UKRAINIAN CATHOLIC UNIVERSITY

MASTER THESIS

Minimal Solvers for Single-View Auto-Calibration

Author:
Yaroslava LOCHMAN

Supervisors:
James PRITTS
Oles DOBOSEVYCH
Rostyslav HRYNIV

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 2020

Declaration of Authorship

I, Yaroslava LOCHMAN, declare that this thesis titled, "Minimal Solvers for Single-View Auto-Calibration" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Master of Science

Minimal Solvers for Single-View Auto-Calibration

by Yaroslava LOCHMAN

Abstract

We introduce a new hybrid minimal solver that admits combinations of radially-distorted conjugate translations and radially-distorted parallel lines from the common scene plane to jointly estimate lens undistortion and affine rectification. The solver is the first to admit complementary geometric primitives for rectification purposes. In addition, a novel solver admitting three pairs of imaged parallel scene lines for the same problem is introduced. The proposed solvers are used with the Manhattan scene assumption to auto-calibrate cameras from a single image. The solvers are generated using elementary methods from algebraic geometry. As a result, they are simple, fast and robust. The solvers are used in an adaptive sampling framework that favors the feature combinations that are most frequently consistent with accurate scene plane rectifications. Auto-calibrations are recovered from challenging images that have either a sparsity of scene lines or scene texture. The method is fully automatic.

Acknowledgements

Yaroslava Lochman acknowledges **ELEKS Ltd.** for generous funding of this project. Yaroslava also thanks **Horst Wildenauer** for sharing the data and code of [40, 41] for evaluation. Yaroslava expresses her deepest gratitude to her supervisors: **James Pritts** for advising, inspiring, encouraging and sharing ideas on developing the proposed approach, infinitely helping during the project work, especially on solvers verification, evaluation strategy and integration of the proposed solvers into the robust framework developed by James Pritts [30, 31]; **Oles Dobosevych** and **Rostyslav Hryniv** for numerous discussions, for sharing ideas and constantly helping during the work on this thesis, especially on circular arcs extraction, circle fitting, examining solvers and runtime experiments. Integration of the subpixel Canny edge detection, arcs connection pipeline, and C++ implementation of the proposed solvers was done by Oles Dobosevych.

Contents

List of Figures

List of Tables

List of Abbreviations

BMSS	**B**est **M**inimal **S**olution **S**election
CCD	**C**harge-**C**oupled **D**evice
IAC	**I**mage of **A**bsolute **C**onic
LAF	**L**ocal **A**ffine **F**rame
LCC	**L**ine of **C**ircle **C**enters
MLE	**M**aximum **L**ikelihood **E**stimation
MSER	**M**aximally **S**table **E**xtremal **R**egions
MSS	**M**inimal **S**ample **S**et
ORUA	**O**rthogonal **R**aster **U**nit **A**spect
PP	**P**rincipal **P**oint
RANSAC	**RAN**dom-**S**ample **C**onsensus
SIFT	**S**cale-**I**nvariant **F**eature **T**ransform

To my family.

Chapter 1

Introduction and Motivation

1.1 Outline of the Problem

This thesis proposes two minimal solvers that jointly estimate affine rectification and lens undistortion from complementary combinations of point correspondences extracted from radially-distorted conjugately-translated coplanar texture and the distorted images of parallel scene lines. In particular, the proposed solvers are the first single-view minimal solvers that admit complementary feature types. Furthermore, the proposed solvers admit all minimal configurations of imaged translations and imaged parallel scene lines for affine rectification of radially-distorted images within the division model of lens distortion [20]. Sampling complementary feature types extends the class of highly-distorted images where high-accuracy rectification and auto-calibration are possible.

Joint undistortion and rectification of imaged scene planes and camera auto-calibration are closely related tasks. Both are notoriously ill-posed single-view geometry estimation problems [29]. Good feature coverage over large spans of the image is necessary to properly constrain these estimation problems. The proposed solvers can leverage combinations of corners, similarity-covariant regions, affine-covariant regions, and contours as feature types. This flexibility to leverage complementary feature types increases the chances of densely sampling constraints from enough image regions so that the joint effects of perspective imaging and lens distortion are sufficiently observable to the estimation framework which, in this work, is a locally-optimized hybrid RANSAC with the proposed minimal solvers generating models from minimal samples [13, 9].

Metric rectification and auto-calibration both are estimations that essentially rely on affine rectification. In fact, there is no minimal solver that jointly undistorts and metrically-rectifies. Thus metric rectification

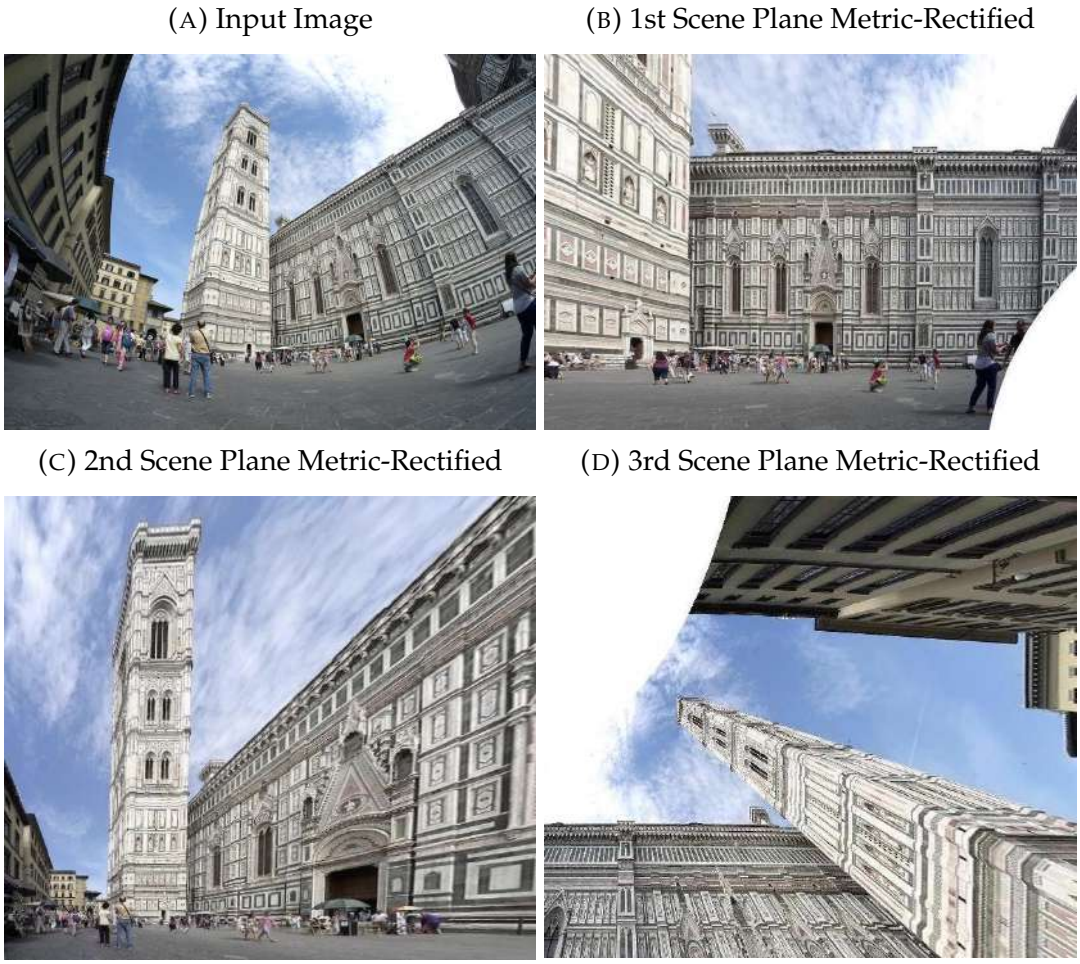


FIGURE 1.1: *Auto-Calibration Result.* Imaged scene planes are rectified using the estimated distortion, focal length, and camera rotation. (A) An input image. (B) The dominant scene plane rectified, which is given by the plane with the most features. (C) The second scene plane rectified. (D) In addition, the third "scene plane" can also be rectified, in the example, there is no plane but the sky. The method is fully automatic.

from distorted images must be achieved by upgrading an affine rectification. Single-view auto-calibration typically relies on the Manhattan assumption, which assumes that two or three recovered vanishing points correspond to orthogonal directions in the scene. Since the vanishing line is sufficient to estimate affine rectification and the vanishing points are estimated from planar features, affine rectification is implicitly needed to auto-calibrate from the Manhattan assumption, which is accomplished by recovering the image of the absolute conic [3, 21].

The proposed solvers directly recover lens undistortion and affine rectification. This eliminates the need to check for pairwise consistency between recovered vanishing points; however, the vanishing points are

also recovered by the solvers as nuisance parameters. Metric rectification can be recovered if reflected or rotated features are present [21, 22, 28]. If the Manhattan world is assumed, then auto-calibration (and metric rectification) can be attempted from the minimal sample used to estimate the affine rectification.

The code for the proposed minimal solvers, evaluation techniques, and experiments has been made available at:
https://github.com/ylochman/auto-calibration.

1.2 Structure of the Thesis

Related Work Chapter 2 gives an overview of the recent methods for radial lens undistortion, affine and metric rectification, vanishing point detection and camera auto-calibration from a single image. A comparison with the proposed work is provided. The contributions of the thesis are enumerated.

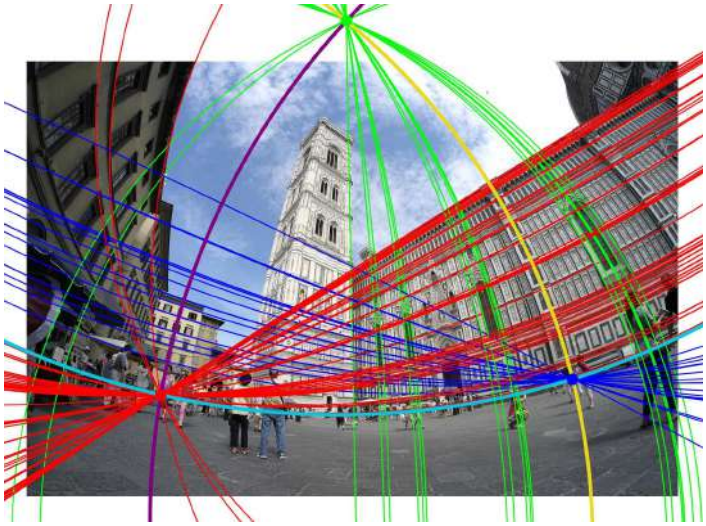
Background Chapter 3 introduces denotations and provides the necessary theoretical background to model radially-distorted cameras viewing scene planes.

Proposed Minimal Solvers Chapter 4 formulates the problem of auto-calibrating distorted cameras using the Manhattan scene assumption with the proposed minimal solvers. The feature configurations that are inputs to the proposed solvers are defined. Best Minimal Solution Selection is introduced, which is an optimization problem to choose the best constraints to use in the minimal solver for overconstrained feature configurations.

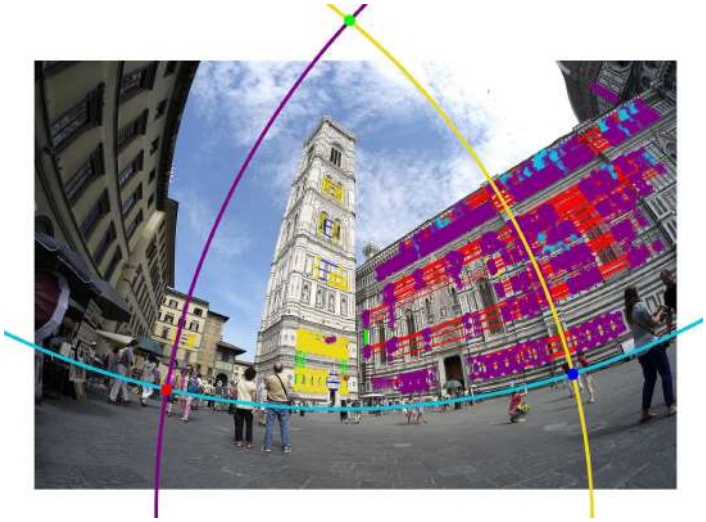
Experiments Chapter 5 discusses the synthetic scene generation and related experiments that are used to test the stability and noise sensitivity of the proposed solvers. The auto-calibration warp error is introduced, which is a summary measure of the quality of auto-calibration. Chapter 5 also describes the pipeline and data used for experiments on real images and provides a comparative analysis of the proposed and related methods from conducted performance benchmarks and experimental results.

Conclusions Chapter 6 summaries the idea, main contribution, and the potential impact of this thesis work. Experimental results are summarized.

(A) Barrel Distorted Image with Arc Labeling



(B) Barrel Distorted Image with Region Labeling



(C) Undistorted Image

FIGURE 1.2: *Manhattan Scene Parsing*. (A-B) Input is a scene containing a plane with features in orthogonal directions (A) labelling of line correspondences with vanishing points of the Manhattan directions (red, green, blue) and (B) labelling of radially-distorted conjugately-translated features on the scene plane *i.e.*, vanishing lines (yellow, magenta, cyan), and the directions of the conjugate translations of the repeats with the vanishing points; (C) image undistorted with the estimated division model parameter.

Chapter 2

Related Work

2.1 Overview and Analysis

We concern ourselves with the single-view geometry problems of radial lens undistortion, affine and metric rectification, vanishing point detection, and camera auto-calibration. In this section, we give an overview of the recent work on these topics, compare it to the proposed work and highlight our contributions.

2.1.1 Rectification and Undistortion

Various techniques for rectification from imaged coplanar repeated pattern were proposed in recent years [16, 12, 2, 45, 23, 1], however, all these methods assume the pinhole camera model *i.e.*, do not model real camera lens distortion. Pritts et al. have proposed several joint undistorting and rectifying minimal solvers that admit various minimal configurations of affine-covariant regions [32, 33, 30]. The solvers assume either translational symmetries or rigidly-transformed scene plane content. The solvers are sensitive, and the noise experiments in [32, 33, 30] report the best solution after RANSAC. Dense planar texture is required to achieve good results with these solvers.

The proposed by Pritts et al. [30] Directly-Encoded Scale (DES) solvers use the invariant that rectified coplanar repeats have equal scales. This invariant is utilized to construct a system of polynomial constraint equations on rectified coplanar repeats to solve for the vanishing line and radial undistortion parameter. The two-direction Eliminated Vanishing Point (EVP) solvers proposed by [31] use the invariant that the undistorted meet of joins of radially-distorted conjugately-translated point correspondences is a vanishing point lying on the vanishing line. The solvers require two similarity-covariant region correspondences that provide two pairs of two point correspondences relating to two vanishing points of the translation direction, respectively. The Gröbner basis method and the hidden-variable trick [15] were used to solve the

systems of polynomial equations arising from the constraints given by the aforementioned geometrical properties.

Antunes et al. [3] and Wildenauer et al. [41] are two methods that affinely rectify lens-distorted images from circles fitted to contours extracted from the image by enforcing the constraint that scene lines are imaged as circles under the division model of radial lens distortion. The problem is simplified by assuming orthogonal raster (zero skew), and unitary aspect ratio, (square pixels), also referred further as ORUA. Sets of circles whose preimages are parallel scene lines are used to induce constraints on lens undistortion and vanishing point location. These methods require two or three distinct sets of imaged parallel lines to estimate rectification, which is a strong scene-content assumption.

2.1.2 Vanishing Point Estimation and Auto-Calibration

An earlier work of Wildenauer et al. [40] proposes minimal solvers that auto-calibrate camera with known principal point; it requires four lines, two of which have the same direction in the scene (or a single vanishing point). However, the method does not account for radial lens distortion which is usually presented on images.

Wildenauer et al. [41] proposed a method incorporating lens distortion estimation that requires five lines, three of which are used to estimate the single parameter of the division model of radial lens distortion and the undistorted vanishing point. The remaining lines are undistorted with estimated distortion model and, together with the first estimated vanishing point, are used in the line-based method of [40] in the undistorted image space.

The method of Antunes et al. [3] estimates the focal length, division model parameter and relative orientation of the camera with respect to the scene plane from seven fitted circles, four of which correspond to lines with the same direction in the scene, and the remaining three correspond to lines in the scene orthogonal to the first four. The auto-calibration pipeline of [3] also estimating the principal point, which, in this work, is assumed to be the image center.

The two-direction EVP solvers of [31] recover two vanishing points, which, if orthogonal, can be used to estimate the auto-calibration parameters under the assumptions made in this work. See Table 2.1 for details about the scene assumptions, features employed, and parameters recovered by the state-of-the-art and the proposed methods.

2.1.3 Comparative Analysis

The proposed minimal solvers complement radially-distorted conjugately-translated scene texture with circles fitted to imaged parallel scene lines. Chapter 4 derives the systems of equations that jointly admit constraints from repeated local features and parallel scene lines. Furthermore, the proposed solvers directly return the vanishing line with lens undistortion, which is sufficient for affine rectification. The vanishing points of the translation directions of the covariant regions and the imaged parallel scene lines are also recovered. Furthermore, the proposed solvers require fewer correspondences than the solvers of [3, 41] and have a much faster time to solution than the solvers introduced in [33, 30].

The fast time to solution and minimal feature correspondence requirements of the proposed solvers make them good candidates for use in a RANSAC framework [19]. However, the classic RANSAC sampling strategy cannot accommodate multiple feature types, *i.e.*, covariant regions and contours. The scene content should jointly inform the sampling strategy: covariant regions are sampled more frequently for texture-rich scenes, while sampled contours are preferred for scenes with many lines. Furthermore, the sampler's preference for a particular hybrid solver should be mitigated if it is relatively less performant than the others. Camposeco et al. [9] propose a method for pose estimation that uses hybrid solvers that admit combinations of 2D-2D and 2D-3D correspondences. However, the approach of [9] assumes that features have a one-to-one correspondence, which is not the case for repeated patterns, which have many-to-many correspondences. We propose a RANSAC framework that accommodates hybrid solvers admitting complementary feature types (distorted lines and covariant regions) with many-to-many correspondences.

As in [40], the proposed framework utilizes a Manhattan scene assumption to auto-calibrate cameras, specifically, to estimate the focal length and scene-rectifying camera rotation. For the proposed minimal solutions, the assumption does not affect either lens undistortion or vanishing line estimation. However, auto-calibrations recovered from the proposed minimal solvers are refined by a local optimizer that imposes joint constraints from radially-distorted conjugate translations and imaged parallel lines on the focal length, lens undistortion and the rectifying rotation, which affects vanishing line detection, too. The scene, as modeled by translated points and parallel lines, is constructed on a plane in calibrated space. The difference between the imaged reconstruction, as viewed by the auto-calibrated camera and the extracted covariant regions and contours, is minimized. A generative model of repeated patterns was used in Pritts et al. [28]; however,

this method does not impose constraints from scene lines and does not auto-calibrate. Instead, the pattern is reconstructed in rectified space, which provides weaker constraints.

Approach	Features	Assumption	Parameters recovered λ	f, R	**c**
Wildenauer et al. [41]	3 arcs in cspond + 2 additional fitted arcs	Manhattan scene, lines parallelism, ORUA, PP at the image center	+	+	-
Antunes et al. [3]	4 arcs in cspond + 3 arcs in cspond	Manhattan scene, lines parallelism, ORUA	+	+	+
EVP of Pritts et al. [30]	2 LAFs in cspond + 2 LAFs in cspond	Manhattan scene, translated repeats, ORUA, PP at the image center	+	+	-
Proposed $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$	2 LAFs in cspond + 2 arcs in cspond or 3 pairs of arc csponds	Manhattan scene, translated repeats and/or lines parallelism, ORUA, PP at the image center	+	+	-
Proposed $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$					

TABLE 2.1: *Scene Assumptions, Feature Configurations, Recovered Calibration Parameters.* The proposed hybrid solver $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ jointly admits two radially-distorted affine-covariant regions in correspondence (cspond in the table) and two corresponded radially-distorted lines for affine rectification of distorted images. Moreover, we propose a solver $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ that jointly undistorts and rectifies from three correspondences of distorted lines. In contrast, [41] requires a corresponded set of three arcs and two arcs corresponding to sets of parallel lines on the scene plane, and [3] requires two distinct set of four and three arcs corresponding to sets of parallel lines on the scene plane.

2.2 Contributions

In this work, several contributions to the state-of-the-art for imaged scene plane rectification and single-view autocalibration are proposed:

- A new minimal hybrid solver is developed that uses combinations of radially-distorted parallel lines and radially-distorted conjugately-translated affine-covariant regions to jointly undistort and rectify the imaged scene plane, estimate the scene plane's vanishing points and, auto-calibrate the camera if the image is of a Manhattan scene. This is the first single-view method admitting such complementary geometric features.
- A new minimal solver admitting three pairs of imaged parallel scene lines for the same problem is introduced. Both types of

solvers are derived in a unified form, utilizing elementary techniques from algebraic geometry. The problem reduces to requiring the solution of a quartic and a small homogeneous linear system, both of which are solved in closed form. This makes the proposed solvers extremely fast and robust.

Moreover, if the image is not of a Manhattan scene, the solvers can be used for the task of undistortion and rectification. The proposed solvers admit different configurations of input features that are consistent with either two or three independent vanishing points. This number of vanishing points present can only be tested with consensus sets, which is done quickly with best minimal solution selection (see next item).

- A technique proposed in [31], called best minimal solution selection (BMSS), is adapted for the single-view auto-calibration problem from complementary features. The configurations of features that are admitted by the proposed solvers provide more than one option for drawing the minimal sample set (MSS) from an input sample set of features required by the proposed solvers. We call a solver incorporating BMSS an *optimal solver*. Minimizing the BMSS objective gives the best sample of minimal constraints. BMSS increases robustness and filters out near degenerate or putative incorrect configurations, which eliminates consensus set construction and hypothesis evaluation in the verification step of RANSAC.

- For measuring the accuracy of camera auto-calibration, a metric warp error is introduced, which is an extension of the measure proposed in [29]. The metric warp error accounts for all parameters relating to auto-calibration: 1. radial undistortion, 2. focal length, 3. and conjugate camera rotation.

- Experiments show that the proposed minimal solvers achieve state-of-the-art results in undistortion, rectification, and auto-calibration. The solvers were integrated into a fully automated robust framework and tested on several benchmarking datasets with challenging images that have either a sparsity of scene lines or scene texture. Highly accurate results were obtained.

Chapter 3

Background

3.1 Notations

Table 3.1 outlines the common denotations used for solver derivation and analysis.

We adapt the solver naming convention of Pritts et al. [31] to the proposed and state-of-the-art solvers studied in this paper. The minimal configuration of region correspondences is given as the subscript to $\mathtt{H}$ denoting a homography. The minimal configuration of parallel lines is given as the superscript to $\mathtt{H}$. E.g., a solver requiring 2 affine- region correspondences and a pair of parallel lines is denoted $\mathtt{H}^{2}_{2}$, a solver requiring three scene lines in correspondence with a single vanishing point plus another pair of scene lines — $\mathtt{H}^{32}$.

The unknowns that are recovered by the solver are suffixed to $\mathtt{H}.$, the unknowns from the auto-calibration upgrade are also suffixed, separated by a hyphen to emphasize that it is possible to recover these parameters with an additional assumption of the orthogonal vanishing points which is not required for the former unknowns. E.g., the proposed solver requiring one region correspondence and one pair of lines returning the vanishing line $\mathbf{l}$, the vanishing points $\mathbf{u}_i$ and the division model parameter λ of lens distortion, and next extracting the focal length and rectifying camera rotation, is denoted $\mathtt{H}^{2}_{2}\mathbf{l}\mathbf{u}\lambda\text{-}\mathtt{fR}$.

3.2 Camera Model

A camera's purpose is to capture rays of light reflected from scene objects to form an image of the scene. Images are formed by projecting points in the scene to points in the image plane. A general camera forming an image is given by

$$\left(x, y\right)^{\top} = \mathbf{h}(\left(X, Y, Z\right)^{\top}, \mathbf{z}), \tag{3.1}$$

Term	Description
$\mathtt{P}$	3×3 camera matrix viewing $z = 0$ (see (3.9))
$\mathbf{c}$	principal point of the camera
$\mathtt{f}$	focal length of the camera
$\mathtt{K}$	camera intrinsics matrix (see (3.6))
$\mathtt{R}$	camera rotation rectifying orthogonal scene planes
λ	division model parameter for undistortion (see Sec. 3.4)
$f(\cdot, \lambda)$	undistortion function under the division model
$f^d(\cdot, \lambda)$	distortion function under the division model
Π, π	the scene plane and image plane (in $\mathbb{RP}^2$)
$\mathbf{X}$	homogeneous scene point in $\mathbb{RP}^2$
$\mathbf{x}, \tilde{\mathbf{x}}$	homogeneous pinhole and distorted image point
$\underline{\mathbf{x}}$	affine-rectified point (see (3.14))
$\mathbf{x} \leftrightarrow \mathbf{x}'$	$\mathbf{x}, \mathbf{x}'$ are in correspondence with a conjugate translation
$\tilde{\mathcal{R}}, \mathcal{R}, \underline{\mathcal{R}}$	distorted, undistorted, and affine-rectified regions
$\mathbf{M}$	homogeneous scene line
$\mathbf{m}$	homogeneous line as imaged by pinhole camera
$\tilde{\mathbf{m}}$	circle corresponding to a radially-distorted (under the division model) image line (see Sec. 3.4)
$\tilde{\mathbf{n}}$	normal of the tangency to the circle $\tilde{\mathbf{m}}$ at some point $\tilde{\mathbf{x}}$
$\mathbf{n}, \mathbf{t}$	undistorted normal of the tangency and the line defined by this normal (see Sec. 4.1.3)
$\mathbf{t}$	(in context of LAFs) a join of undistorted point correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ (see Sec. 4.1.2)
$\mathbf{m} \leftrightarrow \mathbf{m}'$	$\mathbf{m}, \mathbf{m}'$ are from the same pencil of lines
$\tilde{\mathbf{m}} \leftrightarrow \tilde{\mathbf{m}}'$	$\tilde{\mathbf{m}}, \tilde{\mathbf{m}}'$ are from the same pencil of circles or LCC
$\mathbf{U}$,$\mathbf{V}$,$\mathbf{W}$	translational directions on the scene plane and/or directions of the parallel scene plane lines
$\mathbf{u}$,$\mathbf{v}$,$\mathbf{w}$	vanishing points of directions $\mathbf{U}$,$\mathbf{V}$,$\mathbf{W}$
$\mathbf{l}_\infty$	the line at infinity
$\mathbf{l}, \tilde{\mathbf{l}}$	vanishing line and distorted vanishing line
$\mathtt{H}$	affine-rectifying homography
$\mathtt{H}_\mathbf{u}$	conjugate translation in the imaged trans. direction $\mathbf{u}$
$[\cdot]_\times$	skew-symmetric operator for computing cross products

TABLE 3.1: *Common Denotations.* Derivations are in the real projective plane $\mathbb{RP}^2$.

where $\mathbf{h}$ is a vector-valued function defining image capture, vector $\mathbf{z}$ parameterizes the camera, $(X, Y, Z)^\top$ is a scene point and $(x, y)^\top$ is its projection in the image plane by the camera $\mathbf{h}(\cdot)$.

The pinhole camera, also called the camera obscura, is the simplest

camera model. Image formation by a pinhole camera is a composition of central projection through the pinhole onto the image plane followed by a homography that changes the basis to the image coordinate system implicit to the camera's sensor [21].

Here and further, scene and image points will be modeled with homogeneous coordinates. This enables many geometric transformations, *e.g.* perspective projections to be modeled as linear transformations, which simplifies the algebraic representation of the camera.

3.2.1 Perspective Projection

The perspective projection of a 3D point $(X, Y, Z)^\top$ to a 2D point on the image plane $(u, v)^\top$ that is distance $\mathtt{f}$ from the center of projection is given by the perspective projection equation

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

where $(X, Y, Z)^\top$ is the Euclidean representation of a scene point [21].

Perspective projection as defined in (3.2.1) is non-linear but the imaging transformations can be modeled with a linear transformation by representing scene points as homogeneous 4-vectors and image points as homogeneous 3-vectors [21]. Using the homogeneous representation, perspective projection becomes

$$\alpha \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \operatorname{diag}(\mathtt{f}, \mathtt{f}, 1) \left[\mathtt{I}_3 \mid \mathbf{0} \right] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} , \tag{3.2}$$

where $\alpha = 1/z$.

3.2.2 Camera Coordinate System

A scene point $(X, Y, Z, 1)^\top$ is put into the camera's coordinate system by a change of basis given by a transformation defining a rigid transform in Euclidean space

$$\begin{bmatrix} \mathtt{R} & \mathtt{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} , \tag{3.3}$$

where $\mathtt{R} \in \mathbb{R}^{3\times3}$ is a rotation matrix (equivalently an orthonormal matrix), $\mathtt{t} \in \mathbb{R}^3$ is a translation, and $\mathtt{t}_0 = -\mathtt{R}^\top \mathtt{t}$ gives the Euclidean coordinates of the camera's projection center in the scene coordinate system.

3.2.3 Image Coordinate System

Projected points are put into the image coordinate system by applying a homography that encodes the geometry of the camera's sensor. For real cameras, the homography is upper triangular

$$\begin{bmatrix} a_x & a_x \cot\theta & c_x \\ 0 & a_y / \sin\theta & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.4}$$

where a_x and a_y are the scale factors of the image plane in units of pixels/mm, $\mathbf{c} = \left(c_x, c_y\right)^\top$ is the principal point or optical center of the camera in pixels, and θ is the skew of the sensor. For a typical CCD camera with ORUA, the simplifications $\mathtt{f}_x = \mathtt{f}_y = \mathtt{f}$ and $\theta = \pi/2$ can be assumed. For a pinhole camera, in addition to these typical constraints, we have $a_x = a_y = 1$ [21].

The convention is to denote the intrinsics matrix as $\mathtt{K}$ and incorporate the scaling due to the focal length (see (3.2)),

$$\mathtt{K} = \begin{bmatrix} a_x & a_x \cot\theta & c_x \\ 0 & a_y / \sin\theta & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathtt{f} & 0 & 0 \\ 0 & \mathtt{f} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathtt{f}_x & k_c & c_x \\ 0 & \mathtt{f}_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.5}$$

Due to the assumptions made in this work (see Table 2.1) the camera intrinsics matrix $\mathtt{K}$ is assumed to have the following form:

$$\mathtt{K} = \begin{bmatrix} \mathtt{f} & 0 & c_x \\ 0 & \mathtt{f} & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.6}$$

3.2.4 Camera Matrix

Positioning and orienting the camera, projection, and the imaging transformation can be composed into a linear operation given by 3×4 camera matrix [21]

$$\mathtt{P}^{3\times4} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix} = \mathtt{K} \begin{bmatrix} \mathtt{I}_3 \mid \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \mathtt{K} \begin{bmatrix} \mathtt{R} \mid \mathbf{t} \end{bmatrix} \tag{3.7}$$

Columns $\mathbf{p}_j$ have geometric meaning. Columns $\mathbf{p}_j$ where $j \in \{1 \ldots 3\}$ are the vanishing points of the axes of the scene coordinate system and $\mathbf{p}_4$ is the image of the scene origin. The column representation of $\mathtt{P}^{3\times4}$ will play an important role in modeling cameras viewing scene planes, as will be seen in Sec. 3.2.5.

Then the imaging of a scene point by the camera $\mathtt{P}^{3\times4}$ is given as

$$\alpha \left(x, y, 1\right)^{\top} = \mathtt{P}^{3\times4} \left(X, Y, Z, 1\right)^{\top}, \tag{3.8}$$

where $\alpha = 1/Z$.

3.2.5 Camera Viewing a Scene Plane

Without loss of generality, coplanar scene points $\{ \mathbf{X}_i \}$ are assumed to be on the scene plane $z = 0$ (see Fig. 3.1). This permits the camera matrix $\mathtt{P}$ to be modeled as the homography that changes the basis from the scene-plane coordinate system to the camera's image-plane coordinate system in the real-projective plane $\mathbb{RP}^2$ [21],

$$\alpha \underbrace{\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}}_{\mathbf{x}} \underbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix}}_{\mathtt{P}^{3\times4}} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix}}_{\mathtt{P}} \underbrace{\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}}_{\mathbf{X}}, \tag{3.9}$$

where $\mathbf{p}_j = \left(p_{1j}, p_{2j}, p_{3j}\right)^{\top}$ encode the intrinsics and extrinsics of the camera matrix $\mathtt{P}^{3\times4}$. The scene and image planes are denoted Π and π, respectively. Imaged points are denoted $\mathbf{x} = \left(x, y, 1\right)^{\top}$, where x, y are the image coordinates.

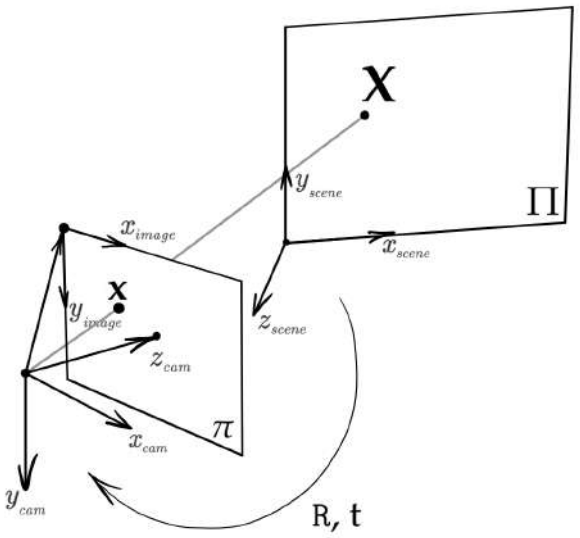


FIGURE 3.1: *Camera Viewing a Scene Plane.* Coplanar scene points $\{ \mathbf{X}_i \}$ are assumed to be on the scene plane $z = 0$. The assumption implies that a homography can transform a point $\mathbf{X}$ on the scene plane in the world coordinate system to an image point $\mathbf{x}$ in the camera's image-plane coordinate system.

3.2.6 Homography Decomposition

The camera $\mathtt{P}$ can be uniquely decomposed into a similarity $\mathtt{S}$ affinity $\mathtt{A}$ and projectivity $\mathtt{H}$

$$\mathtt{P} = \underbrace{\begin{bmatrix} s\mathtt{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{\mathtt{S}} \underbrace{\begin{bmatrix} \mathtt{A}_{2\times 2} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{\mathtt{A}} \underbrace{\begin{bmatrix} \mathtt{I}_{2\times 2} & \mathbf{0} \\ l_1 \;\; l_2 & l_3 \end{bmatrix}}_{\mathtt{H}},$$

where $l_3 \neq 0$, s is non-zero scalar, $\mathtt{R}$ is a rotation, $\mathbf{t}$ is a translation, $\mathtt{A}_{2\times 2}$ is an upper-triangular matrix specifying the anisotropic scaling and skew components such that $\det \mathtt{A}_{2\times 2} = 1$, and the projective components are specified by $\left(l_1, l_2, l_3\right)^\top$, where $l_3 \neq 0$ [21].

Note that since a homography is invertible, (3.2.6) implies that $\mathtt{P}$ can be decomposed as the inverses of a similarity $\mathtt{S}'$, affinity $\mathtt{A}'$ and projectivity $\mathtt{H}'$ as $\mathtt{P} = \mathtt{H}'^{-1}\mathtt{A}'^{-1}\mathtt{S}'^{-1}$.

3.2.7 Rectification

The pre-imaging homography $\mathtt{P}^{-1}$ can be decomposed into a similarity $\mathtt{S}$, affinity $\mathtt{A}$ and projectivity $\mathtt{H}$ as $\mathtt{P}^{-1} = \mathtt{SAH}$ (see (3.2.6)). Metric rectification is invariant to similarity transformations [21]. Thus $\mathtt{AH} = \mathtt{S}^{-1}\mathtt{P}^{-1}$ is metric rectifying. Since The pre-imaging transform $\mathtt{P}^{-1}$ is homogeneous, it has 8 degrees of freedom, 4 of which are eliminated by multiplying it with the similarity $\mathtt{S}^{-1}$. This leaves 4 degrees of freedom for the metric rectifying homography $\mathtt{AH}_\infty$.

3.3 Affine Rectification

We denote the image of the scene plane's vanishing line by $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$. Assuming $l_3 \neq 0$, a projective transformation $\mathtt{H}$ mapping $\mathbf{l}$ back to a line at infinity $\mathbf{l}_\infty = \left(0, 0, 1\right)^\top$ is called an affine-rectifying homography [21]. The transformation of an image point $\mathbf{x}$ to an affine-rectified point $\underline{\mathbf{x}}$ has the form

$$\beta\underline{\mathbf{x}} = \mathtt{H}(\mathbf{l})\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ & \mathbf{l}^\top & \end{bmatrix} \mathbf{x}, \quad \beta \neq 0. \tag{3.10}$$

3.4 Radial Lens Distortion

In this work, we use a known one-parameter division model [20] for lens undistortion

$$\gamma \mathbf{x} = f(\tilde{\mathbf{x}}, \lambda) = \left(\tilde{x}, \tilde{y}, 1 + \lambda(\tilde{x}^2 + \tilde{y}^2)\right)^\top, \tag{3.11}$$

where $\tilde{\mathbf{x}}$ is a distorted homogeneous point that is mapped to a pinhole homogeneous point $\mathbf{x}$.

By substituting (3.13) into (3.10) we get an undistorted affine-rectified point

$$\alpha \underline{\mathbf{x}} = \mathtt{H}(\mathbf{l}) f(\tilde{\mathbf{x}}, \lambda) = \left(\tilde{x}, \tilde{y}, l_1\tilde{x} + l_2\tilde{y} + l_3(1 + \lambda(\tilde{x}^2 + \tilde{y}^2))\right)^\top. \tag{3.12}$$

Affine rectification as given in (3.10) is valid only if $\mathbf{x}$ is imaged by a pinhole camera. Cameras always have some lens distortion, and the distortion can be significant for wide-angle lenses. For a lens distorted point, denoted $\tilde{\mathbf{x}}$, an undistortion function f is needed to transform $\tilde{\mathbf{x}}$ to the pinhole point $\mathbf{x}$. We use the one-parameter division model to parameterize the radial lens undistortion function Sec. 3.4,

$$\gamma \mathbf{x} = f(\tilde{\mathbf{x}}, \lambda) = \left(\tilde{x}, \tilde{y}, 1 + \lambda(\tilde{x}^2 + \tilde{y}^2)\right)^\top \tag{3.13}$$

where $\tilde{\mathbf{x}} = \left(\tilde{x}, \tilde{y}, 1\right)^\top$ is a feature point with the distortion center subtracted.

The strengths of this model were shown by Fitzgibbon [20] for the joint estimation of two-view geometry and non-linear lens distortion. The division model is especially suited for minimal solvers since it is able to express a wide range of distortions (*e.g.*, see Fig. 5.6) with a single parameter (denoted λ), as well as yielding simpler equations compared to other distortion models.

For the remainder of the derivations, we assume that the image center and distortion center are coincident and that $\tilde{\mathbf{x}}$ is a distortion-center subtracted point. While this may seem like a strong assumption, Willson et al. [42] and Fitzgibbon [20] showed that the precise positioning of the distortion center does not strongly affect image correction. No constraints are placed on the location of the principal point of the camera to estimate radial division model parameter and vanishing line, however, we assume it to auto-calibrate camera from two vanishing points restored by the proposed minimal solvers (see Chapter 4).

3.5 Rectification of Radially-Distorted Points

Affine rectified points $\underline{\mathbf{x}}_i$ can be expressed in terms of distorted points $\tilde{\mathbf{x}}_i$ by substituting (3.13) into (3.10), which gives

$$\begin{aligned} \alpha\underline{\mathbf{x}} = \left(\alpha\underline{x}, \alpha\underline{y}, \alpha\right)^\top = \mathtt{H}(\mathbf{l}) f(\tilde{\mathbf{x}}, \lambda) = \\ \left(\tilde{x}, \tilde{y}, l_1\tilde{x} + l_2\tilde{y} + l_3(1 + \lambda(\tilde{x}^2 + \tilde{y}^2))\right)^\top. \end{aligned} \tag{3.14}$$

The rectifying function $\mathtt{H}(\mathbf{l}) f(\tilde{\mathbf{x}}, \lambda)$ in (3.14) also acts radially about the distortion center, but unlike the division model in (3.13), it is not rotationally symmetric.

The distortion function of the lens as parameterized by the division model is denoted $f^d(\cdot, \lambda)$. Under the division model, the radially-distorted image of the vanishing line is a circle and is denoted $\tilde{\mathbf{l}}$ [8, 20, 36, 39].

3.6 Camera Auto-Calibration

Given a camera, we may or may not have access to its intrinsic parameters. But having access to the images taken by a camera, one can recover camera calibration information. A single-view camera auto-calibration is a process of estimating the parameters of image formation from the properties of the observed scene. A camera matrix $\mathtt{K}$ can be restored up to two signs [27] from the matrix:

$$\omega = \mathtt{K}^{-\top}\mathtt{K}^{-1}, \tag{3.15}$$

which is called an *Image of an Absolute Conic* (IAC).

However, ω is unknown and the proposed minimal solvers make assumptions on it structure, so computing the Cholesky decomposition of ω is not an option. Matrix ω is a 3×3 symmetric matrix and thus it has only six independent elements ω_{11}, ω_{12}, ω_{13}, ω_{22}, ω_{23} and ω_{33}. Since the intrinsics matrix have the form of (3.6), an IAC has the following form:

$$\begin{aligned} \omega = \begin{bmatrix} 1/\mathtt{f} & 0 & 0 \\ 0 & 1/\mathtt{f} & 0 \\ -c_x/f & -c_y/f & 1 \end{bmatrix} \begin{bmatrix} 1/\mathtt{f} & 0 & -c_x/f \\ 0 & 1/\mathtt{f} & -c_y/f \\ 0 & 0 & 1 \end{bmatrix} = \\ = 1/\mathtt{f}^2 \begin{bmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ -c_x & -c_y & \mathtt{f}^2 + c_x^2 + c_y^2 \end{bmatrix} \sim \begin{bmatrix} \omega_{11} & 0 & \omega_{13} \\ 0 & \omega_{11} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{bmatrix} \end{aligned} \tag{3.16}$$

The following features configurations can be used to place constraints on the the unknowns of ω as defined in (3.16):

1. orthogonal vanishing points (one constraint for a pair of vanishing points)
2. orthogonal vanishing lines (one constraint for a pair of vanishing lines)
3. vanishing points orthogonal to vanishing lines (two constraints for a vanishing point-vanishing line pair)
4. known principal point (two constraints)

This constraints listed above are not exhaustive. The proposed minimal solvers use the first and fourth constraints to auto-calibrate the camera.

Chapter 4

Proposed Minimal Solvers

This chapter introduces the joint undistorting and rectifying hybrid minimal solver that admits a radially-distorted conjugately-translated covariant region correspondence and a correspondence of circular arcs whose preimages are parallel lines. The solver is denoted $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$. The motivation for introducing the $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ solver is to extend robust rectification to distorted images of scenes where neither texture nor scene lines dominate. In this case, robustly and accurately solving the scene may require sampling from both feature types.

In addition, we introduce a novel joint undistortion and rectifying solver that admits three correspondences of circular arcs, which we denote $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$. The state-of-the-art does not offer a joint undistorting and rectifying solver that admits arc correspondences. Rather, the state-of-the-art requires the corresponded sets of lines, either three for the method of Wildenauer et al. [41] in addition to an arc correspondence or corresponded sets of four and three lines for Antunes et al. [3]. The proposed $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$ method admits the possibility of discovering a third translation direction on the scene plane.

4.1 A Unified Approach

The proposed solvers jointly estimate the division model parameter λ and the vanishing line of an imaged scene plane $\mathbf{l}$. Recall that the recovery of $\mathbf{l}$ is sufficient for affine rectification. Each of the proposed solvers uses the constraints that the meet of imaged parallel scene lines $\mathbf{t}$,$\mathbf{t}'$ is a vanishing point, namely,

$$\mathbf{u} = \mathbf{t} \times \mathbf{t}'. \tag{4.1}$$

and that the vanishing point $\mathbf{u}$ and vanishing line $\mathbf{l}$ are coincident,

$$\mathbf{u}^\top \mathbf{l} = 0. \tag{4.2}$$

The derivations of $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ and $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$ use the same procedure for generating scalar constraint equations on $\mathbf{l}$ by substituting vanishing points constructed from the undistorted measurements into (4.2).

The solver derivations differ only by how lines $\mathbf{t}$ and $\mathbf{t}'$ are constructed from distorted image measurements, *i.e.*, from radially-distorted conjugately translated points or circular arcs whose preimages are parallel scene lines. Since lines $\mathbf{t}$ and $\mathbf{t}'$ are constructed from undistorted measurements, they are functions of λ as well, which is made explicit going forward. Thus the vanishing point $\mathbf{u}$ is determined by polynomials of a certain degree in λ in each coordinate, namely,

$$\mathbf{u}(\lambda) = \mathbf{t}(\lambda) \times \mathbf{t}'(\lambda) = \begin{pmatrix} u_1(\lambda) \in P_1 \\ u_2(\lambda) \in P_1 \\ u_3(\lambda) \in P_2 \end{pmatrix}, \tag{4.3}$$

where P_k is the vector space of polynomials of degree less than or equal to k.

Using (4.2), vanishing points $\mathbf{u}_i(\lambda)$ can be used to generate constraint equations on $\mathbf{l}$ and λ. There are four unknowns to be recovered, namely $\mathbf{l} = \begin{pmatrix} l_1, l_2, l_3 \end{pmatrix}^\top$ and the division model parameter λ (see Sec. 3.4). The vanishing line $\mathbf{l}$ is homogeneous, so it has only two degrees of freedom. Thus three scalar constraint equations (two of which are independent) of the form (4.2) generated by three vanishing points $\mathbf{u}_1(\lambda)$,$\mathbf{u}_2(\lambda)$, and $\mathbf{u}_3(\lambda)$ are needed.

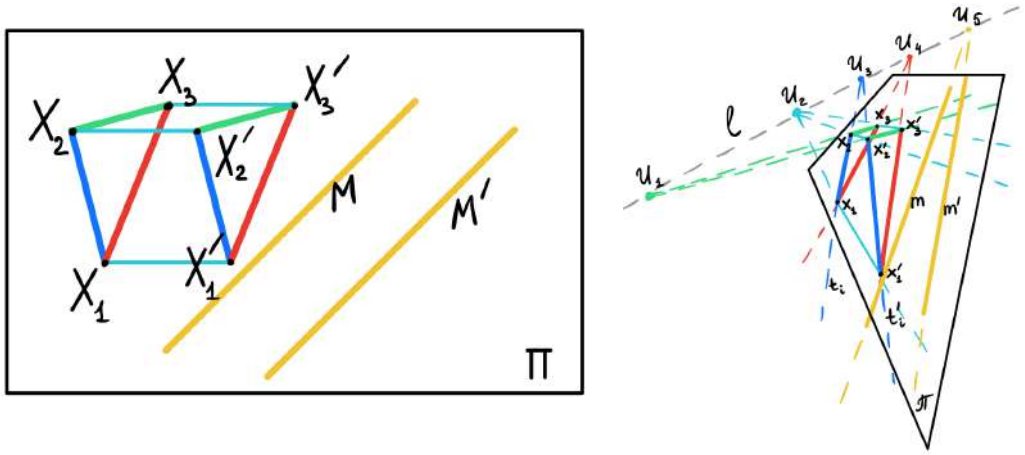
The constraints can be concisely written as a homogeneous matrix-vector equation by introducing $\mathtt{M}(\lambda)$, which is constructed by stacking vanishing points $\mathbf{u}_i(\lambda)$ row-wise such that

$$\mathtt{M}(\lambda)\mathbf{l} = \begin{bmatrix} \mathbf{u}_1^\top(\lambda) \\ \mathbf{u}_2^\top(\lambda) \\ \mathbf{u}_3^\top(\lambda) \end{bmatrix} \mathbf{l} = \mathbf{0}. \tag{4.4}$$

Matrix $\mathtt{M}(\lambda)$ is singular, which generates the additionally needed scalar constraint equation $\det \mathtt{M}(\lambda) = 0$. The determinant constraint defines a univariate quartic with unknown λ, which is solved in closed form. After recovering of λ, the null space of $\mathtt{M}$ is found, which gives the vanishing line $\mathbf{l}$.

This approach gives a unified procedure for generating the $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ and $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$ solvers. Sec. 4.1.2 and 4.1.3 detail how the construction of the vanishing points $\mathbf{u}_i(\lambda)$ differ based on the feature configurations used to construct the lines $\mathbf{t}_i(\lambda), \mathbf{t}_i'(\lambda)$.

Minimal set of features for $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$



Minimal set of features for $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$

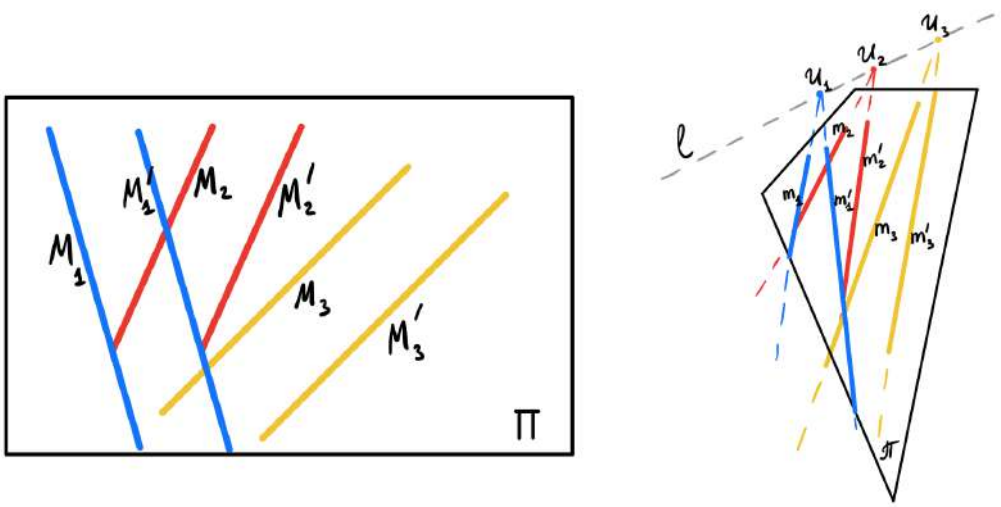


FIGURE 4.1: *Geometry of Proposed Solvers* $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ *and* $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$. Minimal sample sets of features on the scene plane required by the proposed solvers. (left) is a fronto-parallel view of the scene plane, whereas (right) illustrates an undistorted camera view of the scene plane along with a vanishing line and vanishing points arising from the perspective view. (top) The $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ solver requires one radially-distorted conjugately-translated affine-covariant region correspondence and one correspondence of distorted lines. The features provide exactly five scalar constraint equations. Two equations are needed to estimate $\mathbf{l}$ and three are necessary to jointly estimate $\mathbf{l}$ and λ. (bottom) The $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$ solver requires three correspondences of distorted lines. The features form exactly three scalar constraint equations. A degenerate case is possible when two vanishing points coincide (e.g., $\mathbf{u}_1$ and $\mathbf{u}_2$) because all the four lines are in correspondence. This is addressed by incorporating an additional constraint equation (see Sec. 4.1).

4.1.1 Two Coplanar Vanishing Points

If the line pairs $(\mathbf{t}_i, \mathbf{t}'_i)$ constructed from the minimal sample set supplied to a solver do not meet at three distinct vanishing points, then there are not enough independent constraints to recover the undistortion parameter λ and vanishing line $\mathbf{l}$. An example of such a degeneracy is provided by at least two of three correspondences of parallel lines meeting at the same vanishing point. The necessary number of constraints for this configuration is given by adding the constraint that two line correspondences are coincident with the same vanishing point

$$\begin{aligned}\mathbf{u}_i \times \mathbf{u}_j &= \begin{bmatrix} 0 & -u_{3i} & u_{2i} \\ u_{3i} & 0 & -u_{1i} \\ -u_{2i} & u_{1i} & 0 \end{bmatrix} \begin{pmatrix} u_{1j} \\ u_{2j} \\ u_{3j} \end{pmatrix} \\ &= \begin{pmatrix} k_{13}\lambda^3 + k_{12}\lambda^2 + k_{11}\lambda + k_{10} \in P_3 \\ k_{23}\lambda^3 + k_{22}\lambda^2 + k_{21}\lambda + k_{20} \in P_3 \\ k_{33}\lambda^3 + k_{32}\lambda^2 + k_{31}\lambda + k_{30} \in P_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.\end{aligned} \tag{4.5}$$

It is not known apriori if the inputted minimal sample contains 3 distinct vanishing points. A test is proposed in Sec. 4.4 that can detect the configuration by using the remaining unused constraints provided by the feature configuration.

4.1.2 Radially-Distorted Conjugate Translations

Referring to Fig. 4.1, it can be seen that four vanishing points $\{ \mathbf{u}_i \}_{i=1}^4$ can be constructed from just one radially-distorted conjugately-translated affine-covariant region correspondence [31]. The geometry of an affine-covariant region $\mathcal{R}$ is described by a *local affine frame* (LAF) which is a right-handed affine basis in the image coordinate system. LAF is minimally parameterized by three points $\{ \mathbf{o}, \mathbf{x}, \mathbf{y} \}$ (see also [38, 25, 24, 26]). The vanishing points are constructed from each meet of joins of pairs of conjugate-translations that share the same translation direction in the scene plane, which are color-coded in red, green, blue, and cyan in Fig. 4.1. There are six such meets to choose from, three for $\mathbf{u}_1$ and one for each of $\mathbf{u}_2$,$\mathbf{u}_3$ and $\mathbf{u}_4$.

Joins $\mathbf{t}$ and $\mathbf{t}'$ are constructed from either the intra-region conjugate translations, which are red, green, and blue in Fig. 4.1 or inter-region conjugate translations, which are cyan. Without loss of generality, we choose the cyan direction,

$$\mathbf{t} = \mathbf{x}_i \times \mathbf{x}'_i \qquad \mathbf{t}' = \mathbf{x}_j \times \mathbf{x}'_j, \quad i \neq j. \tag{4.6}$$

Solver	Input Set	Assumptions	# sol. $(\lambda, \mathbf{l})$
$\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ (see Sec. 4.1.2)	2 LAFs + 2 arcs	2 LAFs in cspond with $\mathtt{H}_\mathbf{u}$, 2 arcs in cspond with $\mathbf{v}$	4
$\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$ (see Sec. 4.1.1)	2 LAFs + 2 arcs	2 LAFs in cspond with $\mathtt{H}_\mathbf{u}$, 2 arcs in cspond with an inter ($\mathtt{H}_\mathbf{u}$) or intra-rgn conjugate translation	3
$\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ (see Sec. 4.1.3)	6 arcs	2 arcs in cspond with $\mathbf{u}$, 2 arcs in cspond with $\mathbf{v}$, 2 arcs in cspond with $\mathbf{w}$	4
$\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$ (see Sec. 4.1.1)	6 arcs	2 arcs in cspond with $\mathbf{u}$, 4 arcs in cspond with $\mathbf{v}$	3

TABLE 4.1: *Minimal Sample Set of Features.* The assumptions for LAFs and arcs forming a minimal sample set, which is to be drawn for $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solvers. Denoted by $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$ and $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$ are the solvers variants assuming the degeneracy described in Sec. 4.1.1. Further by $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ an optimal solver that chooses the best solution from $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$ will be considered (the same goes for $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$). Three scalar constraint equations formed by the features in MSS are needed for each solver to generate four (for $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$) or three (for $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$ and $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}^*$) solutions of $(\lambda, \mathbf{l})$.

Under the division model, the equation for a join $\mathbf{t}$ in the undistorted space becomes

$$\begin{aligned}\mathbf{t}(\lambda) &= f(\tilde{\mathbf{x}}_i, \lambda) \times f(\tilde{\mathbf{x}}_j, \lambda) = [f(\tilde{\mathbf{x}}_i, \lambda)]_\times f(\tilde{\mathbf{x}}_j, \lambda) \\ &= \begin{bmatrix} 0 & -1-\lambda\tilde{r}_i^2 & \tilde{y}_i \\ 1+\lambda\tilde{r}_i^2 & 0 & -\tilde{x}_i \\ -\tilde{y}_i & \tilde{x}_i & 0 \end{bmatrix} \begin{pmatrix} \tilde{x}_j \\ \tilde{y}_j \\ 1+\lambda\tilde{r}_j^2 \end{pmatrix} \\ &= \begin{pmatrix} k_{11}\lambda + k_{10} \in P_1 \\ k_{21}\lambda + k_{20} \in P_1 \\ k_{30} \in P_0 \end{pmatrix},\end{aligned} \tag{4.7}$$

where $[\cdot]_\times$ is the skew-symmetric operator, k_{ij} are the coefficients of the linear equations in λ occurring in the join. The joins $\mathbf{t}(\lambda)$ and $\mathbf{t}'(\lambda)$

are used in (4.1), where the procedure outlined in Sec. 4.1 is used to generate two scalar constraints on $\mathbf{l}$ and λ.

4.1.3 Distorted Parallel Scene Lines

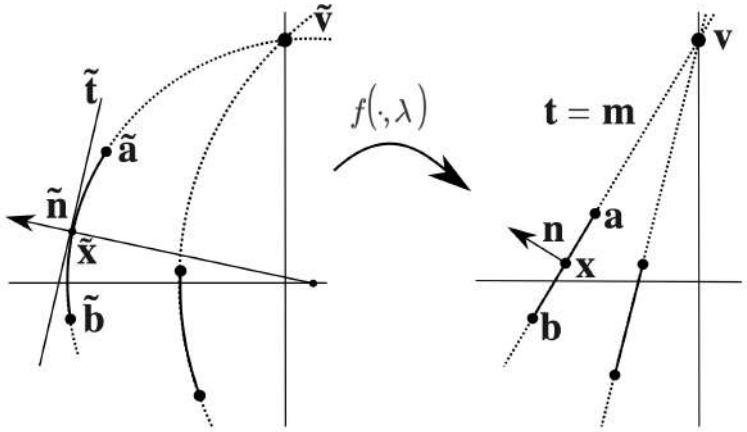


FIGURE 4.2: *Geometry of a Distorted Line.* Barreto et al. [5] showed that under the division model of lens distortion, a straight line is imaged as a circle. When undistorted, a tangent $\mathbf{t}$ to the detected circular arc is transformed such that it coincides with the original straight line [41] in the undistorted image space that passes through the vanishing point under a perspective view of a scene. The figure is adapted from [41]

Wildenauer et al. [41] used elementary differential geometry to derive an expression for how normals to tangents of a circle transform with respect to $f(\cdot,\lambda)$, the division model for undistortion. Barreto et al. [5] showed that under the division model of lens distortion, a straight line is imaged as a circle. Thus parallel scene lines distorted by the division model will be imaged as circles intersecting at a distorted vanishing point. Using this relation, Wildenauer observed that the undistorted normal to the tangent line of a distorted scene line $\tilde{\mathbf{m}}$ (equivalently circle) defines a line $\mathbf{t}$ that is collinear with the scene line $\mathbf{m}$ in undistorted space and thus coincident with its vanishing point (see Fig. 4.2). The expression for $\mathbf{t}$ is derived in terms of the unknown division model parameter λ and the normal of the tangency to the circle $\tilde{\mathbf{n}} = \left(\tilde{a},\tilde{b}\right)^\top$ at point $\tilde{\mathbf{x}} = \left(\tilde{x},\tilde{y}\right)^\top$ estimated from the measurements,

$$\mathbf{t}(\lambda) = \begin{pmatrix} \tilde{a} \\ \tilde{b} \\ -\tilde{\mathbf{n}}^\top\tilde{\mathbf{x}} \end{pmatrix} + \lambda \begin{pmatrix} \tilde{a}\tilde{x}^2 + 2\tilde{b}\tilde{x}\tilde{y} - \tilde{a}\tilde{y}^2 \\ \tilde{b}\tilde{y}^2 + 2\tilde{a}\tilde{x}\tilde{y} - \tilde{b}\tilde{x}^2 \\ 0 \end{pmatrix}, \tag{4.8}$$

Therefore, $\mathbf{t}$ is decomposed as the sum of the λ-independent term $\mathbf{t}_0$ and $\lambda\mathbf{t}_1$ term. Since $\mathbf{t}_1 \times \mathbf{t}_1' = \alpha \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\top$ for $\alpha \neq 0$, we conclude that

$$\mathbf{u}(\lambda) = \mathbf{t} \times \mathbf{t}' = \begin{pmatrix} u_1(\lambda) \in P_1 \\ u_2(\lambda) \in P_1 \\ u_3(\lambda) \in P_2 \end{pmatrix}. \tag{4.9}$$

Again, by substituting (4.9) into (4.2), we create an independent scalar constraint equation.

4.1.4 Constructing the Solvers

By using the unified approach detailed in Sec. 4.1, constraints induced by radially-distorted conjugately-translated affine-covariant regions in (4.7) can be included with constraints induced by distorted scene lines in (4.8) into (4.4) to generate the solvers.

The $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ is simply constructed by stacking two constraints of the form (4.7) with one constraint of the form (4.8). The $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solver requires one radially-distorted conjugately-translated affine-covariant region correspondence and one arc correspondence.

The $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ is constructed by using three constraints of the form (4.8). The $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solver requires three correspondences of distorted lines.

4.2 Vanishing Point Estimation

Having vanishing line $\mathbf{l}$ and the division model parameter λ recovered by either $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ or $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$, vanishing points from the minimal sample set that were not used in the construction of the minimal constraints can be recovered. There are a total of five vanishing points that can be estimated from the input feature set required by $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and there are a total of three (or two for the degenerate case) vanishing points that can be estimated from the input feature set required by $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$.

The relations between region correspondences or line correspondences and the vanishing point $\mathbf{u}$, and the enforcement of the vanishing point-vanishing line incidence constraint $\mathbf{u}^\top\mathbf{l} = 0$ are encoded in the constrained least squares problem,

$$\begin{aligned} &\underset{\mathbf{u}}{\text{minimize}} \ \left\|\mathtt{M}\mathbf{u} - \mathbf{y}\right\|^2 \\ &\text{subject to } \mathtt{C}\mathbf{u} = \mathbf{d}. \end{aligned} \tag{4.10}$$

For the translation direction of $\{ \mathbf{x}_i \leftrightarrow \mathbf{x}'_i \}_i$, (4.10) takes the form that is formulated in [31]:

$$\mathtt{M} = \begin{bmatrix} & \vdots & \\ -\mathbf{l}^\top \mathbf{x}_i & 0 & x'_i(\mathbf{l}^T \mathbf{x}_i) \\ 0 & \mathbf{l}^\top \mathbf{x}_i & y'_i(\mathbf{l}^\top \mathbf{x}_i) \\ & \vdots & \end{bmatrix}, \tag{4.11}$$
$$\mathbf{y} = (\dots, x_i - x'_i, y_i - y'_i, \dots)^\top, \mathtt{C} = \mathbf{l}^\top, \mathbf{d} = \mathbf{0}.$$

For the corresponded lines of $\{ \mathbf{m}_i \leftrightarrow \mathbf{m}'_i \}_i$, which are gotten from the arcs that are imaged lines, the constraint becomes

$$\begin{aligned} \mathtt{M} &= \begin{bmatrix} \mathbf{m}^\top \\ \mathbf{m}'^\top \end{bmatrix}, \quad \mathbf{y} = \mathbf{0}, \\ \mathtt{C} &= \begin{bmatrix} & \mathbf{l}^\top & \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{d} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}. \end{aligned} \tag{4.12}$$

For both types of features, the matrices $\begin{bmatrix} \mathtt{M}^\top & \mathtt{C}^\top \end{bmatrix}^\top$ have linearly independent columns, and $\mathtt{C}$ is row independent. Thus $\mathbf{u}$ is recovered by solving

$$\begin{bmatrix} \mathtt{M}^\top \mathtt{M} & \mathbf{l} \\ \mathbf{l}^\top & 0 \end{bmatrix} \begin{pmatrix} \mathbf{u} \\ z \end{pmatrix} = \begin{pmatrix} \mathtt{M}^\top \mathbf{y} \\ 0 \end{pmatrix}, \tag{4.13}$$

where z is the Lagrangian multiplier [7].

4.3 Auto-Calibration from Vanishing Points

The recovered vanishing points are used for auto-calibration. A Manhattan scene is assumed, which means that the vanishing points correspond to orthogonal directions in the scene. The assumption enables the recovery of the focal length and conjugate rotation of the camera. We also assume zero skew and square pixels for the CCD. Denoting the two orthogonal vanishing points as $\mathbf{u}$, $\mathbf{v}$, and setting the principal point to the image center $\mathbf{c}$, the focal length is computed as follows

$$\mathtt{f} = \sqrt{-\mathring{u}_x \mathring{v}_x - \mathring{u}_y \mathring{v}_y}, \tag{4.14}$$

where $\mathring{\mathbf{u}} = \mathbf{u} - \mathbf{c}, \mathring{\mathbf{v}} = \mathbf{v} - \mathbf{c}$. The camera intrinsics matrix is constructed using (3.6).

The focal length is used to recover the rotation $\mathtt{R}$ of the camera with respect to the scene plane. $\mathtt{R}$ is a rotation matrix whose columns are orthogonal unit vectors of vanishing points pre-imaged to calibrated

ray space by $\mathtt{K}^{-1}$

$$\mathtt{R} = \begin{bmatrix} \mathbf{U} & \mathbf{V} & \mathbf{W} \end{bmatrix}, \tag{4.15}$$

where $\mathbf{U} = \mathtt{K}^{-1}\mathbf{u}$, $\mathbf{V} = \mathtt{K}^{-1}\mathbf{v}$, $\mathbf{W} = \mathtt{K}^{-1}\mathbf{w}$, and the third vanishing point $\mathbf{w}$ is computed from $\mathbf{u}, \mathbf{v}$ to complete the orthonormal basis

$$\mathbf{w} = \mathbf{u} \times \mathbf{v}. \tag{4.16}$$

4.4 Best Minimal Solution Selection

The subset of constraints needed to construct vanishing points from the input set is referred here as *minimal configuration*. The number of minimal configurations for an input set required by $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ (a LAF correspondence and a correspondence of two circular arcs) is 12. If considering a case of two coinciding vanishing points (see Sec. 4.1.1, plus 36 minimal configurations) and also a case of constructing vanishing points from affine-covariant region correspondence only (see Eliminating Vanishing Line solver of [31], plus 10 minimal configurations) the total number is 58. An input feature set required by $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$ has only one minimal configuration. If also considering a degeneracy of two coinciding vanishing points, then there are three additional minimal configurations, so it is four in total for an input set.

Thus a minimal configuration returned by each solver is automatically chosen such that it minimizes the cost of the input sample with respect to a model. The input sample error is computed as a weighted sum (with weight $w = 0.52$ that was empirically chosen) of two terms corresponding to two types of features from the input sample

$$E = w \cdot E_{\text{XFER}} + (1 - w) \cdot E_{\text{LC}}. \tag{4.17}$$

The first term is the symmetric transfer error used to measure the accuracy of the estimated radially-distorted conjugate translation of the point correspondences,

$$E_{\text{XFER}} = \sum_i d(\tilde{\mathbf{x}}_i, f^d(\hat{\mathtt{H}}_{\mathbf{u}}^{-1} f(\tilde{\mathbf{x}}_i', \hat{\lambda}), \hat{\lambda}))^2 + d(f^d(\hat{\mathtt{H}}_{\mathbf{u}} f(\tilde{\mathbf{x}}_i, \hat{\lambda}), \hat{\lambda}), \tilde{\mathbf{x}}_i')^2, \tag{4.18}$$

where $\mathtt{H}_{\mathbf{u}}$ is a conjugate translation $\mathbf{u}$ [31]

$$\hat{\mathtt{H}}_{\mathbf{u}} = \mathtt{I}_3 + \hat{s}^{\mathbf{u}} \hat{\mathbf{u}} \hat{\mathbf{l}}^\top, \tag{4.19}$$

where the scalar $s^{\mathbf{u}}$ is the magnitude of translation in the direction $\mathbf{u}$ for the point correspondence $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ [35].

The correspondences $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ used as input to the minimal solver will have zero symmetric transfer error. The remaining constraints that

are available from the affine-covariant regions correspondence can bee used to estimate the consistency of the correspondence with the proposed model.

The second term of (4.17) is the distorted line consistency measure computed for the parallel scene line segments,

$$E_{\text{LC}} = \sum_{k} \sum_{\mathbf{x} \in A_k} d_l(\mathbf{x}, f_l^d([\bar{\mathbf{e}}_k]_\times \hat{\mathbf{u}}_k, \hat{\lambda}))^2, \tag{4.20}$$

where $\hat{\mathbf{u}}_k$ is an estimated vanishing point (see. Sec. 4.2), $\bar{\mathbf{e}}_k$ is the geometric median of undistorted points of the circular arc A_k, and $d_l(\mathbf{x}, \mathbf{C})$ is an orthogonal distance from the point $\mathbf{x}$ to a circle $\mathbf{C}$. Note that the term $[\bar{\mathbf{e}}_k]_\times \hat{\mathbf{u}}_k$ in (4.20) is the line passing through the estimated vanishing point $\hat{\mathbf{u}}_k$. The line $[\bar{\mathbf{e}}_k]_\times \hat{\mathbf{u}}_k$ approximates the maximum likelihood estimate of the line fitting the undistorted points and is computed as in [37]. However, it is preferable to minimize the error in the distorted space where the measurements come from because neglecting the distortion causes bias in the estimates [21, 18, 36]. Thus the line is back distorted to a circle $\mathbf{C}$ with the estimated division model parameter $\hat{\lambda}$ (see Fig. 4.3) and the geometric error $d_l(\mathbf{x}, \mathbf{C})$ is computed.

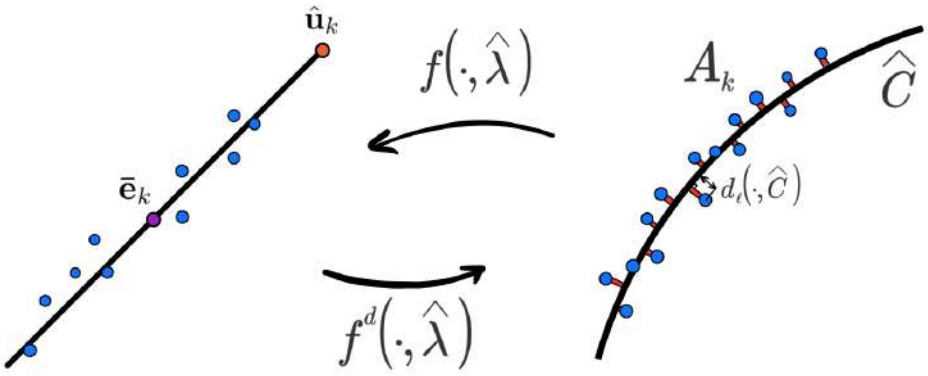


FIGURE 4.3: *Arc Consistency Measure.* An algorithm for computing a circular arc consistency with the model is similar to [37]. An line joining the geometric median of points and the estimated vanishing point $\hat{\mathbf{u}}_k$ of the scene plane line direction is calculated according to (4.20), then distorted (on the right) with an estimated radial division model parameter, and the sum of orthogonal distances to a circle $d_l(\mathbf{x}, \mathbf{C})$ is computed.

BMSS can pre-empt the verification step of the RANSAC estimator. If the minimum total cost is not sufficiently small, then the model is labeled incorrect, and verification against all measurements is pre-empted. An advantage of including BMSS is also shown in the ablation study in the synthetic experiments, where the optimal solvers that use

BMSS are compared to solvers that randomly choose a minimal configuration (see Sec. 5.2.3). Note that BMSS makes all the region-based solvers competitive with the contour-based solvers.

Chapter 5

Experiments

5.1 Evaluation Strategy

We report the numerical stabilities (see Fig. 5.2) and noise sensitivities (see Fig. 5.3) of the proposed minimal solvers $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ and $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$ for the problem of auto-calibrating cameras viewing synthetic scenes. Auto-calibration accuracy is measured with relative errors of undistortion and focal length estimation, as well as with the metric warp error (see. Sec. 5.1.1) on 1000 synthetic images of 3D Manhattan scenes.

The proposed solvers are compared with three state-of-the-art affine-rectifying solvers that admit feature configurations that can be used for auto-calibration. The bench of state-of-the-art affine-rectifying solvers consists of 1. $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{lu}\lambda$ of [30], which affinely rectifies from three correspondences of coplanar affine covariant regions, 2. $\mathtt{H}_{22}\mathbf{lu}\lambda$-$\mathtt{fR}$ [31], which requires two correspondences of radially-distorted conjugately translated covariant regions, and 3. $\mathtt{H}^{32}\mathbf{lu}\lambda$-$\mathtt{fR}$ [41], which requires a corresponded set of three imaged parallel scene lines and a correspondence of imaged parallel scene lines, where each set is consistent with a distinct vanishing point. The Manhattan frame is assumed to upgrade from affine-rectified to calibrated space.

5.1.1 Metrics

Camera auto-calibration accuracy is reported in terms of the estimated parameters. Relative error of the estimated division model parameter is used to report the accuracy of the lens undistortion estimate

$$\mu_\lambda = (\lambda - \hat{\lambda})/\lambda. \tag{5.1}$$

and focal length accuracy is reported as

$$\mu_{\mathtt{f}} = (\mathtt{f} - \hat{\mathtt{f}})/\mathtt{f}. \tag{5.2}$$

Metric Warp Error The accuracy of metric rectification is used to jointly assess the accuracy of the recovered auto-calibration parameters. The metric-rectifying conjugate rotation of the camera is used to rectify points

$$\alpha\underline{\mathbf{x}} = \mathtt{H}\mathbf{x} = \mathtt{K}\mathtt{R}\mathtt{K}^{-1}\mathbf{x} = \mathtt{K}\mathtt{R}\mathtt{K}^{-1}f(\tilde{\mathbf{x}}, \lambda). \tag{5.3}$$

Using the conjugate rotation as the rectifying homography links the accuracy of the metric rectification with the accuracy of the estimations of $\mathtt{f}$, $\mathtt{R}$ and λ, which are directly recovered from the minimal solvers.

We modify the affine warp error introduced by Pritts et al. in [29] to admit conjugate rotations. A scene plane is tessellated by a 10x10 square grid of points $\{\,\mathbf{X}_i\,\}_{i=1}^{100}$ and imaged as $\{\,\tilde{\mathbf{x}}_i\,\}_{i=1}^{100}$ by the lens-distorted ground-truth camera. The tessellation ensures that error is uniformly measured over the scene plane. A round trip between the image space and rectified space is made through the following sequence of transformations: 1. The imaged tessellation $\{\,\tilde{\mathbf{x}}_i\,\}_{i=1}^{100}$ is undistorted using the ground truth division model parameter λ, 2. the undistorted points are back-projected to rays and rotated using $\mathtt{R}^{\top}\mathtt{K}^{-1}$ such that the scene plane is fronto-parallel, giving the rectified points $\{\,\underline{\mathbf{x}}_i\,\}_{i=1}^{100}$, and 3. the metric rectified points are imaged and distorted by the estimated camera using $\hat{\mathtt{K}}\hat{\mathtt{R}}$ and $\hat{\lambda}$.

Ideally, the estimated camera $\hat{\mathtt{K}}\hat{\mathtt{R}}$ images the rectified points $\{\,\underline{\mathbf{x}}_i\,\}_{i=1}^{100}$ onto the distorted points $\{\,\tilde{\mathbf{x}}_i\,\}_{i=1}^{100}$. The metric warp error for estimated camera $\hat{\mathtt{K}}\hat{\mathtt{R}}$ is defined as

$$\Delta_{\text{warp}}^{\text{metric}} = \sum_i d^2(\tilde{\mathbf{x}}_i, f^d(\hat{\mathtt{K}}\hat{\mathtt{R}}\mathtt{R}^{\top}\mathtt{K}^{-1}f(\tilde{\mathbf{x}}_i, \lambda)), \hat{\lambda}), \tag{5.4}$$

where $d(\cdot,\cdot)$ is the Euclidean distance, f^d is the inverse of the division model (the inverse of (3.13)). The root mean square metric warp error, denoted $\Delta_{\text{warp}}^{\text{metric}}$, is used in the sensitivity and stability studies.

5.2 Synthetic Scene Experiments

Cameras with realistic focal lengths and lens distortions are randomly placed such that an imaged scene plane occupies a majority of their view. Image resolution is set to 1000×1000 pixels. The ground-truth division model parameters and focal lengths are generated randomly within realistic bounds such that the synthetic cameras are similar to GoPro Hero 4 cameras (see also [31]). Parallel line segments and translated affine frames are oriented on the scene plane so that they are mutually orthogonal. For the sensitivity experiments, white noise is added to the images of the sampled line segments and translated affine

frames. Fig. 5.1 refers to examples of the synthetically generated Manhattan frames with inlying affine features arranged in the pattern of lattice.

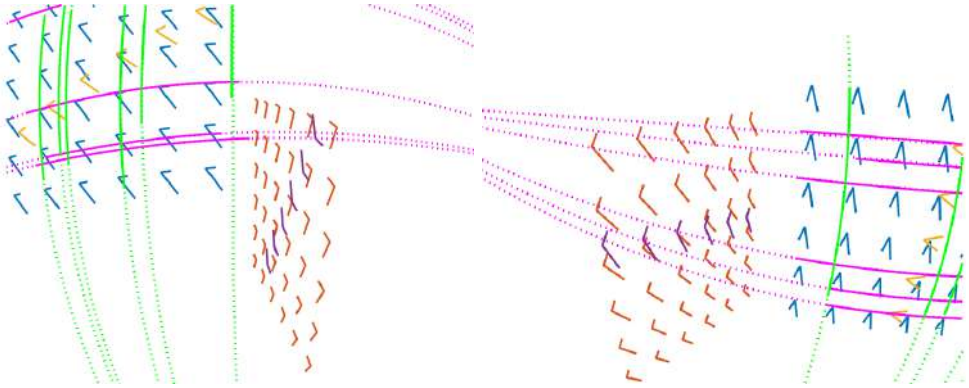


FIGURE 5.1: *Examples of Synthetic Scenes Used for Evaluation.* Illustrated are imaged two orthogonal scene planes each containing affine-covariant features arranged in a lattice (colored blue and red). The dominant scene plane (blue) also contains sets of parallel lines corresponding to orthogonal directions (colored green and magenta). The scene planes also incorporate outliers (yellow and purple) to test the robustness of the framework (see Sec. 5.2).

5.2.1 Numerical Stability

The numerical stability measures the RMS metric warp error $\Delta^{\text{metric}}_{\text{warp}}$ of the two proposed ($\texttt{H}^{2}_{2}\mathbf{lu}\lambda\texttt{-fR}$ and $\texttt{H}^{222}\mathbf{lu}\lambda\texttt{-fR}$) and two state-of-the-art ($\texttt{H}_{22}\mathbf{lu}\lambda\texttt{-fR}$ [32, 30] and $\texttt{H}^{32}\mathbf{lu}\lambda\texttt{-fR}$ [41]) solvers on noiseless features. Configurations of coplanar mutually orthogonal translated affine frames and parallel lines that are consistent with each solver's required inputs are generated for a realistic scene and camera configurations described in the introduction of this section.

Fig. 5.2 reports the distribution of $\log_{10}$ warp errors $\Delta^{\text{metric}}_{\text{warp}}$. All of the proposed solvers demonstrate superior numerical stability, which is consistent with the simple structure of the solvers. The solver $\texttt{H}^{32}\mathbf{lu}\lambda\texttt{-fR}$ [41] admitting fitted circles also has the simple structure and is quite stable. The $\texttt{H}_{22}\mathbf{lu}\lambda\texttt{-fR}$ solver of Pritts et al. [32] has a significant failure frequency. $\texttt{H}_{22}\mathbf{lu}\lambda\texttt{-fR}$ is generated with the Gröbner bases method, which solves a complicated system of polynomial equations. In contrast, the proposed solvers require solving only a quartic and a small linear system (see Chapter 4).

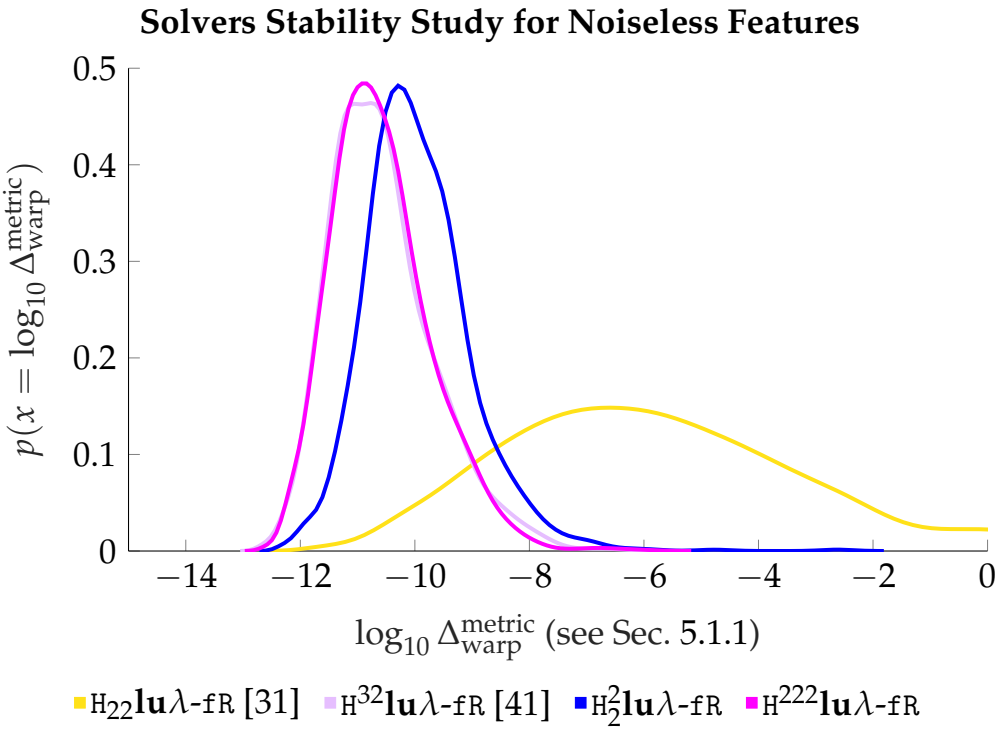


FIGURE 5.2: *Numerical Stability.* For the synthetic experiments, the normalized division model parameter is set to -4, focal length and camera pose are generated randomly within realistic bounds such that the synthetic camera is close to GoPro Hero 4. The RMS metric warp error is reported, which accounts for all the estimated calibration parameters λ, f and R. The proposed solvers are stable.

5.2.2 Noise Sensitivity

The proposed $\mathtt{H}^{2}_{2}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ ans $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solvers and the state-of-the-art $\mathtt{H}^{\text{DES}}_{222}\mathbf{lu}\lambda$, $\mathtt{H}_{22}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and $\mathtt{H}^{32}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solvers, upgraded for auto-calibration as described in Sec. 4.3, are evaluated for their robustness to sensor noise. The RMS metric warp error $\Delta^{\text{metric}}_{\text{warp}}$ is used to measure the accuracy of the auto-calibration (see Sec. 5.1.1). White noise is sampled from a zero-mean Gaussian and added to the arcs and radially-distorted conjugate translations. Solver sensitivity is measured at noise levels of $\sigma_{\text{PT}} \in \{\, 0.5, 1, 2 \,\}$.

The solvers are used in a basic RANSAC estimator that minimizes the RMS metric warp error $\Delta^{\text{metric}}_{\text{warp}}$ over 25 minimal samples for each of the 1000 scenes at each noise level. Each boxplot of Fig. 5.3 has a constant arc noise level and varying affine frame noise level. As expected, the solvers admitting arcs give superior performance since the entirety of the arc is used for fitting. The proposed $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ gives the best performance, handles the case where correspondences are consistent with either two or three vanishing points, and requires only arc pairs (in contrast, to $\mathtt{H}^{32}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ which requires a corresponded set of three

along with an arc pair). The proposed $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ shows significantly improved stability over the state-of-the-art solvers $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{lu}\lambda$ and $\mathtt{H}_{22}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ of Pritts et al. [32, 30, 31]. The state-of-the-art solver $\mathtt{H}^{32}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ of [41], which uses only arcs, also demonstrates good robustness across all the noise levels.

5.2.3 Impact of Best Minimal Solution Selection

The ablation study in Table 5.1 compares the performance of the proposed solvers with best minimal solution selection to variants that randomly select from feature configurations (see Sec. 4.4). BMSS has a significant impact and reduces warp error by 49.2% on average for $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solver and by 30.2% on average for $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ solver. The inclusion of BMSS is further justified by the extremely fast time-to-solution of the proposed solvers.

Solver	$\Delta_{\mathrm{warp}}^{\mathrm{metric}}$ at 1px-σ		Improvement
	Random	BMSS	
$\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$	6.9	3.5	49.2%
$\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$	5.3	3.7	30.2%

TABLE 5.1: *Ablation Study.* The study illustrates an advance of the solvers $\mathtt{H}_2^2\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and $\mathtt{H}^{222}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ that use BMSS, comparing to $\mathtt{H}_2^{2\mathrm{RND}}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ and $\mathtt{H}^{222\mathrm{RND}}\mathbf{lu}\lambda\text{-}\mathtt{fR}$ that randomly choose one of possible configurations given a minimal sample set.

5.3 Performance on Real Images

5.3.1 Robust Estimation

The proposed solvers are utilized in a LO-RANSAC-based robust estimation framework [13, 14, 28]. A RANSAC hypothesis consists of the estimations of radial division model parameter, focal length, and rectifying camera rotation $\mathfrak{H} = \{\lambda, f, \mathtt{R}\}$. The hypothesized models with the best-so-far maximal consensus sets are locally optimized by an extension of the method introduced in [28].

MSS configurations for the proposed minimal solvers are given in Table 4.1 and the examples are illustrated in Fig. 4.1. For each iteration of RANSAC, appearance clusters are selected with the probability which is equal to the cluster's relative cardinality to the other appearance clusters, as well as LCC clusters are chosen with the probability given by its relative cardinality to the other LCC clusters. The required

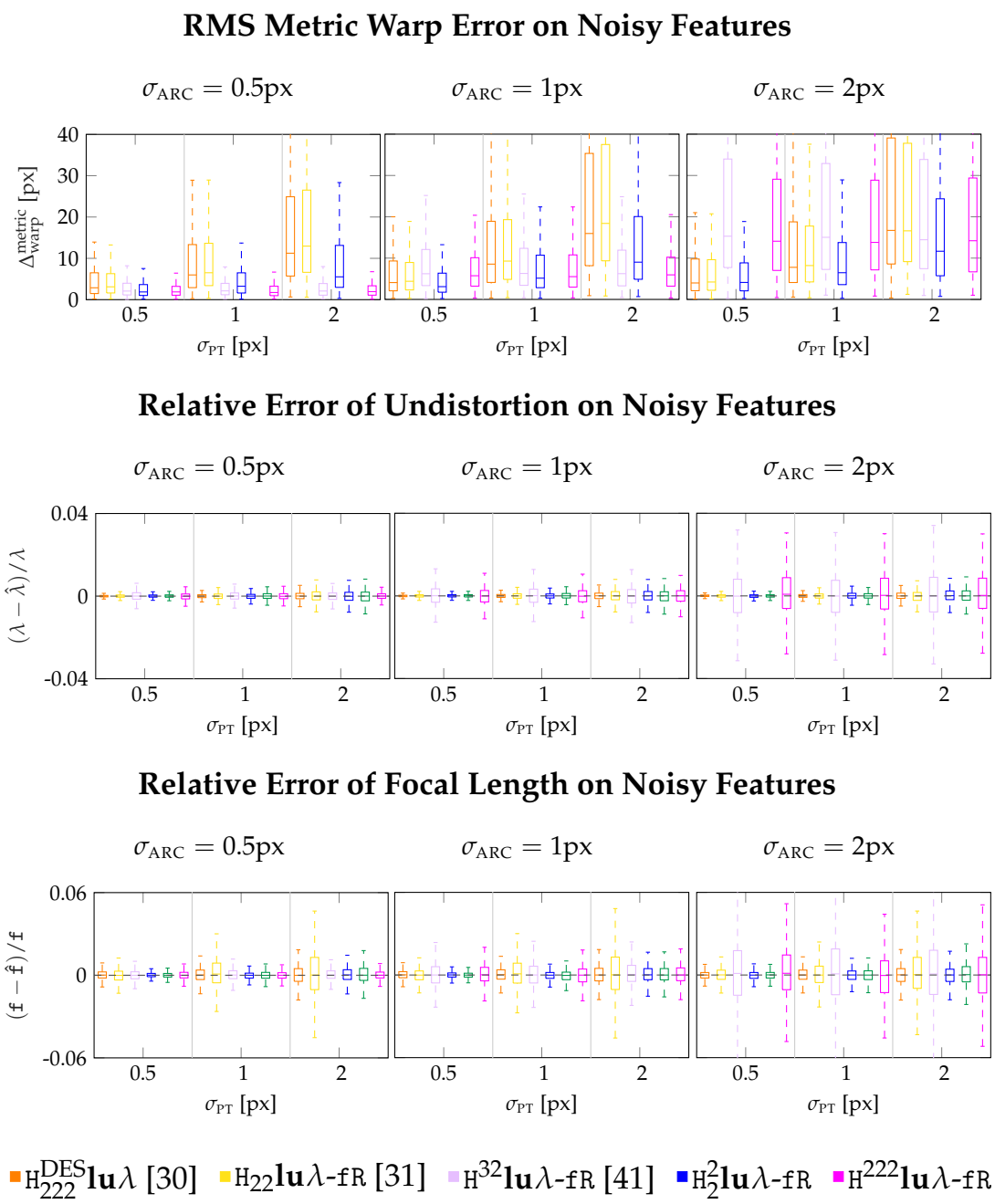


FIGURE 5.3: *Noise Sensitivity Benchmark.* Reported is RMS metric warp error after 25 iterations of a simple RANSAC for the state-of-the-art and proposed minimal solvers that were upgraded for auto-calibration of the Manhattan synthetic scene with increasing levels of the noise σ_{PT} added to the local affine frames, and the noise σ_{ARC} added to the points of the circular arcs in the orthogonal directions of the circles. The synthetic camera parameters were generated in the same way as for solvers' numerical stability experiment.

number of feature correspondences are then drawn from the selected clusters to generate a RANSAC hypothesis using the proposed minimal

solvers and evaluate it based on the cardinality of the consensus set computed.

Covariant Features Extraction and Tentative Correspondences Affine frames are tentatively labeled as a repeated pattern by their appearance. The appearance of an affine frame is given by the RootSIFT description of the image patch local to the affine frame [4]. Affine-covariant regions are also extracted and embedded in the reflected image, where the detections are transformed into the original image space and have a left-handed representation.

The RootSIFT embeddings are agglomeratively clustered, which establishes pair-wise tentative correspondences amongst connected components [31]. Since the proposed hybrid solver does not admit reflections, the appearance clusters are partitioned based on the handedness of the affine frames associated with the clustered embedded regions [31].

(A) Tentative LAFs clusters (B) Tentative LCCs

FIGURE 5.4: *Tentative Correspondences of Features.* (A) The RootSIFT descriptors are used to tentatively cluster LAFs as radially-distorted conjugately-translated affine-covariant regions (an image is taken from [32]), and (B) LCCs are used to tentatively group the circular arcs as radially-distorted imaged parallel scene lines.

Circluar Arcs Extraction and Tentative Correspondences Arc extraction consists of two stages: 1. arc segment detection and 2. circle fitting. The approach is similar to [40]. A subpixel Canny edge detector is used to [10] extract edges from the input image. Short edges are removed, and the remaining contours are split into circular arcs. The circular arcs are segmented with the Ramer–Douglas–Peucker simplification algorithm [17]. Over-segmented arcs (arcs belonging to the same circle) are connected. Each arc is fitted to a circle using Taubin's circle estimator. The Taubin estimate is used as an initial guess to a non-linear

least squares minimizer, which gives the maximum likelihood estimate [11]. The endpoints of contours are projected onto the fitted circles and are used to compute the midpoint and the normal of the circle located at the midpoint. These parameters are the required inputs to the proposed solvers as detailed in Sec. 4.1.3.

Tentative correspondences of detected circular arcs are established by utilizing the geometry of the imaging process of the lines by radially-distorted cameras modeled with the division model. Parallel scene lines are projected by a pinhole camera to a pencil of lines intersecting at a vanishing point and imaged by a radially-distorted camera to a pencil of circles intersecting at two points corresponding to the opposite directions of these lines, where one of these points is a distorted vanishing point. The circle centers of the distorted parallel scene lines are lying on the same line called *Line of Circle Centers* (LCC). Essentially, the LCC can be seen as a circle pencil [3]. Thus different correspondences of distorted scene lines relate to different LCCs.

The join of a pair of circle centers is drawn and used to fit a line. Then, a set of inlying circle centers is found by thresholding the orthogonal distance from the point to the line. All the pairs of circle correspondences are created from this set of inliers. The procedure is repeated for each pair of circles in the set of extracted circles and the union of all the tentative pairs of circles in correspondence is formed.

Fig. 5.4 illustrates an example of the tentative correspondences established for extracted image features.

5.3.2 Experimental Results

On Fig. 5.6 and Fig. 5.5 we present the performance of the proposed solvers on the real images. Integrated into the robust estimation framework, the solvers show an accurate undistortion, as well as focal length estimation and metric rectification. The experiments indicate that the algorithm effectively handles various camera lens distortions, from the cell phone and near rectilinear lens up to fish-eye cameras that have a high distortion, as well as different field of view options. We are also able to label the distorted line correspondences and the directions of the conjugate translations of the regions on the imaged scene planes with the estimated vanishing points of the Manhattan direction; on Fig. 5.6 and Fig. 5.5 labelling is shown as as red, green, blue circles; we also label the repeated regions with estimated vanishing lines corresponding to multiple imaged scene planes. The distorted vanishing lines are colored yellow, magenta, and cyan for the three scene planes.

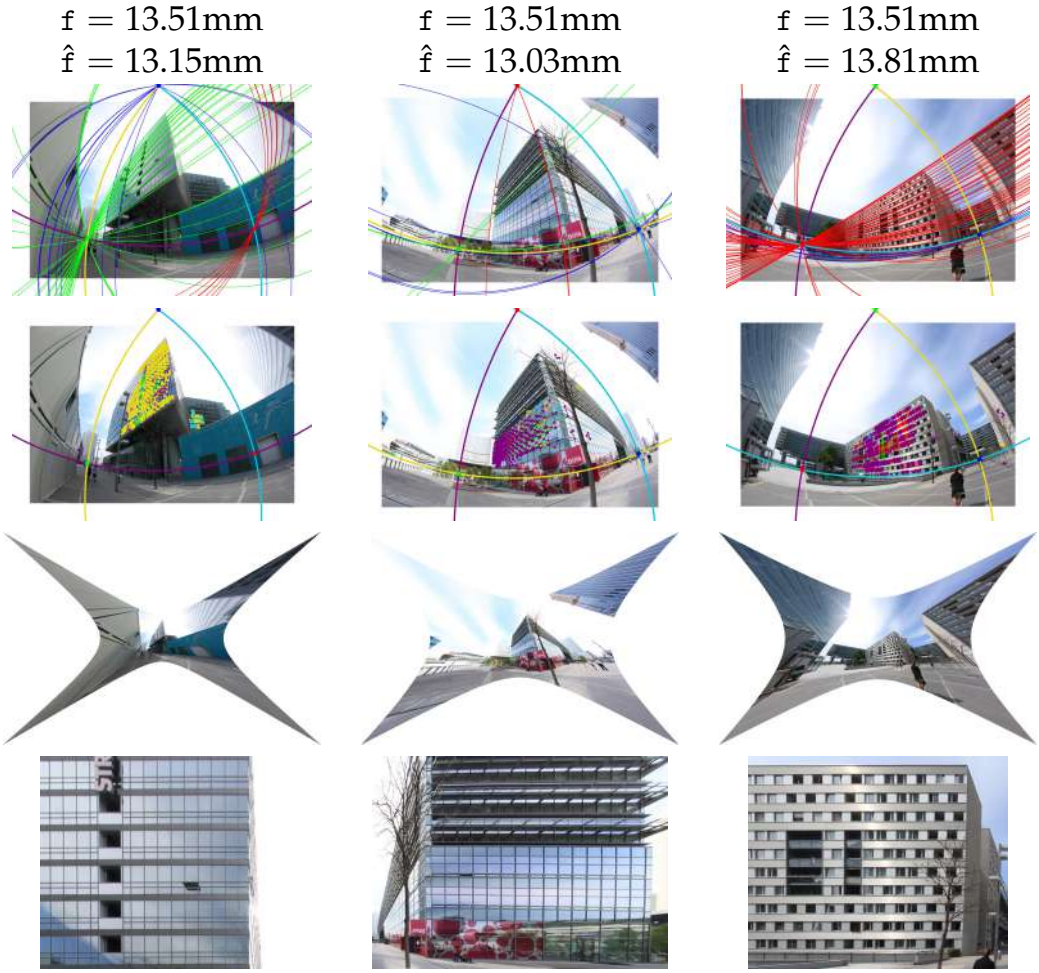


FIGURE 5.5: *Real Data Experiments: AIT.* The proposed solvers accurately estimate distortion and focal length, and metric-rectify even for strong distortion and wide fields-of-view. The outputs are vanishing point labeling of arcs and region correspondences (first row), and vanishing line labeling of regions (second row); undistorted (third row) and rectified images (bottom row). The left-most result represents a very good performance (a highly accurate estimation of undistortion can be seen, as well as very low relative error in focal length). The middle column is for a good performance, and the right-most column one represents a moderate outcome. The experiments were conducted for AIT dataset [41].

5.4 Further Analysis

5.4.1 Computational Complexity of the Solvers

We present a comparison of the mean time-to-solution in the wall clock time of the proposed solvers—$\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$, $\mathtt{H}^{222}\mathbf{lu}\lambda$-$\mathtt{fR}$—with the state-of-the-art solvers $\mathtt{H}_{22}\mathbf{lu}\lambda$-$\mathtt{fR}$ of [32] and $\mathtt{H}_{222}^{\mathtt{DES}}\mathbf{lu}\lambda$ of [30]. All solvers were written and optimized in C++. Relative speeds are reported with respect to the $\mathtt{H}_2^2\mathbf{lu}\lambda$-$\mathtt{fR}$ solver for easy comparison. The proposed solvers

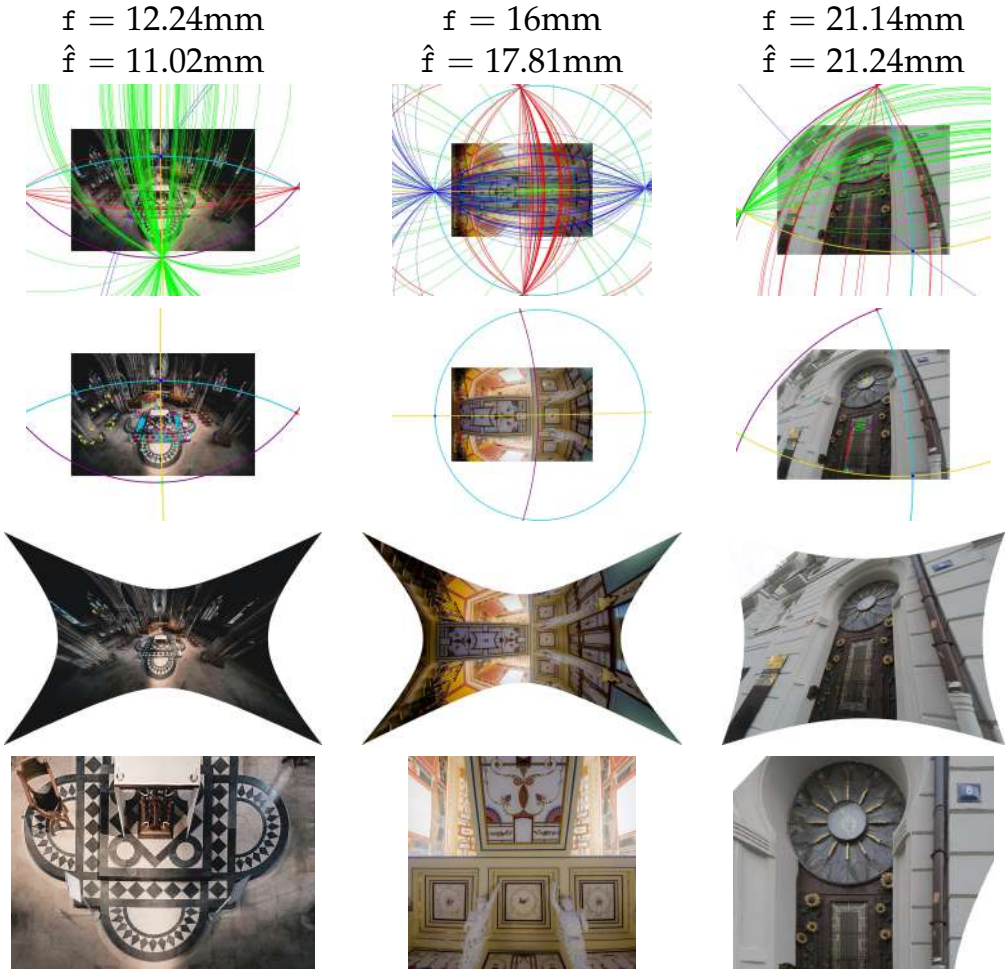


FIGURE 5.6: *Real Data Experiments: Challenging Images.* The proposed solvers were tested on challenging images taken by various cameras with different fields-of-view and distortions. Presented are examples of estimated focal length (top title), visual outputs: vanishing point labeling of arcs and region correspondences (first row) and vanishing line labeling of regions (second row); radially undistorted (third row) and metric-rectified images (bottom row).

have similar algebraic structure and have the same mean time-to-solution of **0.5 µs**. Each gives a 2153.6× speed up over the $\texttt{H}_{222}^{\text{DES}}\mathbf{lu}\lambda$ solver and a 69.2× speed up over the $\texttt{H}_{22}\mathbf{lu}\lambda$-$\texttt{fR}$ solver (see Table 5.2). The slow run times of the $\texttt{H}_{222}^{\text{DES}}\mathbf{lu}\lambda$ and $\texttt{H}_{22}\mathbf{lu}\lambda$-$\texttt{fR}$ solvers can be attributed to their need to solve complicated polynomial systems of equations using the Gröbner bases method. All of the proposed solvers are much more suitable for fast sampling in RANSAC for scenes containing translational symmetries and lines.

Solver	Wall Clock	Relative Time
$\mathtt{H}_2^2\mathbf{lu}\lambda$-fR	**0.5 µs**	**1.0**×
$\mathtt{H}^{222}\mathbf{lu}\lambda$-fR	**0.5 µs**	**1.0**×
$\mathtt{H}_{22}\mathbf{lu}\lambda$-fR	34.6 µs	69.2×
$\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{lu}\lambda$	1076.8 µs	2153.6×

TABLE 5.2: *Runtime Analysis.* Wall-clock times are reported for optimized C++ implementations of the proposed solvers—$\mathtt{H}_2^2\mathbf{lu}\lambda$-fR, $\mathtt{H}^{222}\mathbf{lu}\lambda$-fR—versus $\mathtt{H}_{22}\mathbf{lu}\lambda$-fR and $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{lu}\lambda$ of [32, 33, 30]. The proposed solvers are significantly faster then the state-of-the-art.

Camera Id	Zhang [46]	Minimal Solution		MLE	
	$\mathtt{f}$, mm	$\hat{\mathtt{f}}$, mm	$\mu_{\mathtt{f}}$	$\hat{\mathtt{f}}$, mm	$\mu_{\mathtt{f}}$
Left-Most Camera	1.939	1.893	0.024	1.954	0.008
Right-Most Camera	1.946	1.975	0.015	1.968	0.011
Right-Side Camera	1.944	1.962	0.009	1.929	0.008
Right-Center Camera	1.946	1.994	0.025	1.989	0.022
Right-Front Camera	1.945	1.957	0.006	1.956	0.006
Left-Front Camera	1.937	1.898	0.02	1.939	0.001

TABLE 5.3: *Single-View vs. Multi-View Study.* For the Zhang's calibration technique [46], which is a golden standard, 63 images of the calibration targets were utilized to estimate the focal length whereas the proposed method gives as accurate estimates using a single image. See Fig. 5.7 illustrating typical output of the proposed method for the calibration data used in this experiment.

5.4.2 Single-View vs. Multi-View Camera Calibration

Table 5.3 reports the focal length estimates obtained by Zhang's calibration technique [46] from acquired 63 images versus the proposed technique. Fig. 5.7 gives an example of the input data (the information about the calibration targets was not utilized) and the results obtained by the proposed method in this experiment. It is noteworthy to highlight several advantages of the proposed method: only one input image is needed, and no knowledge of the calibration target is required while the estimates are comparably accurate; still camera calibration from multiple images can be easily incorporated to improve the results.

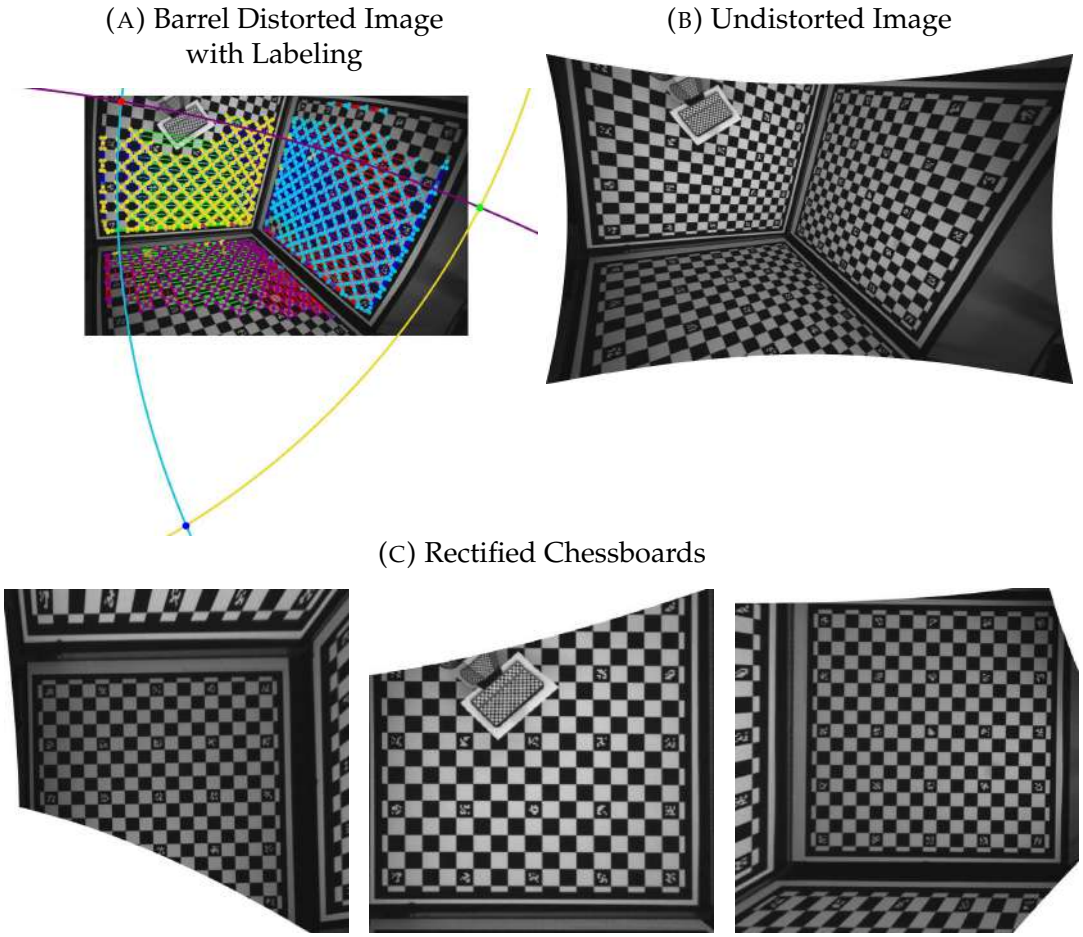


FIGURE 5.7: *Rectification from Radially Distorted Points.* The chessboard scene was undistorted and rectified using a minimal solver that jointly estimates lens distortion and rectification. (A) The corners of each chessboard are color coded with the distorted image of the vanishing line. The radially-distorted conjugate translations used in the estimation are color coded with the distorted vanishing point where they meet. (B) Undistorted with the division model. (C) Each chessboard is metrically-rectified.

Chapter 6

Conclusions

6.1 Summary

This work proposes the first hybrid solvers that jointly admit correspondences of points and circular arcs for camera auto-calibration. This novelty is achieved by incorporating constraints on the tangents of circular arcs from differential geometry with constraints on points induced by radially-distorted conjugate translations. Remarkably, these constraints can be combined and simplified such that the joint estimation of lens undistortion and affine rectification from two or three correspondences of points and arcs is possible by sequentially solving a quartic and a simple linear system. Auto-calibration is also directly recovered if a Manhattan scene is assumed. The simple structure of the equations results in very robust and fast solvers. The solvers extend accurate auto-calibration to images where there is a sparsity of either translational symmetries or scene lines, and they are ideal for use in robust estimators designed to use complementary feature types, like Hybrid Ransac [9].

6.2 Discussion

Many computer vision tasks can be alleviated with known camera calibration parameters. The most widely-known is the problem of 3D reconstruction [43, 34, 44]. Unfortunately, the camera calibration information is rarely available, especially if considering imagery taken from the Internet. A proposed approach to an automatic camera calibration from a single image is based on the geometric properties of imaging a man-made world under several assumptions on camera (zero skew, square pixels and principal point coinciding with an image center) and the scene (a Manhattan frame, but optional if only distortion and horizon estimations are needed). The proposed minimal solvers are simple but efficient, and can be easily integrated into a robust framework such as RANSAC as was illustrated in Chapter 5. A geometric approach

makes the proposed method more accurate than learning techniques *e.g.*, [6], which do not incorporate geometric constraints.

Bibliography

[1] Shahzor Ahmad and Loong-Fah Cheong. "Robust detection and affine rectification of planar homogeneous texture for scene understanding". In: *International Journal of Computer Vision* (2018), pp. 1–33.

[2] Dror Aiger, Daniel Cohen-Or, and Niloy J Mitra. "Repetition maximization based texture rectification". In: 31.2pt2 (2012), pp. 439–448.

[3] Michel Antunes et al. "Unsupervised vanishing point detection and camera calibration from a single manhattan image with radial distortion". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 4288–4296.

[4] Relja Arandjelović and Andrew Zisserman. "Three things everyone should know to improve object retrieval". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 2911–2918.

[5] João Pedro Barreto and Kostas Daniilidis. "Fundamental matrix for cameras with radial distortion". In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Vol. 1. IEEE. 2005, pp. 625–632.

[6] Oleksandr Bogdan et al. "DeepCalib: a deep learning approach for automatic intrinsic calibration of wide field-of-view cameras". In: *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*. ACM. 2018, p. 6.

[7] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[8] Faisal Bukhari and Matthew N Dailey. "Automatic radial distortion estimation from a single image". In: *Journal of mathematical imaging and vision* 45.1 (2013), pp. 31–45.

[9] Federico Camposeco et al. "Hybrid camera pose estimation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 136–144.

[10] John Canny. "A computational approach to edge detection". In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), pp. 679–698.

[11] Nikolai Chernov. *Circular and linear regression: Fitting circles and lines by least squares*. CRC Press, 2010.

[12] Ondřej Chum and Jiří Matas. "Planar affine rectification from change of scale". In: *Asian Conference on Computer Vision*. Springer. 2010, pp. 347–360.

[13] Ondřej Chum, Jiří Matas, and Josef Kittler. "Locally optimized RANSAC". In: *Pattern Recognition*. Ed. by Bernd Michaelis and Gerald Krell. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 236–243. ISBN: 978-3-540-45243-0.

[14] Ondrej Chum, Jirı Matas, and Stepan Obdrzalek. "Enhancing RANSAC by generalized model optimization". In: *Proc. of the ACCV*. Vol. 2. 2004, pp. 812–817.

[15] David A Cox, John Little, and Donal O'shea. "Using algebraic geometry". In: vol. 185. Springer Science & Business Media, 2006.

[16] Antonio Criminisi and Andrew Zisserman. "Shape from texture: homogeneity revisited". In: *BMVC*. 2000.

[17] David H Douglas and Thomas K Peucker. "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature". In: *Cartographica: the international journal for geographic information and geovisualization* 10.2 (1973), pp. 112–122.

[18] Moumen T El-Melegy and Aly A Farag. "Nonmetric lens distortion calibration: closed-form solutions, robust estimation and model selection". In: *null*. IEEE. 2003, p. 554.

[19] Martin A Fischler and Robert C Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Communications of the ACM* 24.6 (1981), pp. 381–395.

[20] Andrew W Fitzgibbon. "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*. Vol. 1. IEEE. 2001, pp. I–I.

[21] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[22] David Liebowitz and Andrew Zisserman. "Metric rectification for perspective images of planes". In: *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231)*. IEEE. 1998, pp. 482–488.

[23] Michal Lukáč et al. "Nautilus: recovering regional symmetry transformations for image editing". In: *ACM Transactions on Graphics (TOG)* 36.4 (2017), p. 108.

[24] Jiri Matas, T Obdrzalek, and Ondrej Chum. "Local affine frames for wide-baseline stereo". In: *Object recognition supported by user interaction for service robots*. Vol. 4. IEEE. 2002, pp. 363–366.

[25] Jiri Matas et al. "Robust wide-baseline stereo from maximally stable extremal regions". In: vol. 22. 10. Elsevier, 2004, pp. 761–767.

[26] Stepan Obdrzalek and Jiri Matas. "Object recognition using local affine frames on distinguished regions". In: *BMVC*. Vol. 1. 2002, p. 3.

[27] Tomas Pajdla. *Elements of geometry for computer vision*. 2013.

[28] James Pritts, Ondrej Chum, and Jiri Matas. "Detection, rectification and segmentation of coplanar repeated patterns". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 2973–2980.

[29] James Pritts et al. "Coplanar repeats by energy minimization". In: 2017.

[30] James Pritts et al. "Minimal solvers for rectifying from radially-distorted conjugate translations". In: *arXiv preprint arXiv:1911.01507* (2019).

[31] James Pritts et al. "Minimal solvers for rectifying from radially-distorted conjugate translations". In: *arXiv preprint arXiv:1911.01507* (2019).

[32] James Pritts et al. "Radially-distorted conjugate translations". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 1993–2001.

[33] James Pritts et al. "Rectification from radially-distorted scales". In: *Asian Conference on Computer Vision*. Springer. 2018, pp. 36–52.

[34] Srikumar Ramalingam and Matthew Brand. "Lifting 3d manhattan lines from a single image". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 497–504.

[35] Frederik Schaffalitzky and Andrew Zisserman. "Geometric grouping of repeated elements within images". In: *Shape, contour and grouping in computer vision*. Springer, 1999, pp. 165–181.

[36] Rickard Strand and Eric Hayman. "Correcting radial distortion by circle fitting." In: *BMVC*. 2005.

[37] Jean-Philippe Tardif. "Non-iterative approach for fast and accurate vanishing point detection". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 1250–1257.

[38] Andrea Vedaldi and Brian Fulkerson. *VLFeat: An open and portable library of computer vision algorithms*. ACM, 2010.

[39] Aiqi Wang, Tianshuang Qiu, and Longtan Shao. "A simple method of radial distortion correction with centre of distortion estimation". In: *Journal of Mathematical Imaging and Vision* 35.3 (2009), pp. 165–172.

[40] Horst Wildenauer and Allan Hanbury. "Robust camera self-calibration from monocular images of Manhattan worlds". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 2831–2838.

[41] Horst Wildenauer and Branislav Micusik. "Closed form solution for radial distortion estimation from a single vanishing point". In: *BMVC*. Vol. 1. 2013, p. 2.

[42] Reg G Willson and Steven A Shafer. "What is the center of the image?" In: *JOSA A* 11.11 (1994), pp. 2946–2955.

[43] Changchang Wu, Jan-Michael Frahm, and Marc Pollefeys. "Repetition-based dense single-view reconstruction". In: *CVPR 2011*. IEEE. 2011, pp. 3113–3120.

[44] Hao Yang and Hui Zhang. "Efficient 3d room shape recovery from a single panorama". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 5422–5430.

[45] Zhengdong Zhang et al. "TILT: Transform invariant low-rank textures". In: *International journal of computer vision* 99.1 (2012), pp. 1–24.

[46] Zhengyou Zhang. "A flexible new technique for camera calibration". In: *IEEE Transactions on pattern analysis and machine intelligence* 22 (2000).