Formulating a Moving Camera Solution for Non-Planar Scene Estimation and Projector Calibration

Katherine Arnold — Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada
Mark Lamm — Christie Digital Systems Canada Inc., Kitchener, ON, Canada
Paul Fieguth — Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada
Email: {k2arnold, pfieguth}@uwaterloo.ca, mark.lamm@christiedigital.com

Abstract

This paper presents the problem formulation for the challenge of projector calibration and non-planar scene estimation when equipped with a moving camera for data capture. This problem formulation assumes no prior information of the scene, moving-camera or fixed projector. This formulation forms the foundation of future handheld moving camera calibration for non-planar scenes and projector calibration.

1 Introduction

As the hardware for data projectors improves, it is anticipated that projectors could become much more commonplace. Both simple at-home and complex performative projector systems rely on the same principles of calibration and some estimation or understanding of the surface being projected upon. In this paper a simple moving camera calibration is explored which can provide both a non-planar scene estimation strategy and projector calibration. Because single uncalibrated cameras (e.g., cell phones) are commonplace, a moving camera projector calibration provides the foundation for a handheld calibration strategy that has the potential to reduce the capital costs associated with cameras for projector calibration. Further, projector-camera scene estimation strategies provide the grounds for accessible 3D scanning.

Current projector calibration methods rely on planar calibration surfaces or targets [1–3], or can require systems of several cameras [4]. Defining the scene can serve as a fundamental component of camera calibration:

- Planar Calibration Targets for 3D locations of known patterns for which homogeneous coordinate systems can be used;
- Non-Planar Calibration Targets where explicit 3D locations are known but homogeneous coordinates may not be used;
- Planar Scenes where points are not explicitly known but systems of structured light can be used with the homogeneous coordinate system;
- Unknown Scenes for which neither known points nor homogeneous coordinates can be assumed.

Unknown scenes may also present other challenges such as occlusions, surface warping of structured light patterns, and do not provide any basis for which to construct geometric understanding of the scene. Overall, a system which allows for non-planar surface estimation and projector calibration without burdensome camera and target requirements would be highly desirable.

Both Single- and Many-Camera projector calibration strategies have been found to be effective. Defining the initial degree of camera calibration serves as a defining characteristic of projector calibration strategies:

- Precalibrated and Known Pose Cameras where the only unknowns are the projector parameters;
- Precalibrated Camera where camera pose and projector must be estimated;
- Unknown Camera where all camera parameters must be found as well as the projector parameters.

Methods providing their own camera calibration gain advantages in environmental flexibility and adaptability not available in the disruption-sensitive and possibly fixed-pose precalibrated cameras. Previous estimation strategies that rely on single-camera and projector pairs [5, 6] treat the projector image plane the same as a captured camera image plane of measurement signals.

Having calibrated the camera across multiple viewpoints, camera measurement is then able to construct a scene understanding. While not explored in this formulation, the foundation of a moving camera allows for the incorporation of additional camera perspectives, potentially overcoming common size and occlusion challenges faced in scene observation by stationary cameras with augmented scene understanding.

This paper presents a clear formulation for a simple moving-camera and projector calibration and surface estimation strategy for non-planar scenes which relies on no parameter priors for devices or scene. The developed formulation provides the building blocks and flexibility to allow for the future incorporation of additional or moving cameras.

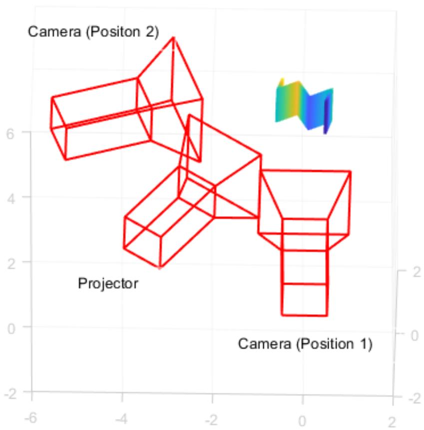

Fig. 1: Moving-Camera and Projector for Non-Planar Scene

2 Problem Formulation

The geometric relation between image planes provides the basis for the extraction of scene information from images and the weak calibration estimates that can be subsequently improved through various optimization strategies [1, 4, 6]. Section 2.1 develops an understanding for the geometric constraints that govern the system. Advantages provided by expected camera parameter structure and a single moving camera are laid out in Section 2.2. Section 2.3 describes the projector constrains and formulation, and Section 2.4 highlights advantages to incorporating reprojection error in the form of Bundle Adjustment as a refinement strategy. This problem formulation assumes unknown scene, unknown moving camera and unknown fixed projector, as plotted in Figure 1.

2.1 Geometry of Image Projection

Projection Model: Both cameras and projectors are modelled with pinhole camera systems where the transformation of any 2D point x on an image plane and its corresponding observed or projected 3D coordinate X is modelled as [7, 8]:

$$x = K \begin{bmatrix} R & c \\ 0^\top & 1 \end{bmatrix} X \tag{1}$$

Here K represents the intrinsic parameters, which includes the principal point (p) and focal length (f). For the extrinsic parameters, R describes the rotation and c describes the baseline vector connecting the position of the camera center within the coordinate system.

Extracting Information from Images: We value the set of information which enables the construction of operations to estimate the geometric relationship between the camera(s), the projector(s), and the scene, as well as the geometry of the scene itself. This is often done with a set *point correspondences*, sets of known points in each image plane. These are typically produced with structured light strategies where reference patterns are projected on the scene surface which encode location information in camera and projector image planes. [9].

These methods allow us to construct our set of point correspondences $x^i_j = (u^i, v^i)$, where $j \in c1, c2, p$ describes the set of points in each camera view or the projector view, and u, v indicate pixel location of correspondence i within the image plane. From these sets of point correspondences, we can begin to construct an understanding of our scene.

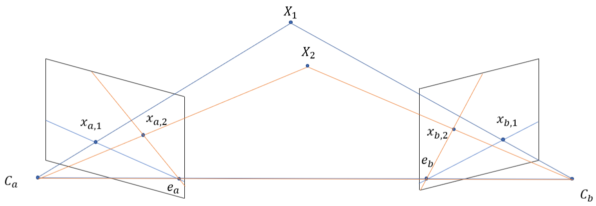

Fig. 2: Epipoles (e_j, C_j) $j = a, b$ connect camera centers C_j and the intersection of line (C_a, C_b) and image planes at e_j. These epipoles connect any observed image point x_a, 3D coordinate X and corresponding image point x_b over a set of epipolar lines $\{(x_a, e_a), (x_b, e_b)\}$ from which geometric understanding of the scene can be constructed.

Epipolar Geometry: Epipolar geometry provides the basic constraint from which we can relate two viewpoints [7, 8]. For two images taken with different camera centers C_a, C_b, points e_a, e_b describe the intersection of line C_a, C_b with the respective image planes. The lines C, e describe the epipoles, which will generate a set of epipolar lines connecting e to any image point x in the image. This behaviour is consistent in both images, and is from this concurrent epipolar behaviour which we begin to relate the images to each other and estimate scene and camera behaviour.
Fundamental Matrix: From this epipolar understanding, the relationship between point x_a and the line (x_b, e) which is a projective and linear consequence of the relationship between x_a, optical ray (C_a, X) and (C_b, X) and it's projection (x_b, e).

$$F x_a \approx (x_b, e) \tag{2}$$

As point x_b belongs to line (x_b, e), it follows that F must satisfy

$$x_b^\top F x_a = 0 \tag{3}$$

This constraint provides the foundation for fundamental matrix estimation strategies which require only the obtained set of point correspondences x_a, x_b, [7, 8]. Estimating F is completed through RANSAC strategies requiring only our point correspondences. F is desirable as it contains a weakly calibrated definition for the system of cameras:

$$F \approx K_a^{-\top} R_a^{-\top} [(C_a, C_b)]_\times R_b^{-1} K_b^{-1} \tag{4}$$

This provides an estimation from which to begin, that may later be improved with optimization strategies.

2.2 Camera Calibration

Two camera image planes provide more measured information from which to construct scene understanding. A moving camera formulation allows for the benefit of multiple camera views while preserving the simplification afforded by requiring only a single set of camera parameters.
Intrinsic Estimate: Where all of the camera parameters are unknown, a set of assumptions are made for the case of a camera-camera calibration problem with two cameras of equal intrinsic parameters.

1. $K_{c1} = K_{c2}$ same intrinsic matrix for both camera views
2. $p_{c1} = p_{c2}$ same principal point for both camera views
3. $p_{c1} = \frac{1}{2}[UV]$ initial principal point for the camera is the centre of the camera image plane for a camera resolution $[U, V]$
4. Assume zero skew and unit aspect ratio

Points 1 and 2 follow from using the same camera to capture both images, provided that the camera is in two different positions. Points 3 and 4 are made based on common camera behaviour of square pixels and center-image principle point axis. From these assumptions, estimation strategies such as Bougnoux's [7] provide an estimate for the focal length f. Having obtained reasonable estimations for all the camera intrinsic parameters, the initial single intrinsic matrix for both camera views is constructed:

$$K_{c1} = K_{c2} \rightarrow K_c = \begin{bmatrix} f & 0 & \frac{1}{2}U \\ 0 & f & \frac{1}{2}V \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

The effort of calibrating the camera is significantly reduced as generalizations about typical camera properties can be made without arduous computation. This allows us to move to estimating pose quite rapidly, where we aim to later recover any loss of accuracy from this generalization.
Extrinsic Estimation: Positioning the world coordinate system such that the camera center for view one aligns with the origin, with zero rotation allows for the following projection matrix definition:

$$P_{c1} = \begin{bmatrix} K_c & 0 \end{bmatrix}, \quad P_{c2} = \begin{bmatrix} K_c' R & K' t \end{bmatrix} \tag{6}$$

where the change in pose is measured with respect to the initial camera view. Provided this assumption and having obtained an estimate of the Fundamental Matrix relating the camera views, it is trivial to extract the Essential Matrix E which describes the pose relating two images for which the calibration is known:

$$E \approx K_c^\top F K_c \tag{7}$$

The essential matrix corresponding to 6 has the form:

$$E \approx [t]_\times R \tag{8}$$

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

2.3 Initial Scene Estimates and Projector Formulation

Unlike camera formulation, projector parameters, particularly the principal point, tend to follow an irregular optical form that make similar generalizations unreliable.

It is common treat the projector image plane the same as a captured camera image plane of measurement signals for the purpose of calibration through fundamental matrix decomposition. In this method, calibration is developed from a set of measurements and projections instead of two sets of measurements. This is further complicated by a irregular projection, and weakened by attempting to solve for all parameters simultaneously under a significant degree of freedom.

Completing this step instead with two camera perspectives, a set of estimated X_E 3D model coordinates can be generated from our obtained point correspondences and known point triangulation methods. As our projector has provided the location encoding pattern for obtaining our camera point correspondences, the locations of corresponding points $x_p \rightarrow X_E$ are known. This strategy provides more measured information from which to construct scene understanding, and allows projector calibration to be constrained by incorporating the projector calibration into a known camera and scene.

2.4 Refinement and Final Scene Estimates

As there are many generalizations made about the parameters and behaviour of the scene, and having rested the algebraic formulation heavily on the weak calibration afforded by the F matrix approach, a recovery strategy is desired to produce a better calibration. Bundle Adjustment (BA) presents a strong framework for adjusting the found calibration of the image planes as well as the scene estimate [1, 8, 10]. By evaluating reprojection error, BA aims to refine a visual reconstruction to produce jointly optimal 3D scene and camera calibration estimates:

$$\min_{\Theta_j, X_E} \frac{1}{2} \frac{1}{N} \sum_j \sum_{i=1}^{N} ||x_j^i - \pi(\Theta_j, X_E^i)||^2 \tag{9}$$

over $i = 1 : N$ image points, where π describes the $3D \rightarrow 2D$ projection (1) of estimated parameters $\Theta_j = \{K_j, d_j, R_j, t_j\}$, $j = \{c1, c2, p\}$. This refinement strategy emphasizes a final calibration and scene that approximates the captured measurements, and provides a level of flexibility that allows for subsequent view incorporation (e.g. before and after projector incorporation, additional camera views).

3 Conclusion

In this paper the problems of projector calibration and scene estimation for non planar scenes has been formulated. This method employs the use of a single moving camera, gaining the scene understanding advantages of multiple camera perspectives, the environmental flexibility and adaptability of an unknown camera initialization, and the simplification of requiring only one set of camera intrinsic parameters. An irregular projector calibration is better constrained by incorporation into a scene with definition. A moving camera formulation forms the foundation of future handheld camera calibration for non-planar scene estimates and projector calibration. There is also potential for incorporating strategies to rapidly assess scene knowledge and overcome scene occlusions, scene size challenges, or similar camera view challenges.

Acknowledgments

This research would not be possible without the support of the Natural Sciences and Engineering Research Council of Canada (NSERC-CRD) and Christie Digital Systems Inc.

References

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

[2] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

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

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

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

[6] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, and A. Wong, "Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 74–83, 2017.

[7] O. Faugeras and Q.-T. Luong, *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2001.

[8] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

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

[10] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment — a modern synthesis," in *Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman, and R. Szeliski, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 298–372.