UKRAINIAN CATHOLIC UNIVERSITY

MASTER THESIS

Corner localization and camera calibration from imaged lattices

Author:
Andrii STADNIK

Supervisor:
Dr. James PRITTS

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2022

Enable

Enable

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Master of Science

Corner localization and camera calibration from imaged lattices

by Andrii STADNIK

Abstract

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Camera calibration is a crucial step in many computer vision applications. Typically, it involves taking a set of images of a calibration pattern, detecting its' features, and estimating the camera parameters. However, under certain conditions (including occlusions, bad lightning, highly distorted images etc.), feature detectors might fail to detect some of the features..

Do I have to justify it?

In this paper, we propose a novel approach to feature detection in the context of camera calibration. After the initial feature detection and calibration, we use the intermediate camera parameters to predict the possible positions of the previously undetected features. Those features are filtered using the binary classifier, and the remaining features are used to further constrain the camera calibration.

Enable

Enable

Enable

Enable

Todo list

1

Remove list of todos

1. Introduction and motivation
   (a) Outline of the problem
   (b) Research objective
       - Finding new features by projecting the board
       - Robustly classifying the true and false positives
   (c) Thesis structure
2. Related work
   (a) Camera calibration
   (b) Calibration boards
   (c) Feature detection on calibration boards
   (d) Camera models
   (e) Camera parameters estimation
   (f) Search of the previously undetected features
3. Background
   (a) Notation
   (b) Camera model
       - Distortion model, inverse distortion model
       - Projection
       - Backprojection
   (c) Feature detection
   (d) Camera calibration (Probably I could move it to the Approach)
4. Approach
   (a) Feature detection
   (b) Camera calibration
   (c) Board projection, features classification

I'll have to merge something to have 5 chapters

5. Experiments
   (a) Synthesizer
   (b) Datasets
   (c) Metrics
   (d) Results
6. Conclusions

Chapter 1

Introduction and motivation

1.1 Outline of the problem

Better camera calibration improves the performance of various downstream tasks by providing a more accurate mapping between 3D world coordinates and 2D image plane coordinates. This improved mapping enables precise alignment, positioning, and scaling of objects within the scene. By determining the camera's intrinsic and extrinsic parameters, algorithms can correct for lens distortion, estimate depth information, and accurately overlay virtual content. Consequently, tasks such as 3D reconstruction, augmented reality, and object detection can achieve better results in terms of precision, spatial consistency, and overall visual quality.

Although manufacturers can estimate camera calibration parameters a priori, fully automatic calibration is often preferred, especially when camera metadata is unavailable. Currently, wide-angle lenses, particularly in mobile phones and GoPro-type cameras, dominate consumer photography. These cameras pose additional challenges due to their requirement for highly non-linear models with numerous parameters. The high distortion of the image plane also makes finding key points robustly challenging.

Typically, camera calibration is obtained by capturing an image of a known calibration pattern, which is then used to estimate the camera parameters. Alternatively, some methods do not use a calibration pattern but instead infer geometric constraints directly from the scene. However, this approach is generally less accurate.

As reported by Duisterhof et al. (2022) on Oct. 5, 2022, the current state-of-the-art methods (**olsonAprilTagRobustFlexible2011a**; Schöps et al., 2020; Krogius, Haggenmiller, and Olson, 2019) fail on images with high distortion. Duisterhof et al., 2022 suggested an iterative the approach of image undistortion and target reprojection, achieving the superior robustness to the noise than the state-of-the-art methods because the feature detection is performed on the undistorted image.

Instead of searching for the features on the undistorted image from scratch, it is possible to utilize the prior knowledge of the geometry of the calibration board, effectively predicting the possible positions of previously undetected features. It can be done by projecting the board onto the image using the intermediate camera calibration, and then filtering the possible positions in order to eliminate false positives.

This additional points will further constrain the camera calibration, improving the accuracy of the calibration parameters.

Explain why more features is better

Missing figure

Show the example of the calibration board and the image

1.2 Research objective

The objective of this research is to improve the accuracy of the camera calibration by finding previously undetected features on the calibration board . For that, we formulate the set of research questions:

Probably rephrase if I won't compare the actual calibration

- How to find additional features on the calibration board which were not detected by the feature detector?
- How to filter out falsely detected features?

Add more

1.3 Thesis structure

This paper has the following structure: in **??**, we will describe the related work, including the literature search method and methodology, various subtopics of the camera calibration, mention conjugate translations, and outline the state-of-the-art solutions. We define the research gap in **??** and outline the proposed approach to solution and evaluation in **??**. We will describe the early results in **??**, including the dataset analysis, feature detector, and conjugately translated points simulator. In **??**, we will summarize the results and outline future work.

Update

Chapter 2

Related work

2.1 Camera calibration

Getting the correspondence between the spatial and the image coordinates requires camera calibration. Camera calibration consists of the geometric camera model and the parameters of this model. That information makes it possible to obtain the 2d image coordinates of any point in the 3d space.

Usually, the geometric camera model is obtained from the domain knowledge of the researcher or the camera manufacturer. Often, they choose the simplified model as a trade-off between accuracy and complexity. The model's parameters are usually obtained by solving the constrained optimization problem, given the set of points with known geometry.

2.2 Calibration boards

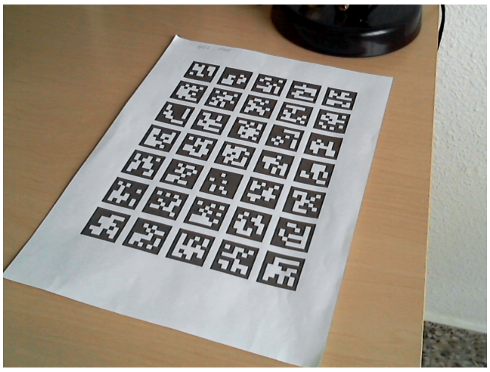
To achieve a robust calibration, images with repeating patterns are usually used. The camera calibration parameters can be found using prior knowledge of the properties of the pattern, such that the pattern invariants hold on the image. Initially, the chessboard (*OpenCV: Camera Calibration* 2023; V. Douskos, I. Kalisperakis, and G. Karras, 2007) patterns were used (fig. 2.1a).

Later, ArUco (Garrido-Jurado et al., 2014) (fig. 2.1b) and AprilTag (**olsonAprilTagRobustFlexible2011a**) (fig. 2.1d) allowed detecting the orientation of the pattern, as well as uniquely identifying each located pattern even under occlusion. Based on ArUco, ChArUcO (*OpenCV: Camera Calibration* 2023) (fig. 2.1c) was proposed as more robust.

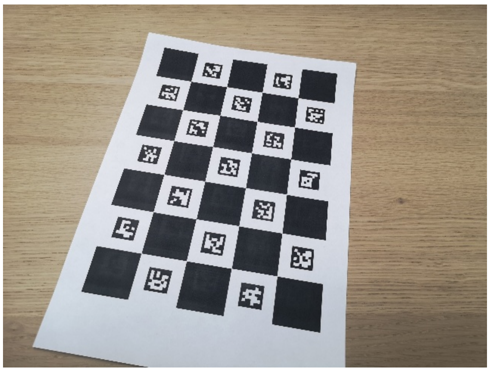
2.3 Camera models

The choice of the camera model depends on the camera's physical properties and the accuracy required. Usually, the parametric models are simpler to use, as they have only a few parameters and deliver good accuracy. The most common are the Double Sphere model (Usenko, Demmel, and Cremers, 2018), the Kannala-Brandt model (Kannala and Brandt, 2006), and the Field-of-View model (Devernay and Faugeras, 2001). In the ill-posed problem of camera calibration, the common choice of the camera model is the division model (Fitzgibbon, 2001). However, Schöps et al., 2020 shows that they tend to have significantly higher errors than the non-parametric (general) models. The Lochman et al., 2021 suggested a framework for converting the parameters of a powerful back-projection Zhang, 2000 model to recover different models' parameters.

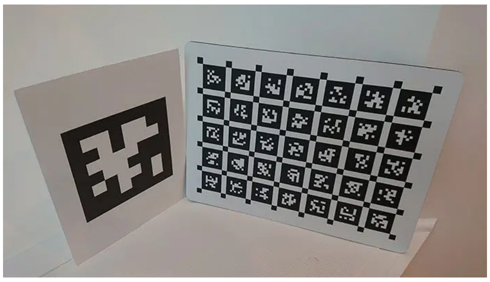
(A) Chessboard *OpenCV: Camera Calibration* 2023



(B) ArUco board *OpenCV: Detection of ArUco Markers* 2023



(C) Charuco board *OpenCV: Detection of ChArUco Boards* 2023



(D) AprilTag board Rosebrock, 2020 (right)

FIGURE 2.1: Calibration boards.

2.4 Camera parameters estimation

Camera calibration using repeating patterns was an important subject for a long time, for example, Schaffalitzky and Zisserman, 1998 in 1998 and Zhang, 2000.

Nevertheless, camera calibration is still an open problem; recently, multiple new approaches have arisen. Lochman et al., 2021 suggest a universal approach to camera calibration, with a separate step of converting between camera models. Hu et al., 2019 used deep learning to detect ChArUcO boards. Recently, on Oct. 5, 2022, Duisterhof et al., 2022 introduced the iterative approach to camera calibration, which outperforms the previous state-of-the-art approaches for wide-angle cameras.

Add the paper about the feature detection

Chapter 3

Background

3.1 Notation

Term	Description
$\mathbf{u} = (u, v, 1)^T$	A point in the board coordinate system

TABLE 3.1: Notation

3.2 Camera model

In this paper, scene and image points are represented using homogeneous coordinates. This approach allows representing many geometric transformations as linear, which simplifies the mathematical representation of the camera model.

3.2.1 Perspective projection

The perspective projection is a mapping from a 3D point $(X, Y, Z)^T$ in the world coordinate to the 2D coordinate $(u, v)^T$ on the image plane which is distance f from the center of projection. It is given by the perspective projection equation:

Should f not be italic?

$$(u, v)^T = \frac{f}{Z} (X, Y)^T .$$

This equation can be written using the homogeneous coordinates as:

$$\alpha \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}^T = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}^T , \tag{3.1}$$

where $\alpha = {}^1/_Z$ is a scale factor.

3.2.2 Scene to camera projection

A 3D scene point $(X, Y, Z)^T$ can be projected onto the image plane as $R\,(X, Y, Z)^T + \mathbf{t}$, where R is a 3×3 rotation matrix and $\mathbf{t}$ is a 3×1 translation vector. Using the homogeneous coordinates, this can be written as:

$$\begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} . \tag{3.2}$$

3.2.3 Camera to image projection

To project a point from the camera coordinate system to the image plane, we need to apply a homography encoding the camera intrinsic parameters. This is a 3×3 upper-triangular matrix:

$$\begin{bmatrix} \alpha_x & \alpha_x \cot\theta & c_x \\ 0 & \alpha_y \sin\theta & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

where:

- α_x and α_y represent the scale factor of the camera in terms of $^{\text{pixels}}/_{\text{mm}}$ in the x and y directions respectively.
- c_x and c_y are the coordinates of the principal point, which is typically the image center.
- $\cot\theta$ and $\sin\theta$ are related to the skew coefficient, which measures the angle between the x and y pixel axes. The variable θ represents this angle.

Add reference

. For a typical camera, $\theta = \pi/2$ and $\alpha_x = \alpha_y = 1$.

Conventionally, the intrinsic matrix incorporates the scaling, introduced by the focal length:

$$K = \begin{bmatrix} \alpha_x f & \alpha_x \cot\theta & c_x \\ 0 & \alpha_y f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_x & k & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} . \tag{3.4}$$

specify them

By incorporating the assumptions into the intrinsic matrix, we can simplify it to:

$$K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} . \tag{3.5}$$

3.2.4 Camera matrix

Add reference

The composition of positioning and orienting the camera, projection, and the imaging transformation can be represented by a 3×4 camera matrix. This matrix can be expressed as:

$$\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix} = K \left[I_3 | \mathbf{0} \right] \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} = K \left[R | \mathbf{t} \right] , \tag{3.6}$$

Hence, the transformation of a point in the scene by the camera $\mathrm{P}^{3\times4}$ can be formulated as:

$$\alpha (u, v, 1)^T = P^{3\times4} (X, Y, Z, 1)^T , \tag{3.7}$$

with α being $1/Z$.

3.2.5 Projection of the points from the scene plane

When working with the coplanar scene points, we can simplify the projection by assuming that the scene plane is located at $Z = 0$. In this case, the projection of the point becomes:

$$\alpha \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3 \quad \mathbf{p}_4] \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \underbrace{[\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_4]}_{P} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}. \tag{3.8}$$

3.2.6 Distortion

The distortion of the image is caused by the lens not being perfectly planar. Typically, the small distoritons caused by lens misalignment are ignored , allowing us to model the distortion as radially symmetric. Then, the function that maps a point $\mathbf{u} = (u, v, 1)^T$ from a retinal plane to the ray direction in the camera coordinate system is given by:

add reference

Specify what's that

$$g(\mathbf{u}) = (u, v, \psi(r(\mathbf{u})))^T, \tag{3.9}$$

where $r(\mathbf{u}) = \sqrt{u^2 + v^2}$ is the radial distance from the principal point.

Back-projection using the Division Model

The division model has a good ability to model the distortion of the wide-angle lenses, and is wildly used . The model is defined as:

Add reference

$$\psi(r) = 1 + \sum_{n=1}^{N} \lambda_n r^{2n}, \tag{3.10}$$

where λ_n are the distortion coefficients.

The function $\psi(r)$ is not invertible in general. Let $\mathbf{X} = (X, Y, Z)^T = \alpha g(\mathbf{u})$ be a ray in the camera coordinate system.

Should I explain somewhere how the rays are defined?

Then,

$$\frac{\mathbf{X}}{Z} = \left(\tfrac{X}{Z}, \tfrac{Y}{Z}, 1\right)^T \tag{3.11}$$

$$= \left(\frac{\alpha u}{\alpha\psi(r(\mathbf{u}))}, \frac{\alpha v}{\alpha\psi(r(\mathbf{u}))}, 1\right)^T \tag{3.12}$$

$$= \left(\frac{u}{\psi(r(\mathbf{u}))}, \frac{v}{\psi(r(\mathbf{u}))}, 1\right)^T \tag{3.13}$$

From 3.13 we see that

$$\begin{cases} \frac{X}{Z} = \frac{u}{\psi(r(\mathbf{u}))} \\ \frac{Y}{Z} = \frac{v}{\psi(r(\mathbf{u}))} \end{cases} \implies \begin{cases} u = \frac{X\psi(r(\mathbf{u}))}{Z} \\ v = \frac{Y\psi(r(\mathbf{u}))}{Z} \end{cases} \tag{3.14}$$

Now, let $\hat{r}$ be a root of $r(\mathbf{u}) = \sqrt{\frac{X*\psi(r)}{Z}^2 + \frac{Y*\psi(r)}{Z}^2} = r$.
Then, $\mathbf{u} = \frac{\hat{r}}{r(\mathbf{X})}\mathbf{X}$.

3.3 Feature detection

Chapter 4

Approach

4.1 Feature detection

4.2 Classifier

Chapter 5

Conclusions

Bibliography

Devernay, Frédéric and Olivier Faugeras (Aug. 1, 2001). "Straight Lines Have to Be Straight". In: *Machine Vision and Applications* 13.1, pp. 14–24. ISSN: 1432-1769. DOI: 10.1007/PL00013269. URL: https://doi.org/10.1007/PL00013269 (visited on 01/15/2023).

Duisterhof, Bardienus P. et al. (Oct. 5, 2022). *TartanCalib: Iterative Wide-Angle Lens Calibration Using Adaptive SubPixel Refinement of AprilTags*. DOI: 10.48550/arXiv.2210.02511. arXiv: 2210.02511 [cs]. URL: http://arxiv.org/abs/2210.02511 (visited on 01/05/2023). preprint.

Fitzgibbon, A.W. (Dec. 2001). "Simultaneous Linear Estimation of Multiple View Geometry and Lens Distortion". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Vol. 1, pp. I–I. DOI: 10.1109/CVPR.2001.990465.

Garrido-Jurado, S. et al. (June 1, 2014). "Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion". In: *Pattern Recognition* 47.6, pp. 2280–2292. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2014.01.005. URL: https://www.sciencedirect.com/science/article/pii/S0031320314000235 (visited on 01/05/2023).

Hu, Danying et al. (July 1, 2019). *Deep ChArUco: Dark ChArUco Marker Pose Estimation*. DOI: 10.48550/arXiv.1812.03247. arXiv: 1812.03247 [cs]. URL: http://arxiv.org/abs/1812.03247 (visited on 01/05/2023). preprint.

Kannala, Juho and Sami Brandt (Sept. 1, 2006). "A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses". In: *IEEE transactions on pattern analysis and machine intelligence* 28, pp. 1335–40. DOI: 10.1109/TPAMI.2006.153.

Krogius, Maximilian, Acshi Haggenmiller, and Edwin Olson (Nov. 2019). "Flexible Layouts for Fiducial Tags". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1898–1903. DOI: 10.1109/IROS40897.2019.8967787.

Lochman, Yaroslava et al. (Oct. 28, 2021). *BabelCalib: A Universal Approach to Calibrating Central Cameras*. DOI: 10.48550/arXiv.2109.09704. arXiv: 2109.09704 [cs]. URL: http://arxiv.org/abs/2109.09704 (visited on 01/05/2023). preprint.

OpenCV: Camera Calibration (2023). URL: https://docs.opencv.org/4.x/dc/dbb/tutorial_py_calibration.html (visited on 01/15/2023).

OpenCV: Detection of ArUco Markers (2023). URL: https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html (visited on 01/15/2023).

OpenCV: Detection of ChArUco Boards (2023). URL: https://docs.opencv.org/3.4/df/d4a/tutorial_charuco_detection.html (visited on 01/30/2023).

Rosebrock, Adrian (Nov. 2, 2020). *AprilTag with Python*. PyImageSearch. URL: https://pyimagesearch.com/2020/11/02/apriltag-with-python/ (visited on 01/30/2023).

Schaffalitzky, Frederik and Andrew Zisserman (Sept. 18, 1998). "Geometric Grouping of Repeated Elements within Images". In: *In Shape, Contour and Grouping in Computer Vision LNCS* 1681. ISSN: 978-3-540-66722-3. DOI: 10.5244/C.12.2.

Schöps, Thomas et al. (June 23, 2020). *Why Having 10,000 Parameters in Your Camera Model Is Better Than Twelve*. DOI: 10.48550/arXiv.1912.02908. arXiv: 1912.02908 [cs]. URL: http://arxiv.org/abs/1912.02908 (visited on 01/05/2023). preprint.

Usenko, Vladyslav, Nikolaus Demmel, and Daniel Cremers (July 24, 2018). *The Double Sphere Camera Model*.

V. Douskos, I. Kalisperakis, and G. Karras (2007). "Automatic Calibration of Digital Cameras Using Planar Chessboard Patterns". In.

Zhang, Z. (Nov. 2000). "A Flexible New Technique for Camera Calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11, pp. 1330–1334. ISSN: 1939-3539. DOI: 10.1109/34.888718.