured Light (SL)[33][32][9].

There are various options for generating correspondences through SL patterns, each with their respective trade-offs. Traditional projector calibration, particularly for pixel-dense 3D reconstruction uses multi-frame SL patterns. Multi-frame patterns like Gray

Code binary patterns (shown in Figure 2.3) can be used to encode unique addresses for each projector pixel, thus producing a densely reconstructed surface. This is preferable for surfaces where smoothness and continuity cannot be assumed.

In cases where assumptions of smoothness and surface continuity can be made, the SL patterns used can be sparse; either multi-frame blobs or spatially encoded patterns (see Figure 2.5 and Figure 2.6. Such configurations can be leveraged to improve the speed and quality of the calibration and reconstruction. For example, one can reduce the solution space from all $\mathbb{R}^3$ to geometric primitives, such as a plane, partial cylinder, or dome. The sparseness is not a limiting factor for projection mapping, as points between explicit correspondences can be interpolated, allowing for a smooth warping of the projector image to the display.

2.4.1 Multi-frame Structured Light

Multi-frame — or temporally-encoded — SL produces a pixel-dense point cloud, but does so at a cost: the image acquisition takes time. The number of required frames is defined by the minimum number of bits needed to uniquely encode every projector pixel, given by:

$$\left\lceil \log_2(h_P) \right\rceil + \left\lceil \log_2(w_P) \right\rceil \tag{2.11}$$

where w_P and h_P are the width and height of the projector in pixels, respectively. Moreover, any cameras used must remain fixed, increasing the price, network traffic, memory requirements, compute and setup times for any such systems.

There is an additional cost that may be incurred as a function of the scene surface properties. Multi-frame binary patterns can be susceptible to signal duplication that can arise from non-lambertian surfaces.

An example of projected Gray Code patterns can be seen in Figure 2.4. Note the ambiguity along the vertical axis; each image axis requires a set of Gray Code frames, as indicated by Equation (2.11). In order to decode each feature descriptor in Gray Code binary patterns, each camera pixel is evaluated along the temporal axis t, such that the binary signal created by the multi-frame SL acts as a feature. Often the projector resolution will be exceeded by the camera resolution, and the one-to-many projector-to-camera pixel mapping requires a centroid calculation. See Figure 2.3 for the patterns along a single axis.

Upon complete capture from each camera position of all necessary structured light frames, the correspondences are then extracted using the algorithm defined in Algorithm 1. The algorithm assumes that the resolution of the camera used in capturing of structured

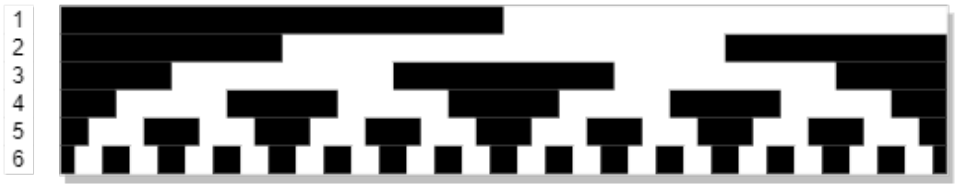


Figure 2.3: 6 Bit Gray Code Patterns

light frames is greater than that of the projector emitting those frames. Therefore, each SL feature in the projector domain should map to multiple camera pixels, hence the need for the centroid of each SL feature correspondence in the camera domain.

2.4.2 Single-Shot Structured Light

An alternative to temporally encoded SL is single-shot, or spatially-encoded SL[32]. In such patterns the feature descriptors are defined spatially and must be detected. An example of this can be seen in Figure 2.6, which shows a set of 6 binary tags marking the vocabulary, and a 3×3 unique address block that defines the correspondence. The tradeoff is self-evident: a faster image acquisition process and a time-independent signal at the cost of structure density. These SL patterns are particularly well suited for applications involving a single moving camera, for which the constraints of multi-frame patterns would be prohibitive.

2.5 Geometric Strata

When performing a camera calibration before scene reconstruction or camera localization, the results will lie within Euclidean space and, by definition, will map to the scene geometry by an $SE(3)$ transform consisting of a rotation $\mathbf{R}$ and translation $\mathbf{t}$.

The structure generated from ≥ 2 camera views lies in one of a set of geometric strata. These strata define the constraints that apply to the transformation of the structure relative to its real-world counterpart. All such general linear transformations are defined as a 4×4 matrix, denoted by ${}^{0}\mathbf{T}_{1}$ for a transform from inertial frame $0 \rightarrow 1$.

Structure and motion obtained from a multi-view configuration will generally occupy one of the following geometric strata: *projective*, *affine*, *metric*, or *euclidean*[27]. The

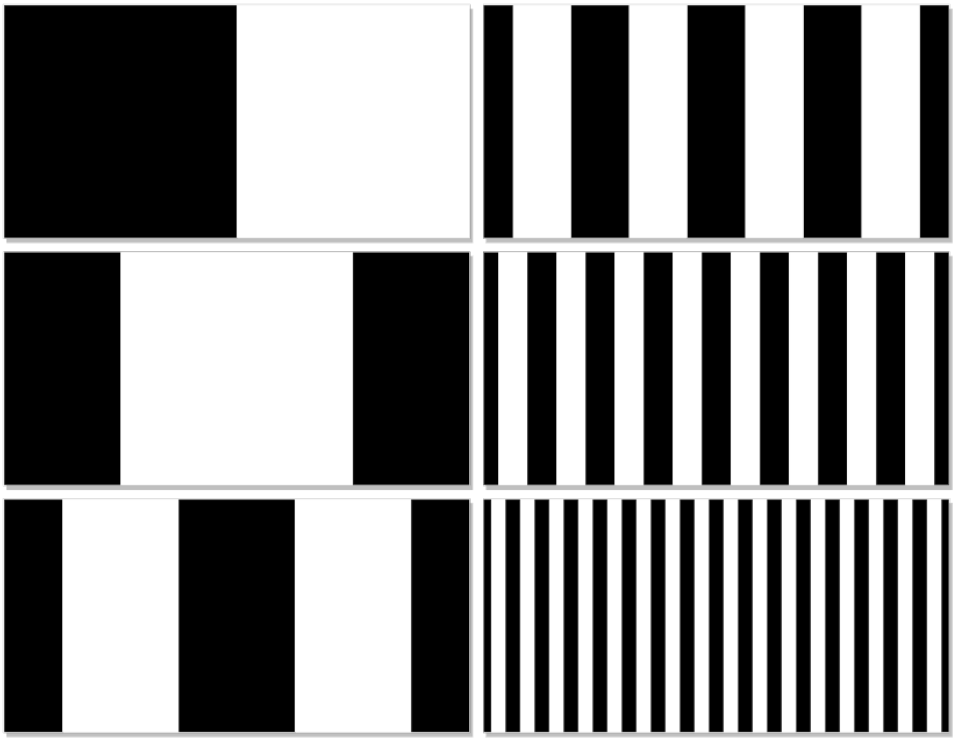


Figure 2.4: 6 bit vertical Gray Code frames.

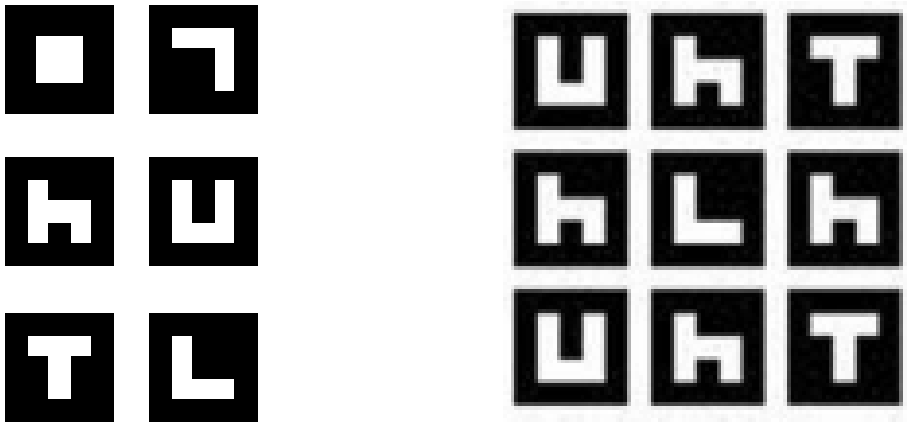


Figure 2.5: Tag-based, spatially encoded structured light pattern. Tag vocabulary (left) and example address block (right). Image taken from [9]

Algorithm 1 Gray Code Structured Light Correspondence Gather

Require: Set of SL frames $F_i = \{f_1, f_2, ..., f_n\}$ for each camera

Ensure: Correspondence table V containing correspondence matches between cameras and projector for each structured light feature

Concatenate all structured light frames in projector domain along dimension 3 to produce a 3D matrix S_p
for each pixel p in the S_p **do**
    Read the binary code b_p along the temporal axis of p
    Store this binary code as the key in a lookup map M mapped to the projector domain coordinates
end for
for each set of SL frames F_i captured by camera C_i **do**
    Apply a binary threshold v to all frames $(0 < v < 1)$
    Concatenate all frames into a 3D matrix S_i
    for each pixel p in the camera-view stack S_i **do**
        Read the binary code b_p along the temporal axis of S_i
        Determine correspondence c_p of the binary code b_p to the projector domain using M
        Append c_p to an intermediate list of correspondences C' for binary address b_p
    end for
    for each correspondence feature b_p mapped in V' **do**
        Calculate the centroid g of all coordinates mapped to b_p in V'
        Add g to list of correspondences V for feature b_p in camera C_i
    end for
end for
return correspondences V

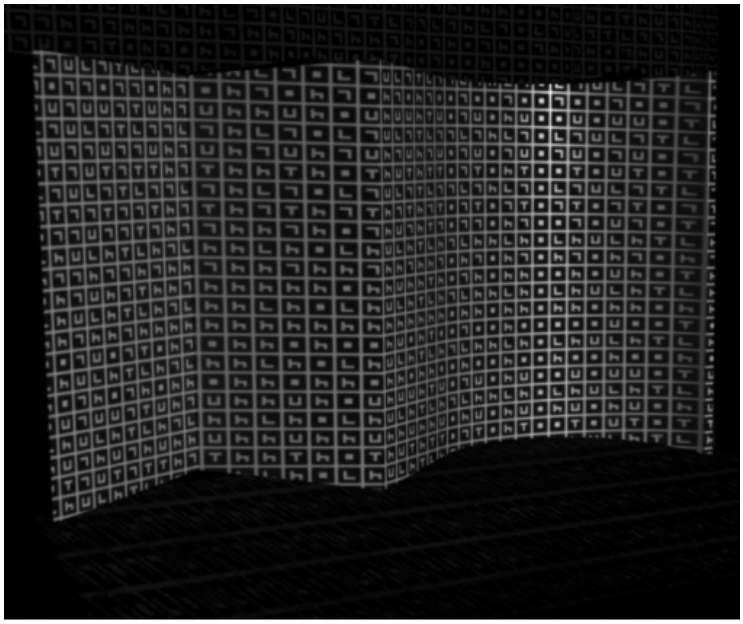


Figure 2.6: Example of tag-based SL. Image taken from [9]

least constrained of these strata, the projective stratum, can be estimated for any multi-view geometry configuration with no prior knowledge of camera(s) or scene[28, 38]. The calibration process is described in detail in Section 2.6. The transform from projective to euclidean space has the following form:

$$\mathbf{H}_{\text{projective}} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \tag{2.12}$$

which has 15 degrees of freedom as it is valid up scale. The desired calibration result is an upgrade to the metric stratum, which maps to the real-world scene by a similarity transform ${}^{M}\mathbf{T}_{W}$ with 7 DoF:

$$\mathbf{H}_{\text{metric}} = \begin{bmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_1 \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_2 \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.13}$$

In many cases, particularly in SL-based reconstructions, any cameras used are calibrated in advance (generally using Zhang's method), producing a Euclidean structure where the

transformation ${}^{W}\mathbf{T}_{E}$ is in the $SE(3)$ Lie group. These inertial frames are therefore related by a rotation $\mathbf{R}$ and translation $\underline{t}$:

$$\mathbf{H}_{\text{euclidean}} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.14}$$

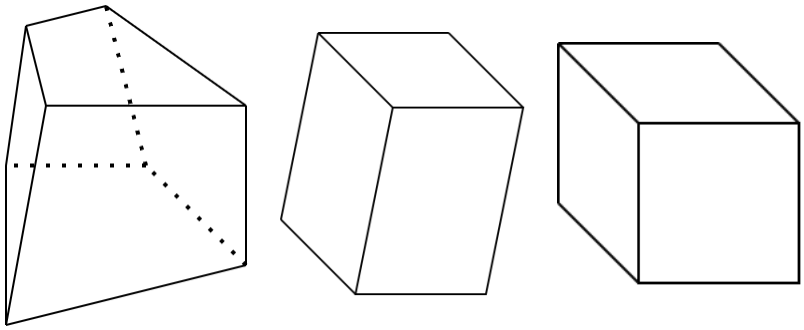


Figure 2.7: Geometric strata examples, from left to right: projective, affine, metric.

Performing the camera calibration using a calibration artifact such as a checkerboard pattern (see Figure 2.8) produces a reconstruction scaled correctly to the real-world scene being scanned. Hence the relationship in Equation (2.14) has only 6 degrees of freedom, for rotation (euler angles ϕ, θ, ψ) and translation (position x, y, z).

2.6 Projective Calibration

Deriving the projection matrix from Equation (2.1) is difficult due to the ambiguity implicit in the multiplication of the forward model. Take any nonsingular 4×4 matrix $\mathbf{H}$. Starting with the forward model, both the calibration matrix $\mathbf{P}$ and structure $\mathbf{X}$ can be transformed by $\mathbf{H}$ while preserving the same image point coordinates $\mathbf{x}$:

$$\begin{aligned} \mathbf{x} &= \mathbf{P}\mathbf{X} \\ &= \mathbf{P}\mathbf{H}\mathbf{H}^{-1}\mathbf{X} \\ &= (\mathbf{P}\mathbf{H})(\mathbf{H}^{-1}\mathbf{X}) \\ &= \hat{\mathbf{P}}\hat{\mathbf{X}} \end{aligned} \tag{2.15}$$

Since Equation (2.15) is true for any nonsingular $\mathbf{H}$ (since $\mathbf{H}\mathbf{H}^{-1} = \mathbf{I}$), we are left with an infinite solution space when lacking additional constraints necessary to solve for $\mathbf{H}$. In order to recover the metric calibration parameters, we initialize our calibration in projective space (with some arbitrary nonsingular $\mathbf{H}$ applied) and upgrade to metric by localizing the plane at infinity $\mathbf{\Pi}_\infty$ and thus recovering $\mathbf{H}$.

2.6.1 Incremental Approach

The projective calibration method developed in [27] is an incremental approach to projective structure and motion that begins with a stereo image pair. To begin the projective calibration from a sequence of images, an image pair must be selected from which to initialize the reconstruction. This initial stereo pair is selected by jointly maximizing two criteria: feature richness and pose difference. Note that these criteria are necessarily inversely correlated, as greater pose difference implies dissimilar views and therefore fewer matching features, and vice versa. The author of [27] proposes maximizing the product of the number of feature matches with the distance measure δ to be described in Equation (2.20).

It should be noted that the uncertainty volume of a reconstructed point $\mathbf{X} \in \mathbb{R}^3$ is directly related to the relative angle between the two views. Unfortunately, in a projective reconstruction the relative pose is unknown. However, we can still optimize for the greatest pose difference through a proxy measure. The method proposed by [27] is to use the reprojection error following a planar-homography $\mathbf{H}$ transformation for each pair of views as the distance measure. The homography is defined by

$$\mathbf{H}^{ij} = [\mathbf{e}]_\times \mathbf{F} + \mathbf{e}\mathbf{a}_{\min}^\top, \quad \mathbf{H}^{ij} : C_i \mapsto C_j \tag{2.16}$$

such that the homography $\mathbf{H}$ provides an approximation of $\mathbf{X}'$:

$$\mathbf{H}\mathbf{X}_k^i \approx \mathbf{X}_k^j \tag{2.17}$$

In equation Equation (2.16), $\mathbf{e}$ is the the homogeneous coordinates of the epipole showing the camera center of C_j in the image plane of C_i, and $[\mathbf{e}]_\times$ is the skew symmetric form of the epipole, and $\mathbf{F}$ is the Fundamental Matrix that satisfies the criterion

$$\mathbf{X}^{i\top}\mathbf{F}\mathbf{X}^j = 0 \tag{2.18}$$

The value of $\mathbf{a}_{\min}$ is then found on the basis of the following minimization:

$$\mathbf{a}_{\min}^{ij} = \arg_{\mathbf{a}}\min \sum_i \left\| ([\mathbf{e}]_\times \mathbf{F} + \mathbf{e}\mathbf{a}^\top)\mathbf{X}_k^i - \mathbf{X}_k^j \right\| \tag{2.19}$$

Here $\mathbf{X}_k^i, \mathbf{X}_k^j$ are the kth point in image i and j. This homography is then used to define a pose-difference measure between two views:

$$\delta^{ij} = \text{median}\Big(\big\{\delta_k\big\}^{ij}\Big), \quad \delta_k^{ij} = \Big\|\mathbf{H}\mathbf{X}_k^i - \mathbf{X}_k^j\Big\| \tag{2.20}$$

With the initial stereo-pair selected, arbitrary projection matrices can be assigned. The initialization for these projection matrices is

$$\begin{aligned} \mathbf{P}_1 &= \begin{bmatrix} \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 1} \end{bmatrix} \\ \mathbf{P}_2 &= \begin{bmatrix} [\mathbf{e}_{12}]_\times \mathbf{F}_{12} + \mathbf{e}_{12}\mathbf{a}^\top & \sigma \mathbf{e}_{12} \end{bmatrix} \end{aligned} \tag{2.21}$$

In general, the projection matrix of the camera used as the frame of reference is aligned with the global coordinate system. The reconstruction is performed using the n-view triangulation method used in [24].

With the reconstruction initialized, it can be extended by traversing the scene graph and solving for the camera matrix of each view using the Direct Linear Transform algorithm outlined in Section 2.3, and triangulating other remaining points again using the method in [24]. This process is repeated until all camera matrices have been estimated and all points with ≥ 2 views have been triangulated.

2.6.2 Factorization Approach

The method of projective calibration used by [37] is a two-step reconstruction process. The preqequisite to this approach is a complete scene graph[2]. This method builds a complete observation matrix of correspondences $\mathbf{W}$ that can be factorized into projective structure $\mathbf{X}_j$ and motion and $\mathbf{P}_i$:

$$\mathbf{W} = \begin{bmatrix} \lambda_{11}x_{11} & \lambda_{12}x_{12} & \dots & \lambda_{1n}x_{1n} \\ \vdots & & & \vdots \\ \lambda_{m1}x_{m1} & \lambda_{m2}x_{m2} & \dots & \lambda_{mn}x_{mn} \end{bmatrix} = \begin{pmatrix} \hat{\mathbf{P}}_1 \\ \hat{\mathbf{P}}_2 \\ \vdots \\ \hat{\mathbf{P}}_m \end{pmatrix} \begin{pmatrix} \hat{\mathbf{X}}_1 & \hat{\mathbf{X}}_2 & \dots & \hat{\mathbf{X}}_n \end{pmatrix} \tag{2.22}$$

If the observations in $\mathbf{W}$ are free of noise, then nothing further need be done. However, considering that the correspondences will have some associated noise, the observation matrix needs to be normalized column-wise and along "triplets-of-rows"[36]. The first step

[2] A graph where cameras are represented as nodes, and the edges indicate the presence of shared correspondences

is described [36, 22], is to recover the projective depths λ_p^i of each correspondence point. These are estimated according to

$$\lambda_p^i = \frac{(\mathbf{e}^{ic} \times \mathbf{u}_p^i) \cdot (\mathbf{F}^{ic}\mathbf{u}_p^c)}{||\mathbf{e}^{ic} \times \mathbf{u}_p^i||^2}\lambda_p^c \tag{2.23}$$

where $\mathbf{e}^{ic}$ represents the epipole, $\mathbf{F}^{ic}$ the Fundamental Matrix, and $\lambda_p^c = 1$ for all known points $\mathbf{u}_p$.

The second step, following the projective depth estimation, is the filling of missing elements for unknown depths. In cases where depth cannot be estimated directly due to occlusions, the missing points may be estimated. The interested reader may find more information on this in [37, 36, 22].

2.7 Camera Autocalibration

The recovery of the internal parameters of a camera is generally performed through Zhang's method[42], using a calibration artifact with known 3D structure, such as a checkerboard(see Figure 2.8). This method allows for high accuracy estimation of camera parameters in Euclidean space, as the dimensions of the checkerboard are known prior to calibration. The simplicity of this approach has made this calibration method arguiably the most ubiquitous camera calibration method.

The process of recovering camera parameters in the absence of such a calibration artifact is known as autocalibration[5, 38, 27, 13], or self-calibration. There are ways to avoid the standard camera calibration approach, and circumstances in which that may be desirable, such as with a zoom lens, or any application that requires the use of unknown cameras. While standard camera calibration involves the use of a calibration artifact with known geometry in order to apply constraints and recover the Euclidean calibration parameters of the camera, with autocalibration these constraints are derived by other means. We employ a set of theoretical geometric entities that remain invariant under rigid motion. These geometric entities are the *plane at infinity* Π_∞ and the *absolute conic* ω_∞^*[28], as well as their derivatives, the *image of the absolute conic* and its dual[14], and the *absolute quadric* and its dual[38].

In the process of self-calibration, the initial reconstruction can only occur in projective space. The forward model of image projection, as shown in Section 2.1 has an implicit ambiguity discussed in see Equation (2.15). This ambiguity prevents us from solving for the metric calibration directly, the solution begins with a projective calibration. Once the

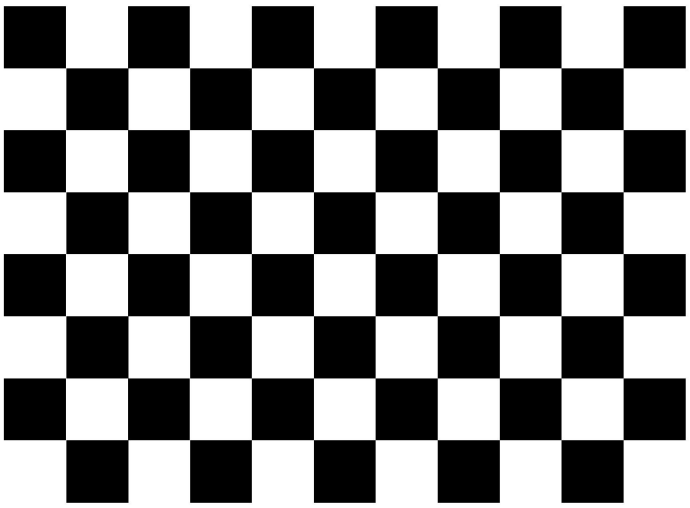


Figure 2.8: Example of a standard checkerboard used for camera calibration using Zhang's method.

projective reconstruction is complete across the entire scene graph, we may constrain the ambiguity to affine or metric by special means.

2.7.1 The Bougnoux Equations

The Bougnoux equations[5] provide an elegant and computationally inexpensive method for the extraction of focal lengths for a stereo camera pair, assuming fairly good principal point estimates for both cameras. The equations are given by:

$$\begin{aligned} f_1^2 &= -\frac{p_2^\top [e_2]_\times \tilde{\mathbf{I}}\mathbf{F}^\top p_1 p_1^\top \mathbf{F} p_2}{p_2 [e_2]_\times \tilde{\mathbf{I}}\mathbf{F}\tilde{\mathbf{I}}\mathbf{F}^\top p_2} \\ f_2^2 &= -\frac{p_1^\top [e_1]_\times \tilde{\mathbf{I}}\mathbf{F}^\top p_2 p_2^\top \mathbf{F} p_1}{p_1 [e_1]_\times \tilde{\mathbf{I}}\mathbf{F}\tilde{\mathbf{I}}\mathbf{F}^\top p_1} \end{aligned} \tag{2.24}$$

where p_i is the principal point and $[e_i]$ the epipole in skew-symmetric form of camera i, and $\tilde{\mathbf{I}}$ is the matrix $\mathrm{diag}(1, 1, 0)$. Given this form, it is an appealing approach for partial autocalibration as most commercially available cameras have an approximately image-centered principal point. However it was pointed out in [23] that this method suffers from a degeneracy when the camera optical axes intersect. While perfect intersection of optical

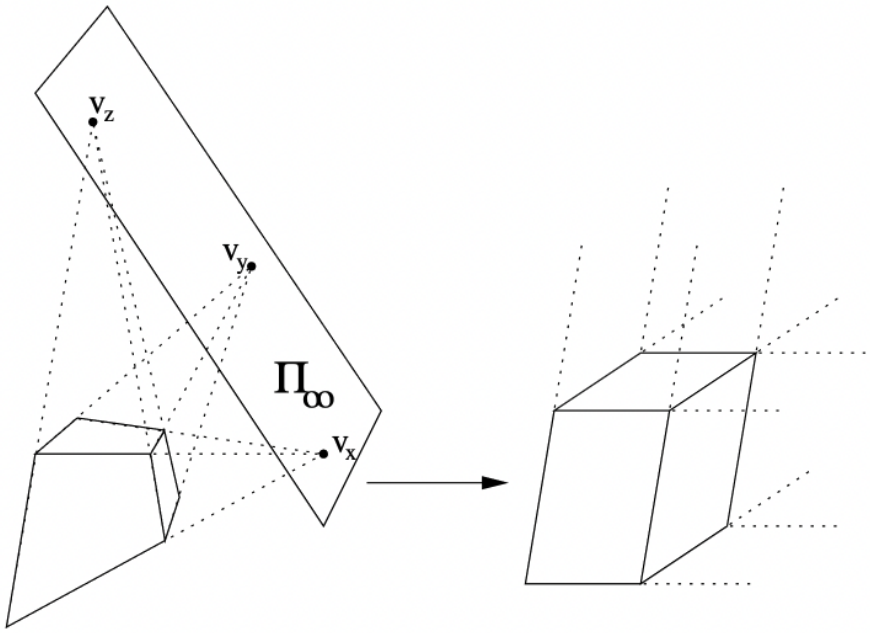


Figure 2.9: The plane at infinity in projective (left) and affine (right) spaces. Figure copied from [27]

axes is inherently only possible in synthetic data, near-intersection is quite common and is of interest for these applications. The behaviour of the Bougnoux equations in such near-intersection cases is investigated in greater detail in Chapter 4 of this thesis.

2.7.2 The Plane at Infinity

The plane at infinity $\mathbf{\Pi}_\infty$ is a theoretic geometric entity that remains invariant under all rigid affine transforms (and consequently under metric and Euclidean transforms), but may be moved under projective transformations[28]. Its canonical position in affine, metric and Euclidean spaces is $[0, 0, 0, 1]^\top$. Further, the geometric interpretation of $\mathbf{\Pi}_\infty$ asserts that the parallel lines of any geometric primitives intersect only at $\mathbf{\Pi}_\infty$. This property of $\mathbf{\Pi}_\infty$ explains its movement away from its canonical position under projective transformations; see Figure 2.9.

Since $\mathbf{\Pi}_\infty$ only varies under projective transformations, identifying the position of $\mathbf{\Pi}_\infty$

in projective space is equivalent to performing an affine calibration. The transformation of $\mathbf{\Pi}_\infty$ to its canonical position can be applied to upgrade the camera matrices $\mathbf{P}_i^P$. This approach is an incremental upgrade, first solving the non-linear problem of locating the plane at infinity, followed by a metric upgrade.

2.7.3 The Absolute Conic

The absolute conic $\mathbf{\Omega}_\infty$ and its image, ω_∞ are elements related to the $\mathbf{\Pi}_\infty$. The ω_∞ is the intersection of $\mathbf{\Omega}_\infty$ with $\mathbf{\Pi}_\infty$, which is invariant under rigid motion and arbitrary rotation[]. The camera intrinsics are encoded in ω_∞ according to

$$\omega \sim \mathbf{K}^{-\top}\mathbf{K}^{-1} \tag{2.25}$$

and its dual, ω_∞^* is given by $\omega \sim \mathbf{K}\mathbf{K}^\top$. Hence solving for the Dual Image of the Absolute Conic (DIAC) ω_∞^* is equivalent to a metric calibration.

2.7.4 The Dual Absolute Quadric

The Dual Absolute Quadric (DAQ) Ω_∞^* is quadric surface whose null vector is the plane at infinity $\mathbf{\Pi}_\infty$:[3]

$$\mathbf{\Pi}_\infty{}^\top \Omega_\infty^* \mathbf{\Pi}_\infty = 0 \tag{2.26}$$

Ω_∞^* is a degenerate, symmetric 4×4 matrix of rank 3 that encodes for metric structure and is invariant under Euclidean transformation. In Euclidean space, it takes the following form:

$$\Omega_{\infty\text{euc}}^* = \begin{pmatrix} \mathbf{I}_{3\times3} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \tag{2.27}$$

Note that the canonical position of $\mathbf{\Pi}_\infty$ in metric space is at $[0,0,0,1]^\top$, which satisfies Equation (2.26). In a warped geometric space, such as the projective space occupied by a projective reconstruction as described in Section 2.6, the Ω_∞^* will map to its metric invariant form by a 4×4 linear transform $\mathbf{H}$ (see Equation (2.15)). This transformation is split into two matrices $\mathbf{B}$ and $\mathbf{c}$ [37]:

$$\mathbf{H} = \begin{bmatrix} \mathbf{B}_{4\times3} & \mathbf{c}_{4\times1} \end{bmatrix} \tag{2.28}$$

[3] For a more in-depth treatment of the DAQ, see [38]

This transform will apply to all the $\mathbf{P}$ across all of the views, assuming constant intrinsics, upgrading them to metric space:

$$\omega^* \sim \mathbf{K}\mathbf{K}^\top \sim \mathbf{P}\mathbf{Q}_\infty^*\mathbf{P}^\top \tag{2.29}$$

In the following section we discuss the use of Equation (2.29) to establish a set of linear constraints to easily estimate Ω_∞^*.

2.8 Camera Calibration using the Dual Absolute Quadric

The structure and motion is initialized via the Factorization method in Equation (2.22). This approach makes use of all the available data, and tends to have better statistical behaviour than the incremental approach, which is subject to the quality of the initial stereo pair.

The expression in Equation (2.29) provides the relationship between the absolute conic via the dual of its image, ω_∞^* (and through it the camera intrinsics $\mathbf{K}$), and the DAQ. To simplify the set of linear equations derived from ω_∞^* to solve for Ω_∞^*, we modify the former by asserting some general prior knowledge about the camera used to capture the image sequence, such as unit aspect ratio ($f_x = f_y$), zero skew, and an image-centered principal point:

$$c_x = \frac{\text{width}}{2} \qquad c_y = \frac{\text{height}}{2} \qquad s = 0 \tag{2.30}$$

We shift correspondences so that the image center falls at $(0, 0)$. Hence, we are left with a diagonal, symmetric matrix ω_∞^*

$$\omega^* = \mathbf{K}\mathbf{K}^\intercal = \begin{bmatrix} f^2 & 0 & 0 \\ 0 & f^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.31}$$

We can use the simplified form of Equation (2.31) to build a set of linear constraints that can be solved using linear least-squares:

$$\begin{aligned} P_i^{(1)}\Omega_\infty^* P_i^{(1)\intercal} &= P_i^{(2)}\Omega_\infty^* P_i^{(2)\intercal} \\ P_i^{(1)}\Omega_\infty^* P_i^{(2)\intercal} &= 0 \\ P_i^{(1)}\Omega_\infty^* P_i^{(3)\intercal} &= 0 \\ P_i^{(2)}\Omega_\infty^* P_i^{(3)\intercal} &= 0 \end{aligned} \tag{2.32}$$

This closed-form solution can be used as an initialization for subsequent refinement procedures, either through additional constraints [7] or by means of bundle adjustment such as in [37]. The constraints established in Equation (2.32) require at least three independent image views[4] via the camera matrices $\mathbf{P}_i$. With an estimate for Ω_∞^*, we can calculate the transform to upgrade the projective calibration to a metric one. From Equation (2.15) we take $\hat{\mathbf{P}}_i = \mathbf{P}_i\mathbf{H}$ where $\hat{\mathbf{P}}_i$ is the projective camera matrix at index i. Let us split $\hat{\mathbf{P}}$ into $\mathbf{M}_{4\times3}$ and $\mathbf{T}_{4\times1}$, coupled with Equation (2.28) such that:

$$\begin{bmatrix}\mathbf{M} & \mathbf{T}\end{bmatrix} = \mathbf{P}\begin{bmatrix}\mathbf{B} & \mathbf{c}\end{bmatrix} \tag{2.33}$$

Here $\Omega_\infty^* = \mathbf{B}\mathbf{B}^\intercal$, where $\mathbf{B}$ can be attained using rank-3 factorization. For $\mathbf{T}^i = \begin{bmatrix}T_x^i & T_y^i & T_z^i\end{bmatrix}$ we have

$$\begin{aligned} T_x^i &= P_x^i\mathbf{c} \\ T_y^i &= P_y^i\mathbf{c} \\ T_z^i &= P_z^i\mathbf{c} \end{aligned} \tag{2.34}$$

resulting in the following relationships to the image correspondences[5]:

$$\frac{T_x^i}{T_z^i} = \frac{\sum_{j=1}^n w_j^i u_j^i}{\sum_{j=1}^n u_j^i} \quad \text{and} \quad \frac{T_y^i}{T_z^i} = \frac{\sum_{j=1}^n w_j^i v_j^i}{\sum_{j=1}^n u_j^i} \tag{2.35}$$

These produce two equations per camera view from which to solve for $\mathbf{c}$. We now have the necessary transform $\mathbf{H}$ to upgrade the projective calibration to metric space according to Equation (2.15).

2.9 Bundle Adjustment

Bundle Adjustment[1, 16, 19, 39] refers to a class of non-linear optimization problems that jointly refine the structure and motion of an image sequence with some number of optical devices, using reprojection error as the primary penalty in the loss function. Modern syntheses leverage the sparsity inherent of these optimization problems to improve performance, as the process can be computationally and memory intensive, particularly at large scales. The naive Levenberg-Marquardt algorithm has a computational complexity of $O((m^3+mn)$ where m represents the number of cameras, n the number of 3D points[12],

[4] Since Ω_∞^* is a rank-deficient symmetric matrix parameterized by 10 unknowns

[5] The derivation for this has been skipped for brevity, the full extent of which can be found in [37]

which can be reduced to $O(m)$ by exploiting the sparsity in primary and secondary matrix structures, and other techniques discussed in [1, 6, 16]. This approach is statistically optimal, provided that the noise function of detected correspondences are Gaussian distributed.

The naive optimization approach uses a loss function derived from the forward model Equation (2.1) similar to the following[39]

$$\arg_{\widehat{\mathbf{P}}}\min \sum_{i} \sum_{j} \rho_j \left(\left|\left| \widehat{\mathbf{P}}_i \widehat{\mathbf{X}}_j - \mathbf{x}_j \right|\right|^2 \right) \tag{2.36}$$

where the camera motion $\mathbf{P}$ is adjusted (either in metric space via $\mathbf{R}$ and $\mathbf{t}$ or directly through projective bundle adjustment) and all structure points $\mathbf{X}_j$ are triangulated from these new camera positions, from which the reprojection error can be calculated.

Chapter 3

Problem Formulation

In this chapter, our aim is to formalize the problem and scope of projector autocalibration. We begin by defining the key components and variables involved in the calibration process, followed by an elaboration of the challenges and assumptions. Finally, we present the mathematical formulation of the problem and discuss the desired outcomes of the autocalibration process.

The task of projector calibration has many components and configurations and can become complex with respect to setup time, hardware, and computational resources. The possible configurations include single or multi-projector systems, photometric or geometric calibration, machine vision or mobile phone cameras, single or multiple cameras, and cameras with fixed or varying intrinsics, among others. The object of this thesis is to determine a practical autocalibration algorithm for projector systems. As a result, the scope of the experiments will be limited to the following: the geometric calibration of a single projector using a single uncalibrated camera with fixed intrinsics, to be calibrated without a checkerboard or calibration artifact and without any priors.

The system will consist of a projector, a camera, and a scene including a piece-wise flat, discontinuous surface. The projector emits structured light patterns onto the scene, which are then captured by the camera from various positions. The parameters that configure this scene are given by

- $\mathbf{K}_P, \mathbf{R}_P, \mathbf{t}_P$: the projector's intrinsic and extrinsic parameters.
- $\mathbf{K}_C, \mathbf{R}_C, \mathbf{t}_C$: the camera's intrinsic and extrinsic parameters.
- $\mathcal{I}$: the set of captured images of the structured light patterns.

where the raw image data in $\mathcal{I}$ is decoded to produce the set of all correspondences $\mathcal{U}$ from which the structure $\mathcal{X}$ along with projector and camera parameters are computed.

3.1 Challenges and Assumptions

The process of projector calibration faces several technical and practical challenges. Some of the main challenges include:

- **Degenerate configurations:** Certain camera motions, such as purely rotational motion, or planar scenes leading to ambiguities in the solution space.
- **Surface properties:** Non-lambertian surface reflection resulting in signal duplication and occlusion via over-blooming.
- **Computational and memory cost:** Large structure-from-motion type problems are very computationally intensive.

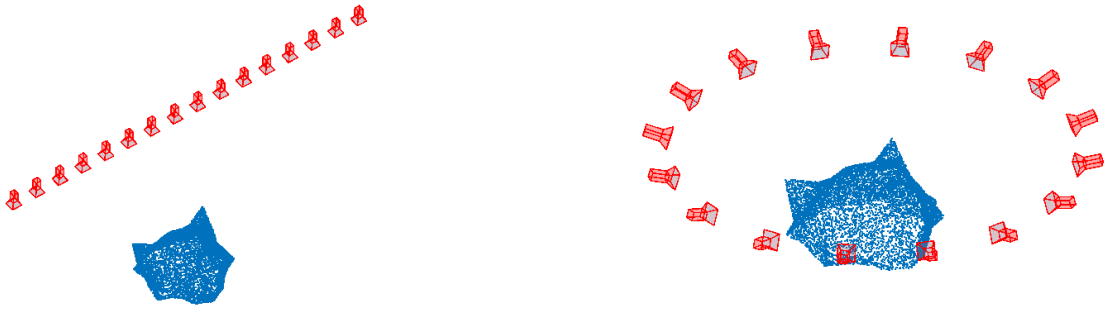
In order to develop a practical autocalibration algorithm for projector systems, we make the following assumptions:

- **General Camera Properties:** We assume that the camera used in these experiments has a unit aspect ratio, zero skew and an image-centered principal point.
- **Constant Camera Properties:** We further assume that for all images in the gathered sequence, the camera and projector internal parameters remain constant.

Note that we make no assumption regarding the projection surface (e.g. properties such as smoothness and surface continuity) that would otherwise compress the solution space.

Key Challenges

The key challenges posed by the projector autocalibration problem are broken down here. We discuss the pitfalls of degenerate camera motions resulting in ambiguities in the solution space, the high computational cost for high correspondence counts and number of views, and the non-lambertian surface properties which affects correspondence noise through signal duplication as well as blooming and occlusion. These challenges are discussed along with ways to mitigate these effects on the autocalibration process.



(a) (b)

Figure 3.1: Examples of degenerate camera motions. These motion sequences, among others, are degenerate in that they do not produce sufficient constraints to recover metric calibration parameters. For a more thorough treatment on these, see [27].

Degenerate Configurations

The process of autocalibration relies on rigidity constraints to infer the necessary transformation to upgrade projective structure and motion to a metric one. As such, the camera views and the correspondences used in this process must be independent and provide sufficient constraints. Degenerate camera motions such as pure translational motion, purely rotational motion, rotation about the optical axis or perfectly orbital motion incur ambiguities that necessarily inhibit good calibration results.

Surface Properties

The surface properties of a scene to be used for projection mapping have an outsized impact on the quality of the resulting calibration. The ideal scenario for any projector calibration and scene reconstruction problem would be that all surfaces within the scene exhibit Lambertian reflectance. A perfectly diffuse reflection of the structured light used for correspondence generation would eliminate signal duplication and result in even lighting and a moderate contrast range. The lack of lambertian surfaces in a scene can result in a series of challenges:

- **Signal duplication.** This can result in non-gaussian noise in the detected point correspondences. This impacts the performance of the refinement process by Bundle Adjustment, as BA is optimal only under conditions of Gaussian noise.
- **Specular reflections**. These can create bright spots, making it difficult to find a single optimal exposure setting for correspondence gather. Poor exposure settings (exposure time and gain values) can exacerbate blooming, which can occlude critical correspondence features and make thresholding for image binarization more difficult. Solving this via high-dynamic range (HDR) significantly diminishes camera mobility, as multiple frames are required for an individual HDR image.

Computational and Memory Complexity

The dimensionality of calibration and structure-from-motion problems increases both with the number of views and the number of correspondences. The iterative nature of these types of algorithms make them quite susceptible to computational and memory complexity. The upper bound on the number of correspondences per camera view is the total number of correspondences encoded in the structured light.

Considering a projector with a resolution of $M \times N$ pixels, the maximum number of correspondences encoded in Grey Code structured light is equal to the total number of pixels in the projector's image plane, which can be represented as:

$$C_{\max} = M \times N. \tag{3.1}$$

As the number of camera views, denoted by V, increases, the overall complexity of the problem grows. The total number of correspondences, C_{total}, across all camera views can be written as:

$$C_{\text{total}} = V \times C_{\max}. \tag{3.2}$$

To quantify this, let us consider the number of edges, E, in the graph as a function of the number of camera views and correspondences. In the worst-case scenario, with fully connected graph, the number of edges can be represented as:

$$E = \frac{V \times (V-1)}{2} + V \times C_{\max}, \tag{3.3}$$

where $C_{\max}$ denotes the maximum number of correspondences per camera view. This equation indicates that the number of edges increases quadratically with the number of camera

views, resulting in a substantial growth in computational complexity. This increase in complexity results in a larger optimization problem that must be solved iteratively, further exacerbating the computational and memory demands. This has resulted in a strong research interest in the direction of efficient optimization problems that leverage sparsity in SfM problems[16, 19], hierarchical approaches through divide-and-conquer methods[4] and sparse graph representations to reduce graph connectivity[34].

3.2 Experiment 1: The Dual Absolute Quadric and Bougnoux Synthetic Data Evaluations

To effectively assess the relative performance between the DAQ and Bougnoux-based methods of autocalibration, it is helpful to test these methods using synthetic data. In experiments with real hardware, some variables are difficult to control and it can be impractical to parameterize an experiment along these axes, especially when ground truth is difficult to attain. These evaluations are discussed and analyzed in Chapter 4.

Bougnoux: Optical Axis Proximity

Recall that the work in [23] discovered that the Bougnoux equations are degenerate in the case of intersecting optical axes. In practice, this perfect intersection is impossible to achieve, yet close proximity of these axes for two cameras observing the same scene is rather expected. The behaviour near these degenerate points is therefore important to understand, as it impacts the viability of methods based on these equations through its impact on robustness.

Take a two camera configuration, a static camera C and another camera C' to pivot about its pitch axis by θ. The experiment is parameterized by the angle θ_i and correspondence noise standard deviation σ_j. For each stereo pair, the Bougnoux equations are used to estimate each camera's focal length. The experimental configuration can be seen in Figure 3.2. Each iteration will take the following form:

1. Generate a synthetic scene with known ground truth geometry and camera parameters.

2. Introduce positional noise by perturbing the ground truth extrinsic camera parameters with zero-mean Gaussian noise.

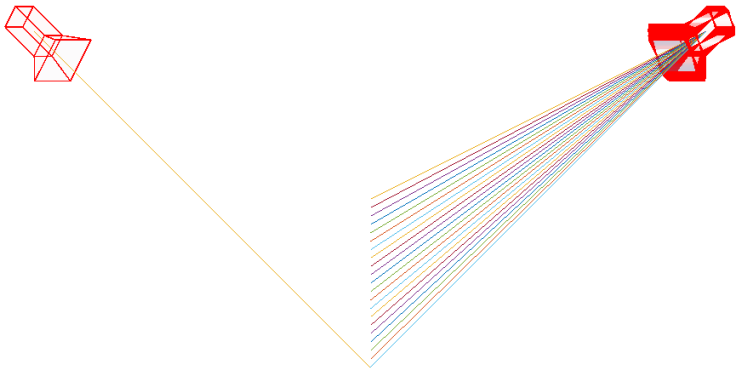


Figure 3.2: Superposition of all camera poses used in Experiment 1 for the Bougnoux evaluation (with principal rays shown). The left camera is kept fixed, while the right camera has a variable pitch. The synthetic data tests are performed in two-view configurations that include the fixed camera on the left combined with one of the views on the right to parameterize the experiment by minimum distance between principal rays.

3. Solve for the camera focal lengths using the Bougnoux equations.

The above experiment results in 1680 unique combinations of σ_{uv} and θ_i, each of which is evaluated $40\times$ to produce a reasonable statistical view of performance.

DAQ: Degenerate Motion Sequences

The degenerate motion sequences (partially illustrated in Figure 3.1 are difficult to achieve by accident. They require a physical setup such as a turntable or a linear camera slider. However, just as in the Bougnoux equations, understanding the behaviour near the degenerate points is eminently desirable in order to develop a robust set of heuristics for projector autocalibration via these methods.

In the DAQ evaluation, we generate a series of virtual scenes with varying levels of correspondence noise, denoted by σ_{corr}, positional noise, σ_{pos} and rotational noise, σ_{rot}. The rotational noise function is represented as perturbations in Euler angles. . The experiments involve the following steps:

1. Generate a synthetic scene with known ground truth geometry and camera parameters.

2. Introduce correspondence noise by perturbing the ground truth correspondences with zero-mean Gaussian noise.

3. Introduce positional noise by perturbing the ground truth extrinsic camera parameters with zero-mean Gaussian noise. The perturbations are applied independently to the translational and rotational components.

4. Perform DAQ autocalibration on the synthetic data.

By systematically varying the levels of correspondence noise and positional noise in the synthetic data, we can analyze the robustness and accuracy of the DAQ autocalibration method for different levels of noise. The results of these experiments can then be used to guide the development of heuristics and improvements to the method, ensuring its robustness and reliability in real-world projector autocalibration scenarios.

DAQ: Minimally Constrained Sequences

The degenerate motion sequences described so far are difficult to achieve without planning. To simulate a more realistic projector calibration configuration, a randomly generated set of camera poses about a point cloud is used and tested under minimally constrained conditions.

For each iteration of this test, three images are used for autocalibration, with increasing levels of correspondence noise.

- Generate a synthetic scene with known ground truth geometry and camera parameters
- Generate a set of random camera poses about the pointcloud with a common focus point for all cameras
- For all possible combinations of three views in the set of camera poses, project the pointcloud to the image plane and apply noise to correspondences
- Perform autocalibration using the DAQ and evaluate against ground truth values

Hypothesis and Expected Outcomes

For the Bougnoux equations, the expectation for this experiment is a poor convergence — even for low values of correspondence noise — near the degenerate configuration that converges on the correct answer as the distance to that configuration increases.

The DAQ method is expected to be more robust to noise, without being impacted by principal ray proximity. It is also expected to behave similarly to the Bougnoux equations in near-degenerate configurations with a similar convergence as the positional noise is increased.

3.3 Experiment 2: Projector Autocalibration via the Dual Absolute Quadric

In this experiment, the DAQ autocalibration method is tested under practical conditions. Drawing upon the insights taken from the synthetic data experiment detailed in Section 3.2, we seek to interpret the findings of this real-world experiment in a comparable context. This experiment is discussed in detail in Chapter 5.

Problem Definition

The crux of this experiment is to undertake the task of autocalibrating a single projector using the DAQ method, under conditions of unknown correspondence noise, non-lambertian surface properties and a densely-connected scene graph. The efficacy of this method will be evaluated by comparing the resulting intrinsic parameter estimates.

This experiment aims to establish the DAQ as an accurate and robust autocalibration method when presented with the practical challenges of projector calibration. While this experiment is limited to a single projector-camera pair, the resource complexity is still significant given the pixel-dense correspondences and multiple camera views processed.

In addition to the DAQ, we will also test the GDPCA method[**?**] of projector autocalibration, juxtaposing its performance with the DAQ against the aforementioned challenges. Since GDPCA uses the Bougnoux equations for initialization, it is expected that the performance between these methods will mirror the results of Section 3.2.

DAQ Projector Autocalibration

The dataset for this experiment is gathered using a single camera and projector setup. By strategically repositioning the camera, a total of 8 views are captured. This dataset then undergoes processing using the DAQ autocalibration method. We have made a conscious effort to avoid purely translational camera poses to steer clear of degenerate configurations.

To draw a comparative analysis, the dataset is also processed using GDPCA[**?**] for each projector-camera stereo pair. The estimated intrinsic parameters derived from this method will be compared against the results obtained from the DAQ.

Hypothesis and Expected Outcomes

It is hypothesized that the DAQ method will demonstrate good convergence in intrinsic parameter estimation across variable camera view counts. In contrast, the GDPCA method is expected to show a significant lack of convergence across all projector-camera pairs. Through this comparison, we aim to demonstrate the relative strengths and weaknesses of these two methods for the task of projector autocalibration.

Chapter 4

Synthetic Data Evaluation

4.1 Motivation

The motivation to conduct a synthetic data experiment for projector autocalibration, comparing both the DAQ and Bougnoux methods, primarily stems from the need for a controlled environment to understand the performance and limitations of these approaches. Synthetic data provide precision and the ability to manipulate individual parameters, allowing us to isolate specific factors of interest, such as noise levels and degenerate camera motions, among others. Moreover, synthetic data serve as a crucial reference point, providing a ground truth against which the performance of the autocalibration methods can be evaluated. By leveraging synthetic data, we can meticulously design scenarios that test these methods under varying conditions for robustness and accuracy, thus revealing their strengths and weaknesses in a controlled, reproducible manner. The insights taken from such an experiment can guide the development of robust heuristics, as well as the fine-tuning of these methods for real-world, practical applications.

4.2 Methodology

4.2.1 Bougnoux: Experimental Setup and Parameterization

Recall from Section 2.7.1 the Bougnoux equations for autocalibration are given by:

$$\begin{aligned} f_1^2 &= -\frac{p_2^\top [e_2]_\times \tilde{\mathbf{I}}\mathbf{F}^\top p_1 p_1^\top \mathbf{F} p_2}{p_2 [e_2]_\times \tilde{\mathbf{I}}\mathbf{F}\tilde{\mathbf{I}}\mathbf{F}^\top p_2} \\ f_2^2 &= -\frac{p_1^\top [e_1]_\times \tilde{\mathbf{I}}\mathbf{F}^\top p_2 p_2^\top \mathbf{F} p_1}{p_1 [e_1]_\times \tilde{\mathbf{I}}\mathbf{F}\tilde{\mathbf{I}}\mathbf{F}^\top p_1} \end{aligned} \tag{4.1}$$

The above equations may, for some combinations of principal points and fundamental matrix estimates, return complex numbers or fail to produce a the correct estimate. This has been seen in applications of the GDPCA method [**?**]. Further, it is expected to show degenerate behaviour in cases where the optical axes of the camera views are intersecting [23]. While perfect intersection is not practically achievable, it is important to understand the behaviour of these equations under conditions of near-intersection. In order to better understand this scenario, we evaluate this with synthetic data using a set of 3D coordinates from which to establish ground truth correspondences. Any arbitrary set of 3D correspondences may be used; for the purposes of this experiment 10,000 points were sampled from a 3D model similar to that shown in Figure 4.3.

Consider a synthetic two camera setup, as depicted in Figure 4.1, parameterized as follows: the minimum distance between the optical axes (on an interval of $[0, 0.2)$ in 21 equal increments) and zero-mean noise applied to the ground truth point correspondences $\mathbf{x}$ to produce estimates $\hat{\mathbf{x}}$ such that $\hat{\mathbf{x}} \sim \mathcal{N}(0, \sigma_X)$. Each combination of parameters will be evaluated 40 times to provide a reasonable sample.

4.2.2 DAQ: Experimental Setup and Parameterization

The DAQ is evaluated in two parts. First, the near-degenerate behaviour of the DAQ is evaluated for one of the critical motion sequences shown in Figure 3.1. Second, it is then evaluated under conditions of minimal constraints and in geometric configurations where the Bougnoux equations fail.

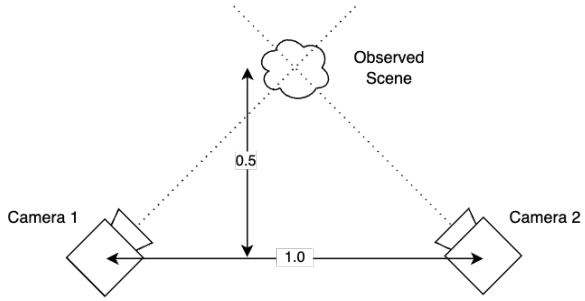


Figure 4.1: Diagram of a synthetic two-view setup for the evaluation of the Bougnoux equations

Degenerate Motion Sequences

The DAQ is evaluated similarly to the Bougnoux equations in the previous section. The near-degenerate behaviour is tested by generating a set of synthetic camera poses in a known degenerate motion sequence, applying noise along multiple parameters and assessing the calibration results against ground truth as a function of the applied parameter noise. The experiment is parameterized by Gaussian noise in the following variables: σ_{uv} for the 2D correspondences $\mathbf{x}$, σ_C the camera position t_c in $\mathbb{R}^3$, and σ_γ for the Euler angles representing the camera orientation.

- Generate a set $\mathcal{T}$ of camera positions t_c, where $\mathcal{T}$ is of size m. These camera positions are placed at regular intervals on a circle of radius r, centered about a pointcloud $\mathcal{X}$, similar to Figure 3.1b.
- Apply noise to all t_c in $\mathcal{T}$ with a distribution of $\mathcal{N}(0, \sigma_C)$
- Determine camera orientation R by aligning the camera z axis with the centroid of $\mathcal{X}$ and aligning the camera image plane x-axis with the XY plane in $\mathbb{R}^3$.
- Apply a rotation by randomly sampling Euler angles distributed according to $\phi, \theta, \psi \sim \mathcal{N}(0, \sigma_\gamma)$.
- Project points of $\mathcal{X}$ onto each camera image plane, producing a set of image points $\mathcal{U}$ for each camera
- Apply a Gaussian noise function to all $\mathbf{x} \in \mathcal{U}$ such that $\mathbf{x} \sim \mathcal{N}(0, \sigma_{uv})$

The above procedure is parameterized on $\sigma_C, \sigma_\gamma, \sigma_{uv}$, such that:

$$\begin{aligned} \sigma_C &\in \{0, 0.05r, 0.1r, 0.2r\} \\ \sigma_\gamma &\in \{0, 0.1, 0.2, 0.4\} \\ \sigma_{uv} &\in \{0, 1, 2, 4\} \end{aligned} \tag{4.2}$$

where each combination of σ values in Equation (4.2) is tested for five (5) iterations, resulting in

Three View Autocalibration

To provide a more explicit contrast between the relative advantages and disadvantages of the discussed autocalibration methods, the DAQ is tested under conditions that are degenerate for the Bougnoux equations. In the previous experiment, the random Gaussian noise applied to position and orientation does not distinguish the DAQ from the Bougnoux equations. However, the DAQ is robust under conditions where the Bougnoux equations are not, which will be shown by the results of this experiment.

In this experiment, the DAQ method is tested using a set of m randomly generated camera poses viewing a common point in $\mathbb{R}^3$, such that all optical axes intersect. From that set, all combinations of three-view configurations are tested to determine the robustness of the DAQ with the minimum possible constraints. The set of camera poses $\mathcal{C}$ is generated as follows:

- Generate a set $\mathcal{C}$ of camera positions C, where $\mathcal{C}$ is of size m. The elements C are sampled randomly from a sphere of radius r where the sphere center coincides with the mean of all points $\mathbf{X}$ in the pointcloud.
- Apply a noise function to C such that $C \sim \mathcal{N}(0, \sigma_C)$
- Determine rotation for camera orientation by aligning the camera z axis with the centroid of $\mathcal{X}$ and aligning the camera image plane x-axis with the XY plane in $\mathbb{R}^3$.
- To preempt ambiguity through linearly dependent images and correspondences, apply noise to each camera's rotation θ about its own optical axis, such that $\theta \sim \mathcal{N}(0, \sigma_\theta)$

This produces a set of transforms $\mathcal{T}$ for each camera where each transform is represented by $\mathbf{T}_i \in SE(3)$. The synthetic cameras with poses from $\mathcal{T}$ are then used to produce correspondences by mapping the points of the pointcloud $\mathcal{X}$ to each camera's image plane. These are the correspondences that are then used to perform autocalibration with the DAQ.

(a) $\sigma_C = 0$, $\sigma_\gamma = 0.0$ (b) $\sigma_C = 0.05$, $\sigma_\gamma = 0.1$ (c) $\sigma_C = 0.1r$, $\sigma_\gamma = 0.2$

Figure 4.2: Examples of increasing positional and rotational variation for near-degenerate motion sequence test of DAQ.

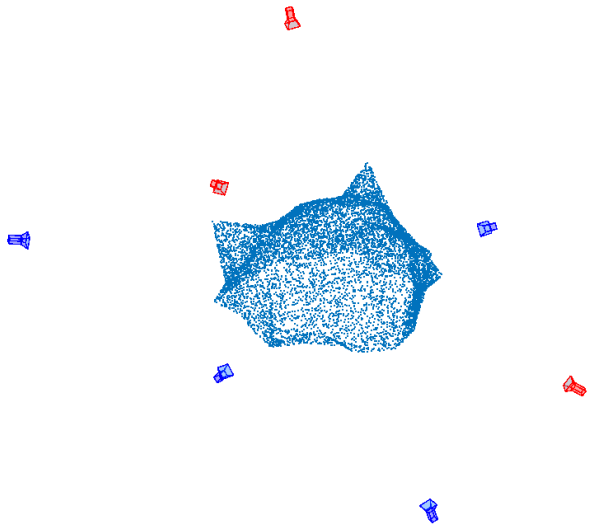


Figure 4.3: Setup for DAQ three-view test. Every combination of three cameras are tested for autocalibration using the DAQ ($n = 7$). In this example, cameras shown in red are selected for DAQ autocalibration. Note that for all these configurations, all camera optical axes intersect. The aim of this is to show the stability of the minimally-constrained problem (i.e., with three views).

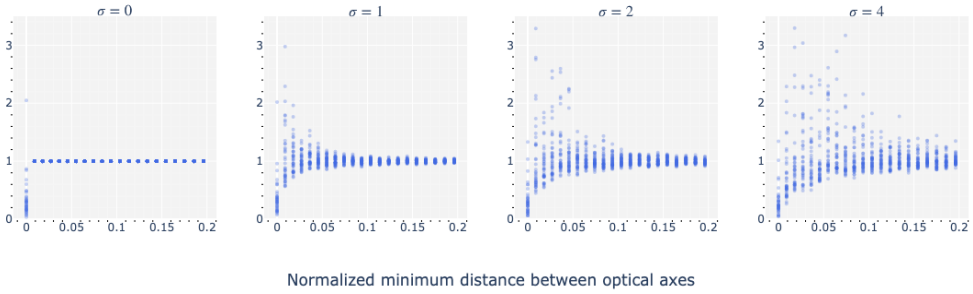


Figure 4.4: Results of Bougnoux equation evaluation for the camera focal length f. In all cases, the correct value is 1.0. At $x = 0.0$ we have the degenerate configuration of intersecting optical axes. In general, the focal length error spread increases with noise σ and with smaller baseline x. The minimum distance is normalized against the camera baseline (shown in Figure 4.1 as 1.0).

4.3 Experimental Results

The results of the tests evaluating the Bougnoux degeneracy are shown in Figure 4.4; the x-axis indicates the normalized minimum distance between the optical axes of the two cameras. It becomes obvious that with low levels of correspondence noise, the focal length estimates approach the ground truth value with increasing distance between the optical axes. However, as $\sigma_u v$ increases, this convergence deteriorates dramatically. This is the core attribute that makes the Bougnoux equations inherently unsuitable for robust autocalibration in [**?**, 40].

The results of the DAQ tests shown in Figure 4.5 and Figure 4.6 show a similar, albeit more promising pattern. Note that in Figure 4.6, even with $\sigma_u v = 4.0$ the estimated focal lengths are within 5% of the ground truth value. The results in Figure 4.5 indicate that varied poses, particularly rotational variance, is a significant contributor to the robustness of the DAQ.

The results of Section 4.2.2 provide a more direct comparison between the DAQ and Bougnoux methods, which can be seen in Figure 4.7. Note that the DAQ focal length estimates are for perfectly intersecting optical axes, while the Bougnoux equations are given for near-intersection, to better reflect real-world scenarios. It is expected, based on this data, that the DAQ will show significantly more robust estimates in the results of Chapter 5.

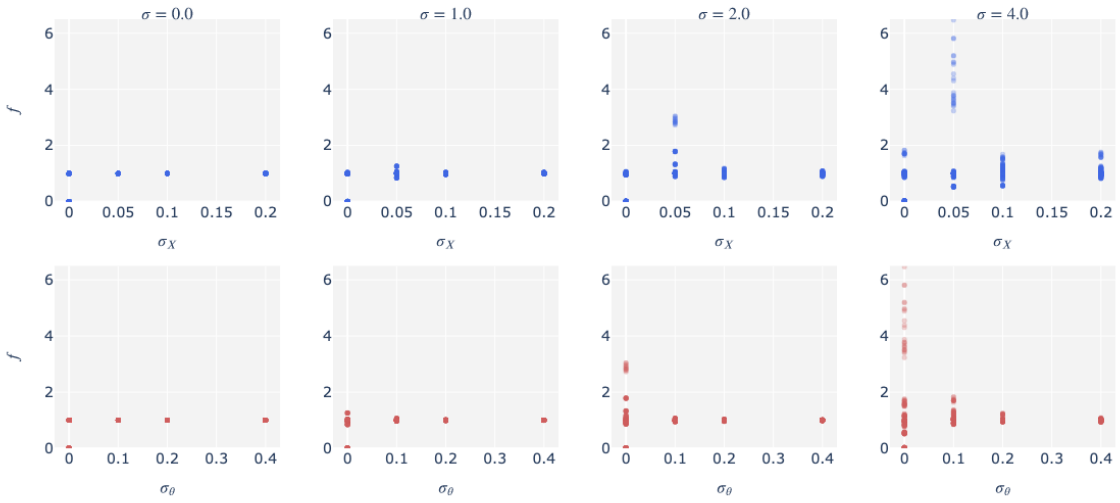


Figure 4.5: Results of DAQ focal length estimates for $\sigma_{uv} = 4$ by pose variation. Note the stronger correlation between rotational variation and improved focal length estimates. The results are normalized such that the ground truth value is $f = 1$.

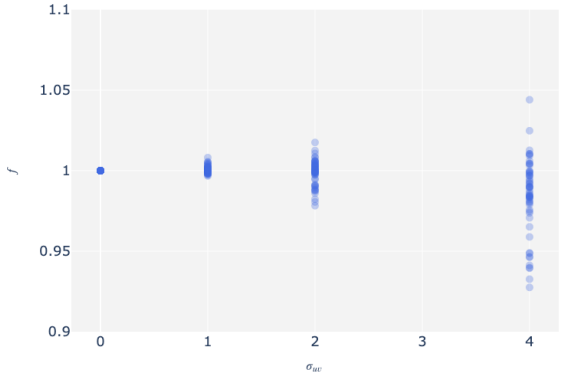


Figure 4.6: Focal length estimation results using the DAQ with maximum tested pose variation ($\sigma_X = 0.2$ and $\sigma_\theta = 0.4$)

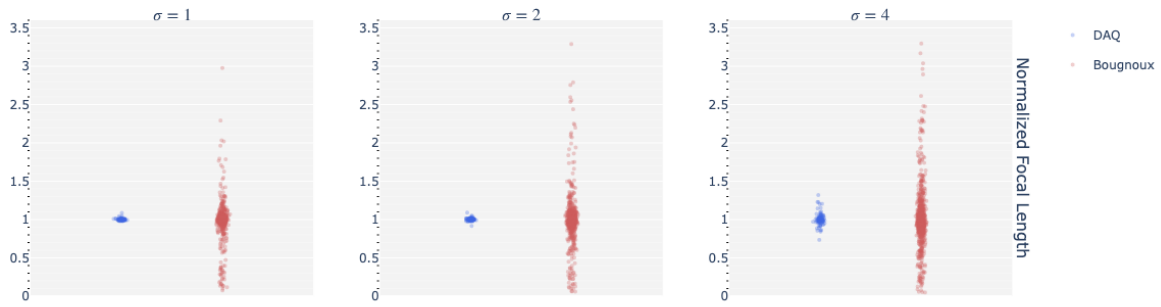


Figure 4.7: Comparison of DAQ and Bougnoux focal length estimate distribution evaluated with synthetic data.

The DAQ method appears preferable for sequences where motion can be randomized without sacrificing image sensor real-estate. Near-degenerate behaviour is similar to Bougnoux equations, however degenerate configurations are significantly more difficult to achieve. In particular, DAQ is not subject to poor calibration results at- or near-principal ray intersection. Overall the DAQ is better suited to the projector-autocalibration problem space.

Chapter 5

Projector Autocalibration

5.1 Motivation

The motivation of this chapter is to effectively evaluate the DAQ as a robust means of autocalibration for projector-camera systems. It will include all the challenges listed in Section 3.1.

5.2 Methodology

The DAQ autocalibration method described in Section 2.8 is evaluated using a single projector-camera pair, to restrict the problem space. Several assumptions will be made regarding these devices, namely that the focal lengths f_x and f_y are equal, as well as an image-centered principal point (for the camera only). These assumptions will be used to develop the set of linear equations used to estimate the DAQ Ω^*_∞ that encodes the transformation upgrading the projective reconstruction to a metric one. To that end, we will capture correspondences from several viewpoints that include sufficient pose and angular variation to sufficiently constrain this set of equations. In addition, we test the GPDCA method of [**?**] on the same dataset to show the relative robustness of the DAQ.

5.2.1 Correspondence Gather

The set of correspondences $\mathcal{U}$ is built using Gray Code structured light (depicted in **??**) by placing the camera at a position with a view of the scene and capturing all 21 SL

frames. This gather process is repeated for a total of 8 camera poses. Once all SL frames are captured, each collection is decoded and mapped to the projector space. Algorithm 1 produces a correspondence table that provides the coordinates of each detected feature in each view, from which the projective calibration can be obtained as described in Section 2.6.

5.2.2 Projector Calibration

To begin the projector calibration process, the single observing camera is calibrated first following the DAQ calibration procedure in Section 2.8, using standard assumptions of $f_x = f_y$, $s = 0$, $u = \frac{\text{width}}{2}$ and $v = \frac{\text{height}}{2}$. The calibration is performed using correspondences from all camera views to build a linear system of equations using the constraints described in Equation (2.32). From this, Ω_∞^* is estimated through linear least squares using singular value decomposition (SVD) and subsequently used to determine the linear transform matrix $\mathbf{H}$ (see Equation (2.28)). The transform matrix $\mathbf{H}$ upgrades the initial projective calibration to a metric one, from which the desired projector calibration parameters can be derived.

Recall the computational and memory complexity challenges discussed in Section 3.1. The dense correspondences generated by the Gray Code Structured Light result in a total correspondence count with an upper bound $C_{\max}$, equal to the total projector resolution. The auto-calibration algorithm using the DAQ first proposed in [37] has a memory complexity on the order of $O(n^2)$, where n is the total number of unique points in $\mathbb{R}^3$. In the worst case, this results in a memory footprint of $5.3 \times 10^{12}k$, where k is some constant[1]. That being said, all available correspondences are not required to solve for Ω_∞^*. In this case, the correspondences are sub-sampled at a constant interval (5% of total correspondences) to reduce the memory and computational requirements.

Having solved for the transform $\mathbf{H}$ to upgrade projective structure $\hat{\mathcal{X}}$ and motion $\hat{\mathcal{P}}$ to metric space, we can now map correspondences in the projector domain to points in the acquired metric structure. The forward model in Equation (2.1) can be estimated via the direct linear transform (DLT) algorithm[12]. The inverse problem takes the form of $\mathbf{A}z = \mathbf{0}$, where

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & w_i\mathbf{X}_i & y_i\mathbf{X}_i \\ w_i\mathbf{X}_i & \mathbf{0} & x_i\mathbf{X}_i \end{pmatrix}_{2n\times 12} \tag{5.1}$$

and z is a 12-element vector representing the camera projection matrix $\mathbf{P}_{\text{projector}}$. From Equation (5.1) we have 2 equations per $\mathbb{R}^3 \to \mathbb{R}^2$ mapping, and as such we need a minimum

[1]This constant is a function of several factors, including the hardware on which the algorithm is executed.

of 6 points to fully constrain the estimate. Naturally, given the pixel-level density of Gray Coded SL we used in this method, $\mathbf{A}$ is significantly over-constrained, and we can sub-sample point correspondences to reduce computational time and memory footprint. Once $\mathbf{A}$ is defined, we can estimate the null space z by eigendecomposition of $A^\intercal A$, which yields $\mathbf{P}_{\text{projector}}$.

5.3 Experimental Results

Setup and Data Collection

We evaluate the performance of our proposed method for the camera using the results of Zhang's method [42] as a ground-truth for comparison. Ground-truth values for the projector calibration parameters were not recoverable by some external means, and hence the projector calibration parameters derived from the proposed method are compared to those recovered using Zhang's method camera calibration parameters. We also evaluate the GDPCA method in [**?**] as a reference for existing projector autocalibration methods. The experimental results are summarized in Figure 5.2. Since the GDPCA method is only designed to calibrate a single projector-camera pair, the evaluation is undertaken by testing it on every combination of projector-camera pairs afforded by our dataset.

The setup follows the example in Figure 5.1, and the equipment used in this experiment consisted of the following items, shown in Figure 5.3(a):

- One 1920x1200 Christie Digital Systems projector
- One 5MP FLIR GigE machine vision camera equipped with a 5mm lens
- One 3D printed model[2]

The performance of the proposed method is evaluated for the camera using the results of Zhang's method [42] as a ground-truth for comparison. The projector and camera are set up to view the Wolf Head model, and correspondences are captured using 21 binary Gray Code frames (11 horizontal, 10 vertical). The camera is repositioned and correspondences gathered for a total of 8 unique views. It is expected that multiple camera pixels in each camera view will map to the same feature in the projector domain, therefore we group all correspondences mapping to the same projector pixel and take the centroid of the coordinates in the camera domain.

[2]The model is the same as the one seen in Figure 4.3

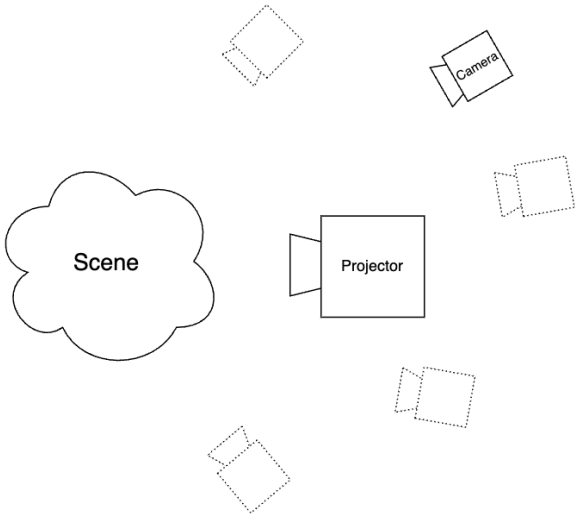


Figure 5.1: A mock up of the experimental setup including an arbitrary scene, a projector, and camera. The camera is moved about to capture the projected structured light from several views.

Results Analysis

The DAQ calibration results for the camera and projector can be seen in Table 5.1. Using Zhang's method results as a baseline, the DAQ shows promising results for the camera parameters, indicated by the $< 5\%$ error for both f_x and f_y. While the projector calibration estimates show even better performance, there is a significant disparity in the coordinates of the principal point (u, v). Specifically, the v coordinate of the projector principal point having an almost 25% discrepancy is cause for further investigation. However, the calibration results are none-the-less quite promising, and could be refined[3] to produce results on-par with pre-calibrated cameras.

When comparing the autocalibration methods, the results produced by GDPCA in Figure 5.2 show the same lack of robustness as seen in the initial evaluation of the Bougnoux method in Figure 4.4.[4] In comparison, autocalibration using the DAQ shows significantly better robustness against correspondence noise, even under minimally constrained config-

[3]Potential methods to refine the autocalibration results include the relaxed optimization method proposed in [7], that enforces the symmetry and positive-semidefiniteness of Ω^*_∞

[4]Note that the method in [40] uses the same underlying autocalibration method as GDPCA to build structure and motion incrementally, and therefore subject to the same degenerate configurations.

Table 5.1: Results for the proposed autocalibration method. Results for Zhang's method are included for reference. The error indicated is the relative difference between the proposed method and the parameters recovered using Zhang's method.

Source	**Method**	f_x	f_y	u	v
Camera	Zhang	2,410.43	2,410.71	1,232.00	1,045.56
	Proposed	2,307.30	2,303.20	1,261.30	1,041.50
	Error %	-4.28%	-4.46%	2.38%	-0.39%
Projector	Zhang	2,410.20	2,420.00	949.35	886.45
	Proposed	2,410.80	2,410.00	945.91	667.37
	Error %	-0.02%	0.41%	0.36%	24.71%

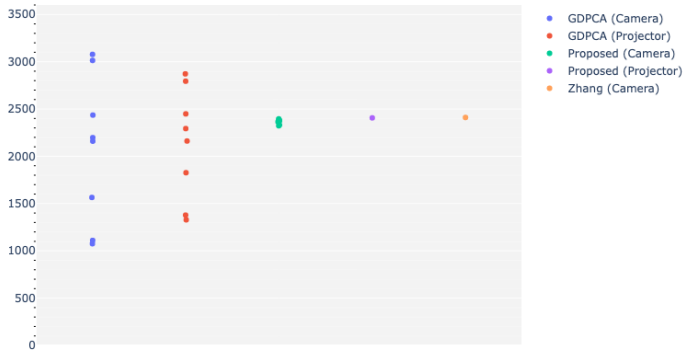
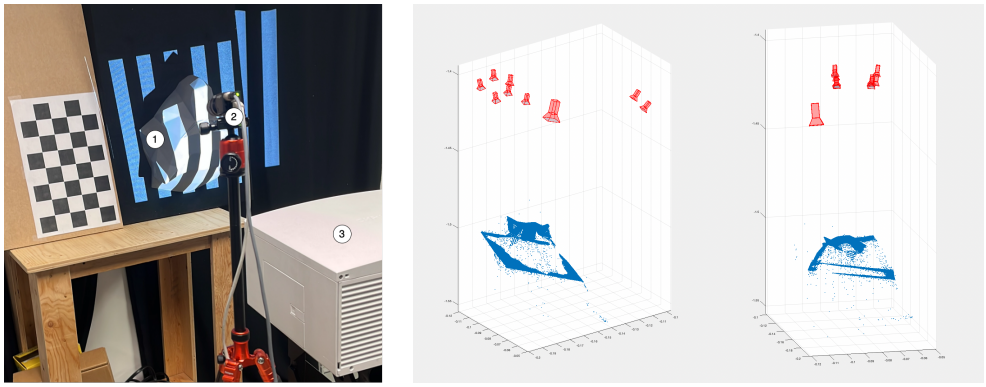


Figure 5.2: Distribution of focal length estimates of GDPCA versus the proposed method. The results for GDPCA evaluate all projector-camera pairs in the real dataset.

urations (i.e., three views). The structure and motion recovered via the DAQ can be seen in Figure 5.3. Note in particular that the camera positions shown in Figure 5.3 can be split into two sets of co-planar poses. This is a particularly positive piece of qualitative feedback, as the camera was moved about on a tripod at two different fixed heights, matching the recovered motion. For further qualitative review, the reconstructed pointcloud is shown in Figure 5.4.



(a) Experimental setup. (b) Estimated structure and motion.

Figure 5.3: The experimental setup (left) and the derived structure and motion (right). The experimental setup includes the following: (1) 3D model, (2) machine vision camera, (3) Christie projector. The large camera shown in the derived results represents the recovered projector pose; smaller cameras show the 8 estimated camera poses used to recover all intrinsic and extrinsic parameters.

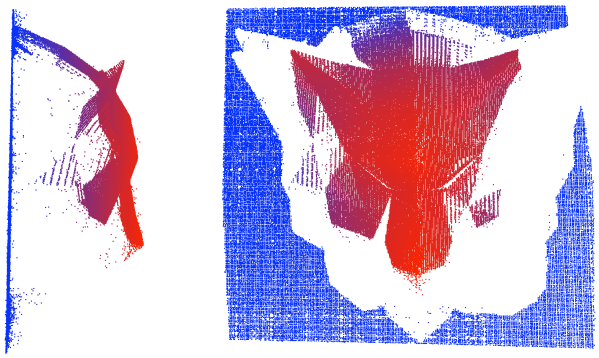


Figure 5.4: Colorized point cloud of the recovered structure. Color scale is based on depth.

Chapter 6

Conclusion

6.1 Summary of Findings

The synthetic data experiments of Chapter 4 corroborate the findings of the practical experiments of Chapter 5. Both methods have inherent degenerate behaviours that impact their reliability, however those of the Bougnoux method are much more likely to be encountered in practical settings for projector autocalibration. There are greater obstacles to encountering the degenerate configurations of the DAQ, as the motion sequences shown in Figure 3.1 require significant setup to achieve and can easily be avoided in projector calibration settings.

The findings showed promising results with less than a 5% discrepancy in camera parameters compared to a baseline established by Zhang's method. However, the projector calibration showed a significant difference in the coordinates of the principal point, with a discrepancy of almost 25% for the vertical coordinate. Despite these issues, the overall results of the study were encouraging, suggesting that with further refinement (such as optimizing Ω_∞^* for symmetry, rank deficiency, and positive-definiteness) this autocalibration method could yield results comparable to an approach using precalibrated cameras.

6.2 Contributions and Impact

The application and study of the DAQ autocalibration method extended to projector calibration is the core contribution of this research. The DAQ autocalibration method

provides a global autocalibration method using an unknown camera that is robust in the presence of correspondence noise, distinguishing it from existing methods. Additionally, probable camera poses for projector calibation have little overlap with Ω_∞^* degenerate motion sequences and can be easily avoided. Mitigating the lack of robustness currently found in similar methods simplifies projector calibration and enables an easier and more reliable autocalibration process.

6.3 Limitations and Future Work

Despite the promising calibration results of this study, some limitations remain. The principal point coordinates of the estimated projector calibration parameters deviated significantly from those produced by the Zhang calibration parameters. Additionally, the current DAQ method does not estimate the lens distortion parameters. Furthermore, the bundle adjustment refinement procedure as currently formulated is not optimal for the non-Gaussian noise distribution of the detected point correspondences of structured light.

There are several promising directions for future work in this area of research. Reducing the resource requirements by applying sparse graph representations[16, 19], divide-and-conquer methods[4] and sparse graph representations[34] would enable the auto-calibration of larger projector-camera systems by extending the single camera gather to greater view counts without reaching hardware memory limits and impractically long computation times.

The method could be further improved by reducing the correspondence gathering time by leveraging single-shot structured light patterns. This would enable a mobile camera for correspondence collection, which would significantly improve the flexibility and speed of the calibration process.

Finally, incorporating additional non-linear optimization to enforce rank-deficiency in Ω_∞^* could improve the estimation of intrinsic parameters such as the method described in [7]. This approach could yield significant improvements and address current limitations such as the significant projector principal point error noted earlier.

References

[1] Sameer Agarwal, Noah Snavely, Steven M Seitz, and Richard Szeliski. Bundle Adjustment in the Large. *European Conference on Computer Vision*, pages 29–42, 2010.

[2] Pablo F. Alcantarilla, Jesús Nuevo, and Adrien Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference*, pages 1281–1298, 2013.

[3] Hafeez Anwar. Calibrating projector flexibly for a real-time active 3D scanning system. *Optik*, 158:1088–1094, 4 2018.

[4] Brojeshwar Bhowmick, Suvam Patra, Avishek Chatterjee, Venu Madhav Govindu, and Subhashis Banerjee. Divide and conquer: A hierarchical approach to large-scale structure-from-motion. *Computer Vision and Image Understanding*, 157:190–205, 4 2017.

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

[6] Martin Byröd and Kalle Åström. Conjugate gradient bundle adjustment. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6312 LNCS(PART 2):114–127, 2010.

[7] Manmohan Chandraker, Sameer Agarwal, Fredrik Kahl, David Nistér, and David Kriegman. Autocalibration via Rank-Constrained Estimation of the Absolute Quadric. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[8] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, I:886–893, 2005.

[9] Sina Farsangi. Single-Shot Direct Block Address Encoding for Learning Screen Geometry. *University of Waterloo*, 2021.

[10] Yuan Gao, Yuyun Zhao, Le Xie, and Guoyan Zheng. A Projector-Based Augmented Reality Navigation System for Computer-Assisted Surgery. *Sensors*, 21(9):2931, 4 2021.

[11] Bingbing Han, Shourui Yang, and Shengyong Chen. Reasearch on comparison and analysis of the projector calibration. *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 1:768–772, 2019.

[12] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[13] E. E. Hemayed. A survey of camera self-calibration. *Proceedings - IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2003*, pages 351–357, 2003.

[14] Imran Junejo and Hassan Foroosh. Dissecting the image of the absolute conic. *Proceedings - IEEE International Conference on Video and Signal Based Surveillance 2006, AVSS 2006*, 2006.

[15] Hiroshi Kawasaki, Hideaki Nakai, Hirohisa Baba, Ryusuke Sagawa, and Ryo Furukawa. Calibration Technique for Underwater Active Oneshot Scanning System with Static Pattern Projector and Multiple Cameras. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 302–310, 2017.

[16] Kurt Konolige. Sparse Sparse Bundle Adjustment. *British Machine Vision Conference*, 10:102.1, 2010.

[17] Shuangyan Lei and Song Zhang. Flexible 3-D shape measurement using projector defocusing. *Optics Letters*, 34(20):3080, 2009.

[18] Francis Li, Hicham Sekkati, Jason Deglint, Christian Scharfenberger, Mark Lamm, David Clausi, John Zelek, and Alexander Wong. Simultaneous Projector-Camera Self-Calibration for Three-Dimensional Reconstruction and Projection Mapping. *IEEE Transactions on Computational Imaging*, 3(1):74–83, 2017.

[19] Manolis I A Lourakis and Antonis A Argyros. SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software (TOMS)*, 36(1):2, 2009.

[20] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 11 2004.

[21] Yi Ma, Stefano Soatto, Jana Košecká, and S. Shankar Sastry. Image Formation. In *An Invitation to 3-D Vision: From Images to Geometric Models*, pages 44–74. 2004.

[22] Daniel Martinec and Tomáš Pajdla. Structure from many perspective images with occlusions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2351:355–369, 2002.

[23] G N Newsam, D Q Huynh, M J Brooks, and H.-P Pan. Recovering Unknown Focal Lengths in Self-Calibration: An Essentially Linear Algorithm and Degenerate Congurations. *Proc. ISPRS-Congress*, 31, 1996.

[24] Sotiris Nousias and Christos Bergeles. Large-Scale, Metric Structure from Motion for Unordered Light Fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[25] Takayuki Okatani and Koichiro Deguchi. Autocalibration of a projector-camera system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1845–1855, 12 2005.

[26] Athanasia Papanikolaou, Sony George, Aditya Suneel Sole, Giorgio Trumpy, and Malgorzata Kujawinska. Multimodal Measurement of Parchment Using a Custom 3D Scanner. *Archiving Conference*, 20(1):7–11, 6 2023.

[27] Marc Pollefeys. *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*. PhD thesis, KU Leuven, Leuven, 1999.

[28] Marc Pollefeys and Luc van Gool. Self-calibration from the absolute conic on the plane at infinity. *International Conference on Computer Analysis of Images and Patterns*, 1296:175–182, 1997.

[29] Andrew Raij and Marc Pollefeys. Auto-calibration of multi-projector display walls. *Proceedings - International Conference on Pattern Recognition*, 1:14–17, 2004.

[30] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2564–2571, 2011.

[31] Behzad Sajadi and Aditi Majumder. Autocalibrating tiled projectors on piecewise smooth vertically extruded surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 17(9):1209–1222, 2011.

[32] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, 8 2010.

[33] Joaquim Salvi, Jordi Pagès, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004.

[34] Noah Snavely, Steven M Seitz, and Richard Szeliski. Skeletal graphs for efficient structure from motion. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[35] Peter Sturm, Srikumar Ramalingam, Jean Philippe Tardif, Simone Gasparini, and João Barreto. Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(1-2):1–183, 2010.

[36] Peter Sturm and Bill Triggs. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *European Conference on Computer Vision*, pages 709–720, 1996.

[37] Tomas Svoboda. A Convenient Multicamera Self-Calibration for Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 14(4):407–422, 2005.

[38] Bill Triggs. Autocalibration and the absolute quadric. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 609–614, 1997.

[39] Bill Triggs, Philip Mclauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle Adjustment-A Modern Synthesis. In *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms*, Corfu, Greece, 1999. Springer Berlin Heidelberg.

[40] Simon Willi and Anselm Grundhofer. Robust Geometric Self-Calibration of Generic Multi-Projector Camera Systems. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2017.

[41] Wei Zhang, Weishi Li, Liandong Yu, Hui Luo, Huining Zhao, and Haojie Xia. Subpixel projector calibration method for fringe projection profilometry. *Optics Express*, 25(16):19158, 2017.

[42] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on pattern analysis and machine intelligence*, pages 1330–1334, 2000.