Methods for the Rectification of Imaged Coplanar Repeated Patterns

A dissertation presented to the Faculty of Electrical Engineering at Czech Technical University in Prague in partial fulfillment of the requirements for the Ph.D. degree in study program No. P 2612 - Electrotechnics and Informatics, branch No. 3902V035 - Artificial Intelligence and Biocybernetics, by

James Pritts

Prague, January 2020

Thesis Advisor

Doc. Mgr. Ondřej Chum, Ph.D.

Thesis Co-advisor

RNDr. Zuzana Kúkelová, Ph.D.

Visual Recognition Group
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo náměstí 13, 121 35 Prague 2, Czech Republic
phone: +420 224 357 666
http://cmp.felk.cvut.cz

Abstract

This thesis presents novel, general and automated methods for the detection, rectification, and segmentation of imaged coplanar repeated patterns. The only assumption made of the scene geometry is that repeated scene elements are mapped onto each other by planar rigid transformations. The class of patterns covered is broad and includes nearly all planar man-made repeated patterns.

Novel minimal solvers are used to hypothesize lens undistortion and plane rectification parameters. A stratum of constraints is derived that defines the necessary configurations of coplanar repeats for each successive level of rectification. The methods work on scenes without straight lines and, in general, relax strong assumptions about scene content made by the state of the art. A synthetic fronto-parallel view of an imaged scene plane (equivalently metric rectification) can be estimated with as few a three coplanar repeats from an image taken with a rectilinear lens or with as few as four coplanar repeats from a radially-distorted lens.

The thesis introduces the first minimal solvers that jointly estimate lens undistortion and affine rectification from imaged coplanar repeated texture. Even with imagery from moderately distorted lenses, plane rectification using the pinhole camera model is inaccurate or invalid. The proposed solvers incorporate lens distortion into the camera model and extend accurate rectification to wide-angle imagery, which is now common from consumer cameras. Accurate rectifications on imagery taken with narrow focal lengths to fisheye lenses demonstrate the broad applicability of the proposed solvers.

In addition, a multi-model estimator is proposed to solve the ill-posed problem of jointly segmenting repetitive texture and regressing the rectification. The estimator encodes a discriminative model in an energy functional that captures global interactions between distinct coplanar repeated patterns and scene planes and combines several features that model how planes with coplanar repeats are projected into images. Energy minimization is achieved by alternately solving labeling and regression problems, which correspond to repetitive texture and plane segmentation and scene geometry estimation, respectively.

Keywords: rectification, repeated patterns, minimal problems, radial distortion, minimal solvers, symmetry, local features

Abstrakt

Disertační práce prezentuje nové a obecné metody pro automatickou detekci, narovnání a segmentaci opakujících se rovinných vzorů. Jediný předpoklad o geometrii scény je ten, že opakující se prvky je možné na sebe transformovat pomocí eukleidovské rovinné transformace. Tuto podmínku splňuje široká škála vzorů, do které spadají téměř všechny lidmi vytvořené opakující se rovinné obrazce.

V práci jsou navržený nové algoritmy pro řešení některých minimálních problémů - výpočet parametrů zkreslení objektivu a parametrů narovnání roviny. Jsou odvozeny nezbytné podmínky pro konfigurace koplanárních opakujících se vzorů pro jednotlivé druhy narovnání. Navržené metody nevyžadují silné předpoklady, jako například stávajícími metodami vyžadovaná přítomnost přímek ve scéně. Nalezené parametry umožňují vygenerovat syntetický čelní pohled snímané scény, a to ze třech korespondujících koplanárních oblastí v případě prosté projektivní kamery a ze čtyřech oblastí v případě radiálně zkreslených snímků.

Tato disertační práce prezentuje první algoritmy pro minimální problémy, které z obrazu opakujícího rovinného vzoru počítají současně parametry radiálního zkreslení a parametry afinního narovnání roviny. I v případě malého radiálního zkreslení narovnání roviny za použití prostého projektivního modelu kamery vede k nepřesným či chybným výsledkům. Navržené algoritmy proto začleňují radiální zkreslení do modelu kamery a poskytují přesné odzkreslení roviny i pro širokoúhlé kamery, které jsou v současné době běžné používané. Přesné narovnání roviny z kamer s úzkým zorným polem stejně tak jako širokoúhlých kamer s velkým radiálním zkreslením demonstruje širokou použitelnost navržených metod.

V práci je dále navržena metoda pro souběžný odhad několika modelů, řešící špatně podmíněný problém současné segmentace opakujících se textur a odhadu parametrů narovnání. Navržený estimátor obsahuje diskriminativní model v objektivní funkci, který zachycuje globální vazby mezi různými koplanárními opakujícími se vzory, různými rovinami ve scéně, a jejich projekcemi do obrazů. Funkce je minimalizována střídavým opakovaným řešením problémů značkování a regrese, což odpovídá problémům segmentace rovin s opakující se texturou a odhadování geometrie scény.

Klíčová slova narovnání roviny, opakující se vzory, minimální problémy, radiální zkreselní, systémy polynomiálních rovnic, symetrie

Acknowledgements

James Pritts was supported by the European Regional Development Fund under the project Robotics for Industry 4.0 (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000470) and the grants MSMT LL1303 ERC-CZ, GACR P103/12/2310, and SGS17/185/OHK3/3T/13.

Contents

List of Figures

1 Introduction

Scene-plane rectification is a fundamental task of computer vision and is a prerequisite for many classic computer-vision tasks. Rectification restores properties of the scene that make it possible to detect parallel scene lines, measure ratios of areas, angles and ratios of lengths. The task of restoring these properties is a gateway to many sophisticated computer-vision applications. In particular, this thesis proposes methods for the robust estimation of rectification from imaged coplanar repeated texture. The importance of detecting and modeling imaged repeated scene elements grows with the increasing usage of scene-understanding systems in urban settings, where man-made objects predominate and coplanar repeated structures are common. Methods that jointly detect coplanar repeated texture and estimate imaged scene-plane rectification serve as powerful tools for scene understanding.

There are several tasks where rectification is essential and coplanar repetitions are assumed. *E.g.*, single-view 3D reconstruction of buildings and facade parsing relies on comparing metric properties of matched features on the building facade [96]. Inpainting and image beautification are symmetry-aware image editing tasks that benefit from planar symmetry labeling, which is performed in a rectified space [61].

State-of-the-art rectification methods that use coplanar repeated texture as input ignore lens distortion. However, wide-angle imagery that has significant lens distortion is common since consumer photography is now dominated by mobile-phone and GoPro-type cameras. High-accuracy rectification from wide-angle imagery is not possible with only rectilinear camera models [45, 94]. Lens distortion can be estimated by performing a camera calibration a priori, but a fully automated method for scene-plane rectification is desirable. Furthermore, in the case of Internet imagery, the camera and its metadata are often unavailable for use with off-line calibration techniques. A primary goal of the proposed methods in this thesis is to extend accurate rectification to lens-distorted images containing coplanar repeated texture.

Augmented reality applications require calibrated cameras to place virtual objects in the imaged scene, and calibration data may not be readily available for Internet imagery or from mobile phones. Rectification is a necessary task of single-view auto-calibration methods. The state-of-the-art single-view auto-calibration methods for lens-distorted images rely on the presence of parallel scene lines to estimate vanishing points [93, 94, 3]. Expanding rectification to lens-distorted images with sparse scene lines but available coplanar repeated scene texture also expands the class of images that can be auto-calibrated.

In particular, the affine rectification of a scene plane transforms the camera's principal plane so that it is parallel to the scene plane. This restores the affine invariants of the imaged scene plane, which include parallelism of lines and translational symmetries [32, 74]. There is only an affine transformation between the affine-rectified imaged scene plane and its real-world counterpart.

Chapters 5 and 6 propose minimal solvers that jointly estimate lens undistortion and affine rectification from local features extracted from repeating coplanar texture. The proposed solvers are the first solvers that can directly affinely rectify from the radially-distorted image of points or regions extracted from coplanar repeated texture. The input to the solvers are intra-image correspondences of local features. Geometrically, the local features are represented by *local affine frames*, that is, by triplets of (semi-) locally measured image points (see Section 3.2).

The solvers can be differentiated by the assumptions made with respect to the configurations of the inputted local features. Chapter 5 introduces solvers that jointly undistort and affinely rectify from the imaged translations and reflections. This feature configuration is shown in Figure 1.1a. Chapter 6 generalizes joint undistortion and affine rectification to work with the images of rigidly-transformed local features. This feature configuration is shown in Figure 1.1b. All of the proposed solvers eliminate the intermediary undistortion step that is required by the state-of-the-art solvers using repeated texture as input. The best solver can be chosen based on the expected scene content or speed requirements of the application. In general, the solvers are fast and robust to image noise, so they work well in robust estimation frameworks like RANSAC [24].

Metric rectification restores the metric invariants of the imaged scene plane, which include length ratios and angles [32, 74]. In general, the removal of the effects of perspective imaging is helpful for understanding the geometry of the scene plane, and the recovery of metric invariants greatly helps with tasks such as detecting symmetries and repeated image content. Metric rectification is used throughout the thesis to synthesize fronto-parallel views of scene planes. Section 2.10.3 introduces linear minimal solvers to estimate either a semi-metric upgrade from the affine-rectified images of glide-reflected coplanar repeated texture (see Figure 1.1a) or a metric upgrade from the affine-rectified images of rigidly-transformed coplanar repeated texture (see Figure 1.1b).

Imaged scene plane rectification is a poorly constrained problem and verifying the restoration of affine or metric invariants is typically insufficient to correctly assign measurements extracted from imaged scene planes to the model that generated it in the multi-plane setting. Furthermore, good feature coverage over large spans of the imaged scene plane is necessary to properly constrain rectification estimation. Chapter 7 proposes an energy functional that combines several features that model how planes with coplanar repeats are projected into images and captures global interactions between different coplanar repeated texture and scene planes. In particular, regularization terms are incorporated that encourage the assignment of measurements to models such that they conform to the expectations of how a physical scene containing planes must look. These scene prior terms benefit rectification estimation by assuring smooth and dense coverage of measurements over contiguous spans of the imaged scene plane. Minimal solvers for rectification, *e.g.*, proposed in Chapters 5 and 6, are easily plugged into the energy minimization framework to provide scene plane proposals. The model proposals are jointly and globally evaluated, which prevents a model's validity from being biased by the order of its evaluation, which is a common problem with greedy methods like sequential RANSAC. These properties of the energy function and minimization proposed in Chapter 7 enable rectification solvers to be used in difficult scenes with multiple planes lacking a dominant plane.

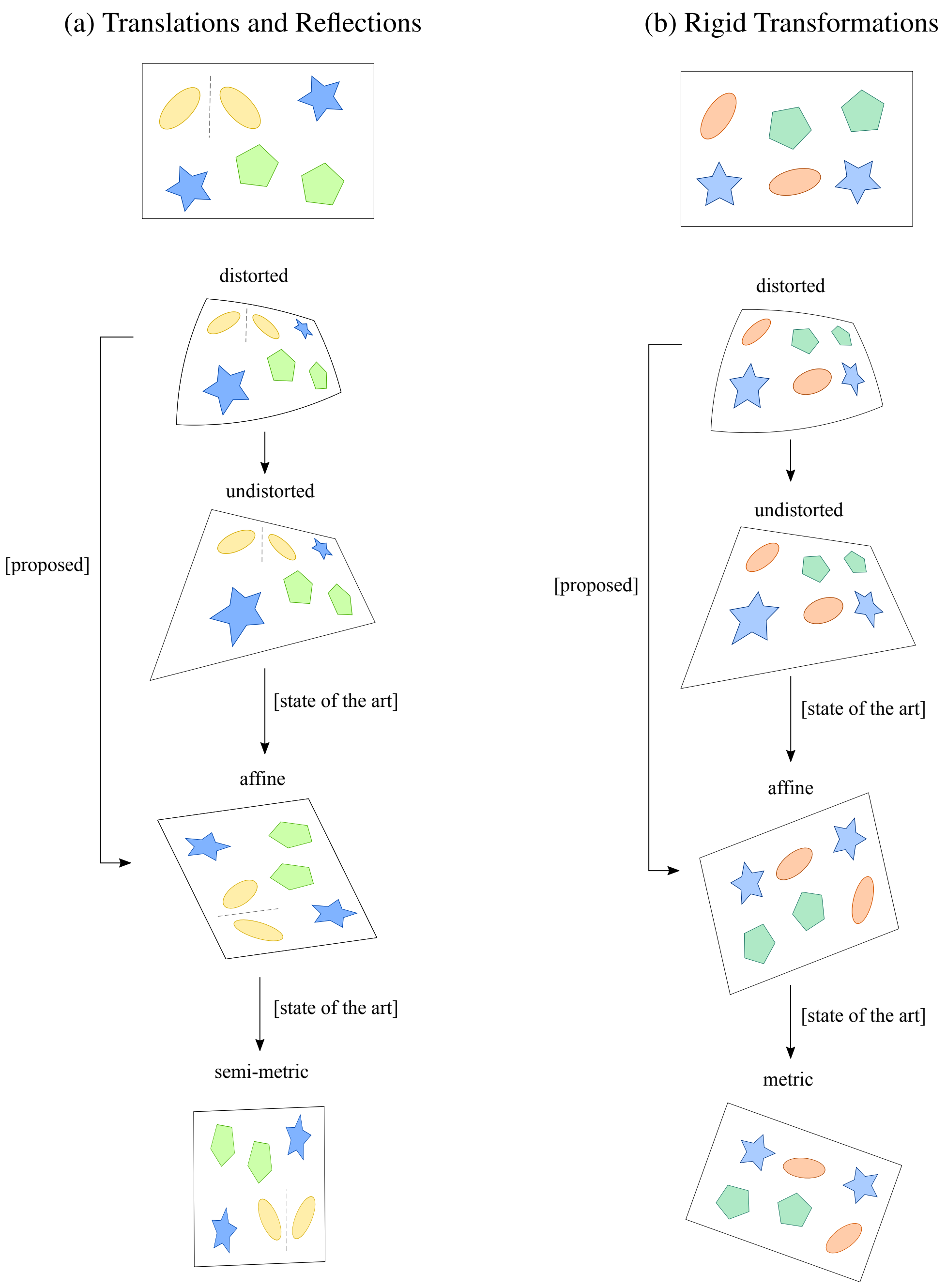


Figure 1.1: *Rectifications of Coplanar Repeated Texture.* The top row is a scene plane with (a) translated and reflected regions, which is the assumed configuration for the solvers of Chapter 5, and (b) rigidly-transformed regions, which is the assumed configuration for the solvers of Chapter 6. The state-of-the art requires an intermediate undistortion estimation, while the proposed solvers can directly affine-rectify from the distorted image of coplanar repeated texture. Affine-rectifications are metrically upgraded with the linear solvers introduced in Section 2.10.3.

1.1 Contributions

This thesis introduces the first minimal solvers of polynomial systems of equations for single-view geometry. In particular, the thesis introduces several novel methods for rectifying imaged scene planes from coplanar repeated texture.

The thesis derives novel constraints on lens undistortion and scene-plane rectification parameters using different configurations of radially-distorted conjugately-translated and reflected texture. The complexities of the generated solvers are compared with respect to the choice to eliminate particular unknowns from the polynomial system of equations in their derivations. One of the proposed solver variants can jointly undistort and rectify in only 0.5 microseconds.

In addition, the thesis generalizes the problem of lens undistortion and imaged scene-plane rectification to admit imaged rigidly-transformed coplanar repeats. In particular, derivations of constraints on rectification parameters that either directly use the undistorting and rectifying transform or its linearization are given. The solvers are generated with either elementary methods or the Gröbner basis method. A method adapted from [54] is used to sample feasible monomial bases to maximize the numerical stability of the solvers generated with the Gröbner basis method. The constraints derived from the linearized rectifying transform are used to calculate the dense relative change-of-scale due to the imaging of a scene plane, which gives the relative change of scale at any point of the imaged scene plane.

The code repository associated with this thesis at $\text{https://github.com/prittjam/repeats}$ provides solvers that cover all minimal configurations of these problems.

The thesis proposes several adaptations to the RANSAC robust estimation framework for the problem of rectifying imaged coplanar repeats [24]. In particular, a criterion for pre-empting consensus set construction is introduced for candidate solutions that are generated from measurements that provide redundant constraints. The pre-emptive strategy, called *best minimal solution selection*, eliminates the need to construct the consensus sets for all but one candidate solution. Best minimal solution selection significantly increases rectification accuracy compared to a strategy of random selection. In addition, a sequential two-stage RANSAC verification strategy is proposed that: (i) verifies that affine-rectified coplanar repeats are consistent with affine invariants, and (ii) metrically-upgraded coplanar repeats respect metric invariants. The combination of tests for scale consistency and congruence greatly increases the precision of covariant regions that are labeled as coplanar repeats and rectification accuracy.

The methods proposed in this thesis extend accurate rectification to a new class of imagery. Accurate rectificatoins of imaged scene planes from challenging wide-angle and fisheye images are achieved using the new solvers.

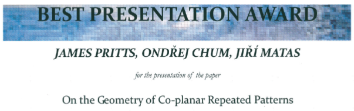
The thesis proposes and novel energy function for modeling scenes containing imaged coplanar repeated texture. The energy function is designed such that efficient inference can be achieved with state-of-the-art methods in discrete optimization. The global context of the energy function enables accurate scene plane segmentation and rectification of scene containing multiple planes and lacking a dominant plane. Rectifying solvers can easily be integrated into the minimization framework to provide scene plane models. A challenging dataset is introduced that is used for quantitative evaluation against the state of the art.

1.2 Publications

The content of this thesis is based on the material from the following articles published during the time of the PhD candidacy,

- [74] J. Pritts, O. Chum, and J. Matas. Detection, rectification and segmentation of coplanar repeated patterns. In *CVPR*, 2014.

  The article *Detection, Rectification and Segmentation of Coplanar Repeated Patterns* won the Computer Vision Winter Workshop 2014 *Best Presentation* Award.



BEST PRESENTATION AWARD

JAMES PRITTS, ONDŘEJ CHUM, JIŘÍ MATAS

for the presentation of the paper

On the Geometry of Co-planar Repeated Patterns

- [77] J. Pritts, D. Rozumnyi, M. P. Kumar, and O. Chum. Coplanar repeats by energy minimization. In *BMVC*, 2016.

- [75] J. Pritts, Z. Kukelova, V. Larsson, and O. Chum. Radially-distorted conjugate translations. In *CVPR*, 2018.

- [76] J. Pritts, Z. Kukelova, V. Larsson, and O. Chum. Rectification from radially-distorted scales. In *ACCV*, 2018.

  The article *Rectification from Radially-Distorted Scales* won the Saburo Tsuji *Best Paper* Award at the 14th Asian Conference on Computer Vision (ACCV) 2018.

ACCV 2018

2 – 6 December 2018 Perth Western Australia

Saburo Tsuji Best Paper Award

Rectification from Radially-Distorted Scales

James Pritts

Zuzana Kukelova

Viktor Larsson

Ondrej Chum

Sponsored By

Springer

- [80] J. Pritts, Z. Kukelova, V. Larsson, Y. Lochman, and O. Chum. Minimal solvers for rectifying from radially-distorted scales and change of scales, 2019. arXiv: 1907.11539 [cs.CV].

  The journal article *Minimal Solvers for Rectifying from Radially-Distorted Scales and Change of Scales* is accepted to The International Journal of Computer Vision (IJCV).

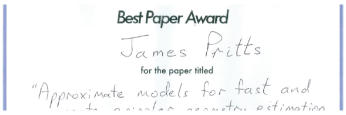
- [79] J. Pritts, Z. Kukelova, V. Larsson, Y. Lochman, and O. Chum. Minimal solvers for rectifying from radially-distorted conjugate translations. In 2019. arXiv: 1911.01507 [cs.CV].

  The journal article *Minimal Solvers for Rectifying from Radially-Distorted Conjugate Translations* was submitted for review to IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).

The following publication was from the time of the PhD candidacy but is not included in the thesis. The publication was omitted because the content is irrelevant to the thesis topic.

- [78] J. Pritts, O. Chum, and J. Matas. Approximate models for fast and accurate epipolar geometry estimation. In *IVCNZ*, 2013.

  The article *Approximate Models for Fast and Accurate Epipolar Geometry Estimation* won the Image and Vision Computing New Zealand 2013 (IVCVNZ) *Best Paper* Award.



Best Paper Award

James Pritts

for the paper titled

"Approximate models for fast and

epipolar geometry estimation

1.3 Structure of the Thesis

Radially-Distorted Cameras Viewing a Scene Plane Chapter 2 gives denotations, terminology and concepts from single-view geometry that unify the novel solvers proposed in Chapter 5 and 6. In particular, Chapter 2 relates the problem of imaged scene-plane rectification to the pre-imaging operation of a projective camera with radial lens distortion that can be parameterized with the division model [26]. Fundamental properties of the real-projective plane are discussed, which are prerequisites for following the derivations of the proposed minimal solvers. State-of-the-art techniques for computing affine and metric rectification are discussed. Linear solvers for computing the metric upgrade are derived. While this was a contribution of the author prior to the PhD, the proposed method for computing metric upgrades has been extended and used in several novel contexts.

In addition, Chapter 2 introduces the *warp error*, which is a novel measure of rectification accuracy that is used in several performance benchmarks in Chapter 5, 6, and 7 to compare the proposed methods against the state-of-the-art methods discussed in Chapter 2.

The Correspondence Problem for Imaged Coplanar Repeats Chapter 3 formalizes the concept of a coplanar repeat. Coplanar repeats are related to planar symmetry groups such as frieze and wallpaper groups and rotational symmetries. A local feature detection, description, and matching pipeline is proposed for the tentative identification of coplanar repeats. Ultimately, the tentative correspondences of image patches are used to induce constraints on parameters for radial lens undistortion and scene plane rectification. Since the proposed feature pipeline must be robust to viewpoint change, lighting change, sensor noise and occlusions, the formal definition of a coplanar repeat is relaxed so that the problem formulation can be posed in the context of the local-feature representations that are extracted by the proposed pipeline.

Solving Systems of Polynomial Equations Chapter 4 provides the prerequisites in algebraic geometry that are required to derive and generate the solvers introduced in Chapters 5 and 6. In particular, the Gröbner bases method ant the hidden-variable trick are discussed in detail.

Rectifying from Radially-Distorted Conjugate Translations Chapter 5 explores novel problem formulations for the affine rectification of imaged translated coplanar repeats (*i.e.*, radially-distorted conjugate translations) and imaged reflections. Furthermore, Chapter 5 establishes the connection between joint undistortion and affine rectification and radially-distorted conjugate translations, and several statements that are necessary to derive the solvers are proved. The chapter shows how covariant regions extracted from radially-distorted conjugately-translated and reflected texture can be used to satisfy the derived constraints.

Rectifying from Radially-Distorted Scales and Change of Scales Chapter 6 explores novel problem formulations for the affine rectification of imaged rigidly-transformed coplanar repeats. Furthermore Chapter 5 derives how the scale constraint—that two instances of rigidly-transformed coplanar repeats occupy identical areas in the scene plane and in the affine rectified image of the scene plane—can be used to affinely rectify imaged rigidly-transformed coplanar repeats.

Each of Chapters 5 and Chapter 6 provides: (i) a detailed analysis of the degeneracies of the solvers, (ii) experiments that evaluate the stability, noise sensitivity, and wall-clock time to solution of the proposed solvers with respect to a bench of the state-of-the-art solvers, and (iii) rectifications using the new solvers on challenging images.

Coplanar Repeats by Energy Minimization Chapter 7 introduces a global energy function for modeling scenes containing coplanar repeated texture. An energy minimization framework is described, which is a block-coordinate descent that alternates between labeling and regression problems. Results using a rectifying solver on scenes containing multiple planes are presented. A new annotated dataset is introduced that enables quantitative evaluation against the state of the art.

1.4 How to Read the Thesis

The prerequisites in single-view geometry and algebraic geometry that are needed for understanding the derivations of the the minimal solvers proposed in this thesis are contained in Chapters 2 and 4. The concept of a coplanar repeat and the types of local features that are used for their representation are detailed in Chapter 3. Chapters 5 and 6 introduce novel methods of jointly undistorting an rectifying imaged scene planes from coplanar repeated patterns. Chapter 7 proposes the energy function that formulates a global scene model for how coplanar repeated textures are imaged. The minimization of the energy using model proposals from recityfing solvers is detailed. Any of the state-of-the-art or proposed solvers can be used to hypothesize the model proposals.

Knowledge of the real-projective plane on the level of [32] will help with the understanding of the content of the thesis. This thesis takes an algebraic approach to planar geometry, meaning that geometric primitives are parameterized in terms of coordinates and algebraic entities. *E.g.*, for concision of language, a point is synonymous with a vector with respect to some basis; a line is also a vector, and a conic is a symmetric matrix.

1.4.1 State of the Art

The state of the art is not explicitly broken out as a chapter. Rather it is mostly grouped with the relevant topics in Chapters 2, 3, and 4.

1.5 Authorship

I, James Pritts, hereby certify that the results presented in this thesis are my novel research, which was done with the cooperation of my thesis advisors Ondřej Chum [77, 74, 75, 76, 80, 79] and Zuzana Kukelova [75, 76, 80, 79]. I am also grateful for the scientific collaboration of coauthors Viktor Larsson [75, 76, 80, 79], Pawan Kumar [77], Denys Rozumnyi [77], Yaroslava Lochman [80, 79] and Jiři Matas [74].

2 Radially-Distorted Cameras Viewing a Scene Plane

This chapter introduces the single-view geometry that will be needed to model cameras viewing coplanar repeated patterns. The affine and metric rectifying homographies will be introduced as well as rectification under lens distortion. State of the art methods for computing rectifying homographies will be reviewed.

2.1 Notation

In this section we provide a brief review of the notations that are used to formulate rectifying solvers in this thesis. The concepts denoted here will be introduced throughout this chapter, but a comprehensive introduction to the notation is given here as a reference to the reader.

For most of the text, points are modeled with homogeneous coordinates and are denoted $\mathbf{x}_i = \left(x_i, y_i, 1\right)^\top$, where x_i, y_i are the image coordinates. For particular derivations such as for the linear solvers for a metric upgrade in Section 2.10.3 or for the joint undistorting and affine rectifying change-of-scale solvers in Section 6.4, it is convenient to use inhomogeneous points, which are denoted in serif font as $\mathsf{x}_i = \left(x_i, y_i\right)^\top$. The affine-rectified images of homogeneous points and inhomogeneous points are denoted as $\underline{\mathbf{x}}_i = \left(\underline{x}_i, \underline{y}_i, 1\right)^\top$, and $\underline{\mathsf{x}}_i = \left(\underline{x}_i, \underline{y}_i\right)^\top$, respectively.

The image of a scene plane's vanishing line is denoted $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$ and the line at infinity is $\mathbf{l}_\infty = \left(0, 0, 1\right)^\top$. The phrase *vanishing point of the translation direction* is motivated by the fact that all imaged scene point correspondences translating in the same direction meet at a vanishing point. A vanishing point is denoted by either $\mathbf{u}$ or $\mathbf{v}$ that are the vanishing points of the translation directions $\mathbf{U}$ or $\mathbf{V}$ on the scene plane as imaged by $\mathtt{P}$, respectively. Matrices are in typewriter font; *e.g.*, an affinity is $\mathtt{A}$, a homography is $\mathtt{H}$, and a conjugate translation (also a homography) with vanishing point $\mathbf{u}$ is denoted $\mathtt{H}_\mathbf{u}$ (see Section 2.9.2). In general, a point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ is two points $\mathbf{x}$ and $\mathbf{x}'$ that are related by some geometric transformation. The notation is summarized in Table 2.1.

2.2 Solver Naming Convention

The naming convention for the rectifying solvers proposed in this thesis is based on the number of required correspondend regions and the unknowns returned by the solver. The minimal configuration of region correspondences is given as the subscript to $\mathtt{H}$ (denoting a homography);

Term	Description
$\mathtt{P}$	3×3 camera matrix viewing $z = 0$ (see (2.8)).
$\mathbf{X}$	homogeneous scene point or metric-rectified point in $\mathbb{RP}^2$
$\mathbf{x}, \tilde{\mathbf{x}}$	homogeneous rectilinear and distorted image point
$\underline{\mathbf{x}}$	homogeneous affine-rectified point (see (2.36))
X	inhomogeneous scene point or metric-rectified point
$\mathsf{x}, \tilde{\mathsf{x}}$	inhomogeneous rectilinear and distorted image point
$\underline{\mathsf{x}}$	inhomogeneous affine-rectified point
$\mathbf{x} \leftrightarrow \mathbf{x}'$	$\mathbf{x}$,$\mathbf{x}'$ are in correspondence with some transformation
$\mathbf{U}, \mathbf{V}$	translations in the scene plane
$\mathbf{u}, \mathbf{v}$	vanishing points of the trans. $\mathbf{U}$,$\mathbf{V}$ as imaged by $\mathtt{P}$
$\mathbf{m}_i$	join of undistorted point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$
$\mathbf{m}_{ij}, \mathbf{m}'_{ij}$	joins of $\mathbf{x}_i \leftrightarrow \mathbf{x}_j$ and $\mathbf{x}'_i \leftrightarrow \mathbf{x}'_j$, respectively
$[\cdot]_\times$	skew-symmetric operator for computing cross products
$\mathtt{T}$	homogeneous rigid-transformation matrix
$\mathbf{l}, \tilde{\mathbf{l}}$	image of vanishing line and distorted vanishing line
$\mathbf{l}_\infty$	the line at infinity
$\mathtt{H}$	affine-rectifying homography
$\mathtt{H}_\mathbf{u}$	conjugate translation in the imaged trans. direction $\mathbf{u}$
$\mathtt{A}$	an affinity, change of basis or the metric upgrade in metric rectification
λ	division model parameter for undistortion (see Section 2.11)
Π, π	the scene plane and image plane (in $\mathbb{RP}^2$)
$\tilde{\mathcal{R}}, \mathcal{R}, \underline{\mathcal{R}}$	distorted, undistorted, and affine-rectified regions

Table 2.1: *Common Denotations.*

e.g., a solver requiring 3 region correspondences is denoted $\mathtt{H}_{222}$. The unknowns that are recovered by the solver are suffixed to $\mathtt{H}_{\cdot}$. An additional superscript may be added to denote that the constraints used to derived the solver belongs to a family of solvers using similar constraints.

E.g., the solver of Chum et al. , which requires two region correspondences and returns the vanishing line, is denoted $\mathtt{H}^{\mathrm{CS}}_{22}\mathbf{l}$, where CS is used to denoted that it is a change-of-scale solver (see Section 2.9.1). The proposed solver in Section 5.4.2 requiring one region correspondence and returning the vanishing line $\mathbf{l}$ and division model parameter λ of lens distortion is denoted $\mathtt{H}_2\mathbf{l}\lambda$.

2.3 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

$$\left(x, y\right)^{\top} = \mathbf{h}(\left(X, Y, Z\right)^{\top}, \mathbf{z}), \tag{2.1}$$

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

The pinhole camera, also called the camera obscura, is perhaps 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. The following sections develop the algebraic relations that are sufficient to model this geometry.

Perspective Projection

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

$$\left(x, y\right)^{\top} = \frac{f}{Z}\left(X, Y\right)^{\top},$$

where $\left(X, Y, Z\right)^{\top}$ is the Euclidean representation of a scene point.

Perspective projection as defined in (2.3) 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. Using the homogeneous representation, perspective projection simply becomes

$$\alpha \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \mathrm{diag}(f, f, 1)\left[\mathtt{I}_3 \mid \mathbf{0}\right] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{2.2}$$

where $\alpha = 1/Z$.

For the following sections scene and image points will be modeled with homogeneous coordinates. This enables, *e.g.*, rigid transforms and perspective projections to be modeled as linear transformations, which simplifies the algebraic representation of the camera.

Camera Coordinate System

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

$$\begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^{\top} & 1 \end{bmatrix}, \tag{2.3}$$

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

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

$$\begin{bmatrix} a_x & a_x \cot\theta & p_x \\ 0 & a_y / \sin\theta & p_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.4}$$

where a_x and a_y are the scale factors of the image plane in units of pixels/mm, $\left(p_x, p_y\right)^\top$ is the principal point or optical center of the camera in pixels, and θ is the skew of the sensor.

The convention is to denote the intrinsics matrix as $\mathtt{K}$ and incorporate the scaling due to the focal length (see (2.2)),

$$\mathtt{K} = \begin{bmatrix} a_x & a_x \cot\theta & p_x \\ 0 & a_y / \sin\theta & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} k_x & k_c & p_x \\ 0 & k_y & p_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.5}$$

For a typical CCD camera with orthogonal raster and unit aspect ratio, the simplifications $k_x = k_y$ and $\theta = \pi/2$ can be assumed. For a pinhole camera, in addition to these typical constraints, we have $a_x = a_y = 1$.

Camera Matrix

Thus positioning and orienting the camera, projection, and the imaging transformation can be composed into a linear operation given by 3×4 camera matrix

$$\mathtt{P}^{3\times4} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix} = \mathtt{K} \begin{bmatrix} \mathtt{I}_3 \mid \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \mathtt{K} \begin{bmatrix} \mathtt{R} \mid \mathbf{t} \end{bmatrix} \tag{2.6}$$

Columns $\mathbf{p}_j$ have geometric meaning. Columns $\mathbf{p}_j$ where $j \in \{ 1 \dots 3 \}$ 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 Section 2.4.

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

$$\alpha \left(x, y, 1\right)^\top = \mathtt{P}^{3\times4} \left(X, Y, Z, 1\right)^\top. \tag{2.7}$$

The methods presented in this thesis work for affine cameras as well. There is nothing in the derivations that follow that preclude the use of affine cameras. However, affine cameras cannot move the line at infinity, and we are interested in modeling physical cameras with shorter focal lengths viewing scene planes at oblique angles, which is not a use case of affine cameras [20]. Thus, let us assume that the camera is modeled as one of the finite projective cameras defined above.

2.4 Camera Viewing a Scene Plane

Without loss of generality, a coplanar scene point $\left(X, Y, Z, 1\right)^\top$ is assumed to be on the scene plane $z = 0$. 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 $\mathbb{RP}^2$,

$$\underbrace{\alpha \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}}_{\mathbf{x}} \underbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix}}_{\mathtt{P}^{3\times 4}} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix}}_{\mathtt{P}} \underbrace{\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}}_{\mathbf{X}} . \tag{2.8}$$

The scene and image planes are denoted Π and π, respectively.

2.5 Affine-Rectifying Homography

Affine rectification restores affine invariants such as parallelism of lines and ratios of areas. An affine-rectifying homography $\mathtt{H}$ transforms the image of the scene plane's vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$ to the line at infinity $\mathbf{l}_\infty = \left(0, 0, 1\right)^\top$ [32]. Thus any homography $\mathtt{H}$ satisfying the constraint

$$\eta \mathbf{l} = \mathtt{H}^\top \mathbf{l}_\infty = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \eta \neq 0, \tag{2.9}$$

and where $\mathbf{l}$ is an imaged scene plane's vanishing line, is an affine-rectifying homography. Constraint (2.9) implies that $\mathbf{h}_3 = \mathbf{l}$, and that the image of the line at infinity is independent of rows $\mathbf{h}_1^\top$ and $\mathbf{h}_2^\top$ of $\mathtt{H}$. Thus, assuming $l_3 \neq 0$ [32], the affine-rectification of image point $\mathbf{x}$ to the affine-rectified point $\underline{\mathbf{x}}$ can be defined as

$$\begin{aligned} \alpha \underline{\mathbf{x}} &= \left(\alpha \underline{x}, \alpha \underline{y}, \alpha\right)^\top = \mathtt{H}(\mathbf{l}) \mathbf{x} \\ \text{s.t.} \quad \mathtt{H}(\mathbf{l}) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ & \mathbf{l}^\top & \end{bmatrix} \quad \text{and} \quad \alpha \neq 0. \end{aligned} \tag{2.10}$$

2.6 Metric-Rectifying Homography

Metric rectification restores metric invariants such as angles and ratios of lengths. Analogous to Section 2.5, where affine rectification is achieved by transforming the vanishing line to its canonical position at the line at infinity, metric rectification is achieved by transforming the imaged circular points to their canonical positions at $\left(1, \pm i, 0\right)^\top$. The vanishing line can be encoded as the join of the imaged circular points [16, 56]. Let the images of the circular points be $\mathbf{i}_\pi = \left(a+ib, c+id, 1\right)^\top$ and $\mathbf{j}_\pi = \left(a-ib, c-id, 1\right)^\top$. Then the join of the images of the circular points is scaled so that $\mathbf{l} = \frac{1}{2\pi}\left(\mathbf{i}_\pi \times \mathbf{j}_\pi\right) = \left(d, -b, bc-ad\right)^\top$. Thus an affine-rectifying homography can be constructed from the coordinates of the imaged circular points as

$$\mathtt{H}(\left(d, -b, bc-ad\right)^\top) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d & -b & bc-ad \end{bmatrix} \tag{2.11}$$

where the affine-rectified imaged circular points have the form

$$\mathtt{H}(\left(d, -b, bc-ad\right)^\top)\mathbf{i}_\pi = \xi\mathbf{i}'_\pi = \xi\left(a+ib, c+id, 0\right)^\top = (p+iq)\left(a+ib, c+id, 0\right)^\top, \tag{2.12}$$

where $\xi = (p+iq)$.

Furthermore, there exists an affine transformation, call it $\mathtt{A}^{-1}$, that moves the circular points from their canonical positions, namely $\mathbf{I} = \left(1, i, 0\right)^\top$ and $\mathbf{J} = \left(1, -i, 0\right)^\top$ to their transformed position in the affine-rectified space [16]. Let the parameterization of $\mathtt{A}^{-1}$ be

$$\mathtt{A}^{-1} = \begin{bmatrix} pa-qb & qa+pb & t_x \\ pc-qd & qd+pd & t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.13}$$

Note that $\det \mathtt{A}^{-1} = (p^2+q^2)(ad-bc)$, which is non-zero if either $p \neq 0$ or $q \neq 0$. Also note that the translation parameters t_x, t_y may be set arbitrarily since they have no effect on the ideal points or the invertibility of $\mathtt{A}^{-1}$. Unknowns q, t_x, t_y can be eliminated by setting $q = t_x = t_y = 0$. Setting $p = (ad-bc)^{-1}$ and inverting $\mathtt{A}^{-1}$ gives a parameterization of the metric-rectifying homography strictly in terms of the coordinates of the imaged circular points

$$\begin{aligned} \mathtt{H}_M(a,b,c,d) = \mathtt{AH} &= \begin{bmatrix} d & -b & 0 \\ -c & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d & -b & bc-ad \end{bmatrix} \\ &= \begin{bmatrix} d & -b & 0 \\ -c & a & 0 \\ d & -b & bc-ad \end{bmatrix}. \end{aligned} \tag{2.14}$$

Note that by QR decomposition, $\mathtt{A}$ can be replaced with $\mathtt{R}\,\mathtt{K}$, where $\mathtt{R}$ is a rotation matrix and

$\mathtt{K}$ is upper triangular so that

$$\mathtt{H}_M = \underbrace{\begin{bmatrix} \frac{d}{\sqrt{c^2+d^2}} & \frac{c}{\sqrt{c^2+d^2}} & 0 \\ \frac{-c}{\sqrt{c^2+d^2}} & \frac{d}{\sqrt{c^2+d^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathtt{R}} \underbrace{\begin{bmatrix} \sqrt{c^2+d^2} & -\frac{ac+bd}{\sqrt{c^2+d^2}} & 0 \\ 0 & \frac{ad-bc}{\sqrt{c^2+d^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathtt{K}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d & -b & bc-ad \end{bmatrix}}_{\mathtt{H}}. \tag{2.15}$$

Since metric rectification is invariant to rotation, $\mathtt{R}$ can be eliminated leaving $\mathtt{R}^\top\mathtt{H}_M = \mathtt{KH}$, where $\mathtt{R}^\top\mathtt{H}_M$ is metric rectifying. Up to a uniform scale factor, this coincides with the expression for $\mathtt{S}^{-1}\mathtt{P}^{-1}$ derived with the decomposition of (2.18) in the introduction,

$$\mathtt{R}^\top\mathtt{H}_M = \underbrace{\begin{bmatrix} \sqrt{c^2+d^2} & -\frac{ac+bd}{\sqrt{c^2+d^2}} & 0 \\ 0 & \frac{ad-bc}{\sqrt{c^2+d^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathtt{K}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d & -b & bc-ad \end{bmatrix}}_{\mathtt{H}}. \tag{2.16}$$

Since $\{\, \alpha\mathtt{R}^\top\mathtt{H}_M \,\}_{\alpha\neq 0}$ is the set of all metric-rectifying homographies, the metric upgrade $\mathtt{A}$ is upper-triangular only if the metric-rectifying homography is of the form $\alpha\mathtt{KH}$, where $\mathtt{K}$ $\mathtt{H}$ are strictly functions of the coordinate of the image of the circular points.

Suppose that $\alpha \neq 0$. By letting $a' = \sqrt{\alpha}a, b' = \sqrt{\alpha}b, c' = \sqrt{\alpha}c$ and $d' = \sqrt{\alpha}d$, homography $\alpha\mathtt{H}_M$ can be written as a function of the uniformly scaled coordinates of the imaged circular points as

$$\alpha\mathtt{H}_M = \mathtt{H}_M^\alpha = \mathtt{H}_M^\alpha(a',b',c',d') = \begin{bmatrix} d' & -b' & 0 \\ -c' & a' & 0 \\ 0 & 0 & \sqrt{\alpha} \end{bmatrix} \begin{bmatrix} \sqrt{\alpha} & 0 & 0 \\ 0 & \sqrt{\alpha} & 0 \\ d' & -b' & -\frac{a'd'-b'c'}{\sqrt{\alpha}}. \end{bmatrix} \tag{2.17}$$

From (2.17), tt can be seen that $\mathtt{H}_M^\alpha$ takes the form of $\mathtt{H}_M$ only if $\alpha \in \{\, -1, 1 \,\}$. In other words, $\mathtt{H}_M$ is inhomogeneous with respect to the parameters a, b, c, d.

2.7 Homography Decomposition

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

$$\mathtt{P} = \underbrace{\begin{bmatrix} s\mathtt{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{\mathtt{S}} \underbrace{\begin{bmatrix} \mathtt{A}^{2\times 2} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{\mathtt{A}} \underbrace{\begin{bmatrix} \mathtt{I}_{2\times 2} & \mathbf{0} \\ l_1 \quad l_2 & l_3 \end{bmatrix}}_{\mathtt{H}}, \tag{2.18}$$

where $l_3 \neq 0$, s is non-zero scalar, $\mathtt{R} \in SO(2)$ is a rotation, $\mathbf{t} \in \mathbb{R}^2$ 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 $l_3 \neq 0$ (see Hartley and Zisserman [32]).

Note that since a homography is invertible, (2.18) implies that $\mathtt{P}$ can be decomposed as the

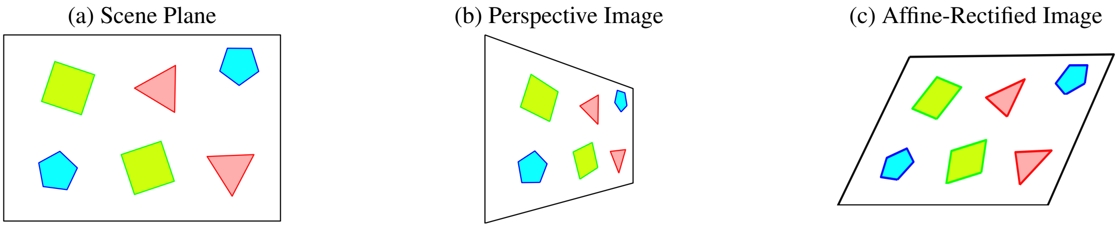


Figure 2.1: *The Equal-Scale Affine Invariant.* (a) Two instances of rigidly-transformed coplanar regions occupy identical areas in the scene plane. Rigidly-transformed coplanar regions are the same color. (b) Scene plane viewed by a perspective camera. (c) Affine-rectified image of the scene plane restores the invariant that rigidly-transformed coplanar regions occupy identical areas.

inverses of a similarity $\mathtt{S}'$, affinity $\mathtt{A}'$ and projectivity $\mathtt{H}'$ as $\mathtt{P} = \mathtt{H}'^{-1}\mathtt{A}'^{-1}\mathtt{S}'^{-1}$.

2.8 Pre-imaging and Rectification

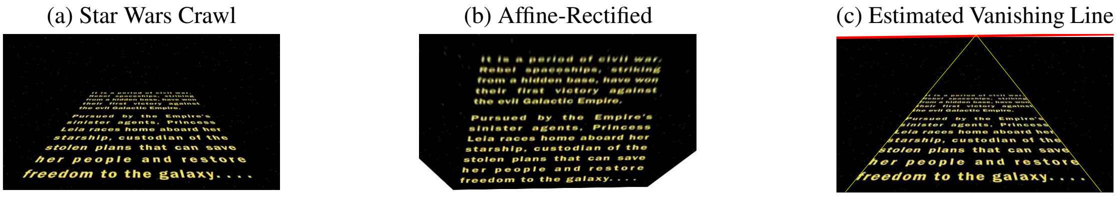
As shown in Section 2.7, the pre-imaging homography $\mathtt{P}^{-1}$ can be decomposed into a similarity $\mathtt{S}$, affinity $\mathtt{A}$, and projectivity $\mathtt{H}$ as $\mathtt{P}^{-1} = \mathtt{SAH}$. Metric rectification is invariant to similarity transformations [32]. Thus, $\mathtt{AH} = \mathtt{S}^{-1}\mathtt{P}^{-1}$ is also metric rectifying. Since the pre-imaging transform $\mathtt{P}^{-1}$ is homogeneous, it has eight degrees of freedom, four of which are eliminated by multiplying it with the similarity $\mathtt{S}^{-1}$. This leaves four degrees of freedom for the metric rectifying homography $\mathtt{AH}$, where the matrix $\mathtt{H}$ is the affine-rectifying homography, and the matrix $\mathtt{A}$ is the metric upgrade.

2.9 Computing Affine Rectification

The form of the affine-rectifying homography parameterized by the coordinates of the imaged vanishing line is derived in (2.10). The image of the vanishing line is typically estimated from affine invariants encoded by algebraic constraints that are a function of the unknown image of the vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$ [14, 20, 56, 82]. The following sections will review some state-of-the-art solvers for computing the vanishing line.

2.9.1 Change-of-Scale Solvers

The solvers introduced in this section exploit the scale constraint of affine-rectified space: two instances of rigidly-transformed coplanar regions occupy identical areas in the scene plane and in the affine-rectified image of the scene plane. The scale constraint is also called the *equal-scale invariant* of affine-rectified space. The invariant is shown in Figure 2.1. The equal-scale invariant of affine-rectified space is also used by the proposed solvers in Chapter 6 to extend affine rectification from minimal solver to lens-distorted images.


(a) Star Wars Crawl (b) Affine-Rectified (c) Estimated Vanishing Line

It is a period of civil war.
Rebel spaceships, striking
from a hidden base, have won
their first victory against
the evil Galactic Empire.
Pursued by the Empire's
sinister agents, Princess
Leia races home aboard her
starship, custodian of the
stolen plans that can save
her people and restore
freedom to the galaxy. . . .

Figure 2.2: *Affine Rectification from Change of Scale.* (a) Projective warp of the text placard gives the Star Wars crawl (b) The same letters cover the same area on the placard and are used to affinely rectify the crawl. (c) The estimated vanishing line used to construct the affine-rectifying homography is colored in red. Imaged parallel lines converge at a vanishing point on the vanishing line. This figure is taken from [14].

The change-of-scale solvers use the Jacobian determinant of the affine-rectifying transformation to induce constraints on the imaged scene plane's vanishing line. The Jacobian determinant measures the local change-of-scale of a differentiable transformation. The first solver to exploit the change-of-scale constraint for affine rectification was from Ohta et al. [71]. In fact, Ohta et al. did not explicitly derive their solver using a linearization of the rectifying transform, but arrived at an affine approximation for imaging local regions by geometric construction. The change-of-scale of is estimated from the local affine transformations of repeated coplanar texture.

Criminisi et al. [20] were the first to impose a constraint on the vanishing line from the Jacobian determinant; however, they did not use it to compute the vanishing line. Rather, they used the fact that the level sets of the derivative of the Jacobian of the rectifying transform are parallel to the direction of the vanishing line [20]. The vanishing line's position is recovered ex-post. Chum et al. [14] were the first to formulate a linear solver from the Jacobian determinant constraint. Imaged regions whose preimages are the same on the scene plane are used with the change-of-scale constraint to construct the solver. The derivation of the change-of-scale linear solver that follows unifies the derivations of Criminisi et al. and Chum et al. .

The inhomogeneous coordinates $(\underline{x}, \underline{y})^\top$ of the rectified point $\alpha\underline{\mathbf{x}} = \alpha\left(\underline{x}, \underline{y}, 1\right)^\top = \mathtt{H}\mathbf{x}$ (refer to (2.10)) of the imaged point $\mathbf{x} = (x, y, 1)^\top$ on the scene plane is given by the vector-valued nonlinear function

$$\underline{\mathbf{x}}(x, y) = \left(\underline{x}(x, y), \underline{y}(x, y)\right)^\top = \left(\frac{x}{\mathbf{l}^\top\mathbf{x}}, \frac{y}{\mathbf{l}^\top\mathbf{x}}\right)^\top .$$

The function $\underline{\mathbf{x}}$, which returns the inhomogeneous coordinates of the rectified point $\left(\underline{x}, \underline{y}\right)^\top$, can be linearized at $(x, y)^\top$ with the first-order Taylor expansion,

$$\underline{\mathbf{x}}(x + \delta_x, y + \delta_y) = \underline{\mathbf{x}}(x, y) + \mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l})|_{(x,y)} \cdot \left(\delta_x, \delta_y\right)^\top .$$

The Jacobian determinant $\det\left(\mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l})|_{(x_i,y_i)}\right)$ gives the approximate change of scale of the rectifying function $\underline{\mathbf{x}}$ near the point $(x, y)^\top$. Let s_i be the scale of an image region $\mathcal{R}_i$ with its centroid at $\left(x_i, y_i\right)^\top$, where the preimage $\underline{\mathcal{R}}_i$ of $\mathcal{R}_i$ is on some scene plane Π. Let $\underline{s}_i$ be the rec-

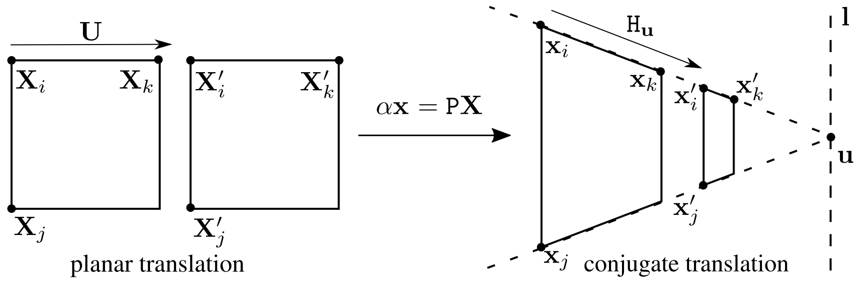


Figure 2.3: *Conjugate Translation.* A translation of coplanar scene points $\{\,\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k\,\}$ by $\mathbf{U}$ induces a conjugate translation $\mathtt{H}_\mathbf{u}$ in the undistorted image as viewed by camera $\mathtt{P}$. Joined conjugately-translated point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$ and $\mathbf{x}_k \leftrightarrow \mathbf{x}'_k$ must meet at the vanishing point $\mathbf{u}$. Vanishing line $\mathbf{l}$ is the set of all vanishing points of translation directions.

tified scale of $\underline{\mathcal{R}}_i$ and constrain the vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$ to the affine subspace $l_3 = 1$. Then the unknown rectified scale $\underline{s}_i$ can be expressed in terms of the imaged scale s_i and the Jacobian determinant as

$$\underline{s}_i = s_i \det\left(\mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l})|_{(x_i,y_i)}\right) = \frac{s_i}{(l_1x_i + l_2y_i + 1)^3}. \tag{2.19}$$

Let regions $\mathcal{R}_i$ and $\mathcal{R}_{i'}$ have the same preimage on the scene plane. Then $\underline{s}_i = \underline{s}_{i'}$ and the unknown rectified scale can be eliminated by setting equal

$$s_{i'}(l_1x_i + l_2y_i + 1)^3 = s_i(l_1x_{i'} + l_2y_{i'} + 1)^3, \tag{2.20}$$

which implies

$$(\sqrt[3]{s_{i'}}x_i - \sqrt[3]{s_i}x_{i'})l_1 + (\sqrt[3]{s_{i'}}y_i - \sqrt[3]{s_i}y_{i'})l_2 = \sqrt[3]{s_i} - \sqrt[3]{s_{i'}}. \tag{2.21}$$

Each pair of regions with the same preimage gives one constraint equation of the form (2.21). There are two unknowns, namely l_1, l_2, thus two pairs of regions with the same preimage are needed to solve for the orientation of the vanishing line. Note that in the overdetermined case, (2.21) is an algebraic least squares problem, so a whitening transform should be applied to the measurements [31]. An affine whitening transform will change all imaged regions by a global scale factor, which can be eliminated from (2.21). Finally, using the constraint $l_3 = 1$ gives the position of the vanishing line $\mathbf{l} = \left(l_1, l_2, 1\right)^\top$.

The set of solvers of [14, 20, 71] and the unifying derivation provided above is similar to the change-of-scale solvers that incorporate lens distortion that are proposed in Chapter 6. We call this group of solvers the *change-of-scale solvers* and acronymize them as (CS).

2.9.2 Conjugate Translations

Assume that the scene plane Π and a camera's image plane π are related point-wise by the camera $\mathtt{P}$ (see (2.8)) so that $\alpha\mathbf{x}' = \mathtt{P}\mathbf{X}'$, where α is a non-zero scalar, $\mathbf{X}' \in \Pi$ and $\mathbf{x}' \in \pi$. Furthermore, let $\mathbf{X}$ and $\mathbf{X}'$ be two points on the scene plane Π such that $\mathbf{U} = \mathbf{X}' - \mathbf{X} = \left(u_x, u_y, 0\right)^\top$. By encoding $\mathbf{U}$ in the homogeneous translation matrix $\mathtt{T}$, the points $\mathbf{X}$ and $\mathbf{X}'$ as imaged by camera $\mathtt{P}$ can be expressed as

$$\begin{aligned} \alpha\mathbf{x}' =& \mathtt{P}\mathbf{X}' = \mathtt{PT}\mathbf{X} = \mathtt{PTP}^{-1}\mathbf{x} = \mathtt{H}_\mathbf{u}\mathbf{x} \\ &\text{s.t.} \quad \mathtt{T} = \begin{pmatrix} 1 & 0 & u_x \\ 0 & 1 & u_y \\ 0 & 0 & 1 \end{pmatrix}, \end{aligned} \tag{2.22}$$

where the homography $\mathtt{H}_\mathbf{u} = \mathtt{PTP}^{-1}$ is called a conjugate translation because of the form of its matrix decomposition, and points $\mathbf{x}$ and $\mathbf{x}'$ are in correspondence (denoted $\mathbf{x} \leftrightarrow \mathbf{x}'$) with respect to the conjugate translation $\mathtt{H}_\mathbf{u}$, [32, 82].

Decomposing $\mathtt{H}_\mathbf{u}$ into its projective components gives

$$\begin{aligned} \alpha\mathbf{x}' = \mathtt{H}_\mathbf{u}\mathbf{x} =& \left[\mathtt{PI}_3\mathtt{P}^{-1} + \mathtt{P}\begin{pmatrix} u_x \\ u_y \\ 0 \end{pmatrix}\left[\mathtt{P}^{-\top}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right]^\top\right]\mathbf{x} \\ =& [\mathtt{I}_3 + s^\mathbf{u}\mathbf{u}\mathbf{l}^\top]\mathbf{x} \end{aligned} \tag{2.23}$$

where $\mathtt{I}_3$ is the 3×3 identity matrix, and, also consulting Figure 2.3 to relate the unknowns to the geometry,

- line $\mathbf{l}$ is the imaged scene plane's vanishing line,
- point $\mathbf{u}$ is the vanishing point of the translation direction,
- and scalar $s^\mathbf{u}$ is the magnitude of translation in the direction $\mathbf{u}$ for the point correspondence $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ [82].

Detection and Grouping of Repeated Elements Using Conjugate Translations

The method of Schaffalitzky and Zisserman [82] uses constraints induced by conjugate translations to detect and group repeated coplanar scene content in images. A conjugate translation is proposed in a hypothesize and verify framework, where at any point during sampling, the verified conjugate translation is the proposal consistent with the largest subset of an imaged lattice. The vanishing points of the imaged lattice are used to recover the vanishing line and to construct conjugate translations. The proposed conjugate translations are used in a guided search to discover conjugately-translated texture.

The use of translated planar scene texture to estimate conjugate translations is similar to method proposed in Chapter 5, which formulates the concept of a radially-distorted conjugate translation and proposes solvers to estimate it.

2.10 Computing Metric Rectification

The metric-rectifying homography, as parameterized by the coordinates of the image of the circular points, is given in (2.17) and 2.16. Metric rectification is either directly computed or estimated as a metric upgrade to an affine rectification. Metric invariants are expressed as algebraic constraints, which are typically parameterized as the image of the circular points as their unknowns or the conic dual to the image of the circular points, from which the image of the circular points can be recovered [32]. The next sections discuss methods from the state-of-the-art for computing metric-rectifying homographies.

2.10.1 Estimation From Imaged Orthogonal Line Segments

The parameters of $\mathtt{H}_M$, namely the coordinates of the imaged circular points, can be estimated linearly from correspondences of imaged orthogonal line segments [56, 32]. The conic dual to the circular points is $\mathtt{C}^*_\infty = \mathbf{IJ}^\top + \mathbf{JI}^\top$, where $\mathbf{I} = \big(1, i, 0\big)^\top$ and $\mathbf{J} = \big(1, -i, 0\big)^\top$, and $\mathtt{C}^*_\infty$ is a rank two matrix. The conic dual to the circular points as imaged by $\mathtt{P}$ is $\mathtt{PC}^*_\infty\mathtt{P}^\top$. Estimation of $\mathtt{C}'^*_\infty$ also determines the coordinates of the imaged circular points from which $\mathtt{H}_M$ can be constructed (see (2.14)). It can be shown that orthogonal lines are conjugate with respect to $\mathtt{C}'^*_\infty$, *i.e.*, if $\mathbf{m}$ and $\mathbf{m}'$ are images of orthogonal lines, then

$$\mathbf{m}^\top \mathtt{C}'^*_\infty \mathbf{m}' = 0. \tag{2.24}$$

Five orthogonal line pairs, where at least two pairs have different orientations, are sufficient to estimate $\mathtt{C}'^*_\infty$. If the rank two constraint is imposed on $\mathtt{C}'^*_\infty$, then four lines are sufficient, but the problem becomes nonlinear [56, 32].

If the affine rectifying homography $\mathtt{H}$ is known, then two degrees of freedom can be eliminated since the vanishing line is known. The affine-rectified image of the conic dual to circular points $\mathtt{HC}'^*_\infty\mathtt{H}^\top$ is rank two by construction, thus only two lines are needed to estimate the affine upgrade [56, 32].

2.10.2 Estimation From Imaged Circles

Two conics in general position whose preimages are circles intersect at the image of the circular points. The points of intersection of two ellipses can be found by solving a quartic, which gives two pairs of complex conjugate solutions. The pair that lies on the vanishing line of the scene plane are the image of the circular points. Once these coordinates are recovered, the metric rectifying homography is easily computed, *e.g.* by constructing $\mathtt{H}_M$ from (2.14).

This technique is not particularly useful. Typically there is not an abundance of circles in the scene. Unbiased estimation of ellipses requires a nonlinear least squares fit or a bias-renormalized fit such as the one proposed by Taubin [88]. Most importantly, ellipses do not distort to a simply parameterized closed curve under lens distortion, which complicates joint undistortion and rectification—a major topic of this thesis.

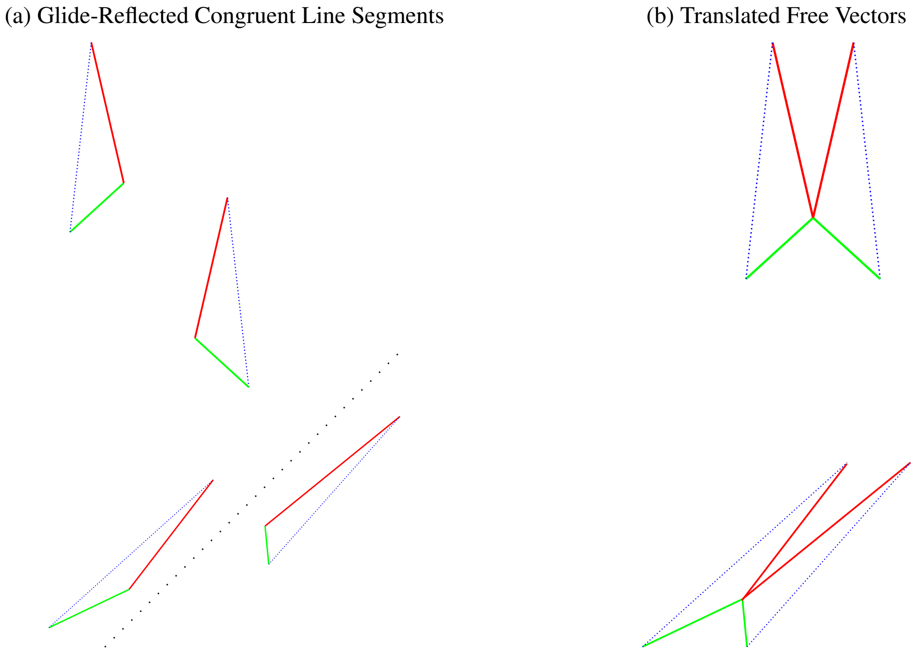

(a) Glide-Reflected Congruent Line Segments
(b) Translated Free Vectors

Figure 2.4: *Glide-Reflected Congruent Line Segments.* The first row is in a semi-metrically rectified frame, and the second row is in an affine frame. (a) Correspondences of line segments that are congruent in the scene are colored the same. (b) The line segments are interpreted as free vectors and are translated to the origin. (c) Corresponded congruent line segments are on the perimeter of the same circle in the world space and on the perimeter of the same ellipse in the affine-warped space.

2.10.3 Estimation from Equal Angles and Length Ratios

Metric rectification is typically computed as the metric upgrade of an affine rectification since fewer feature correspondences are required from this sequential estimation than directly estimating the metric rectification. The following paragraphs describe the process of upgrading an affine-rectified imaged scene plane to a metrically-rectified imaged scene plane using constraints on the metric upgrade induced by metric invariants.

Solving Quadratic Systems

Liebowitz et al. [56] derived quadratic constraints on the coordinates of the affine-rectified image of the circular points, from which the metric upgrade can be constructed (see (2.14)). Liebowitz et al. showed that pairs of line segments with 1. known angle 2. equal but unknown angle, and 3. known length ratio can be combined in a quadratic system of equations to determine the affine-rectified image of the circular points. The approach of Liebowitz only admits minimal samples, and thus precludes estimating the affine upgrade using additional measurements.

Linear Solvers

This thesis introduces the linear solvers for metric upgrades from glide-reflected and rigidly-transformed line segments, which are used throughout the thesis, especially in Chapters 5 and 6 to synthesize fronto-parallel views of affine-rectified imaged scene planes. These solvers are simple to implement and admit over-determined solutions.

Ratios of distance are not invariant to an affine transformation: the length of vectors in different directions are affected differently by an affine transformation. The derived constraints are linear, so they are simple to implement and efficient to use in a RANSAC-like robust estimator [24]. Fast estimation can be achieved from minimal sampling, or a more accurate least-squares solution can be obtained from many sets of line segments with the same lengths in the scene.

The length of corresponding imaged line segments will be used to design constraints for estimating the metric upgrade $\mathtt{A}$. Depending on the arrangement of the line segments, either a semi-metric or metric upgrade is possible. The semi-metric upgrade has a scale ambiguity in one direction. Let the semi-metric upgrade be denoted $\mathtt{A}$. Then the relation between the metric-rectified point $\mathbf{X}$, the affine-rectified point $\underline{\mathbf{x}}$, and the imaged point $\mathbf{x}$ is given by

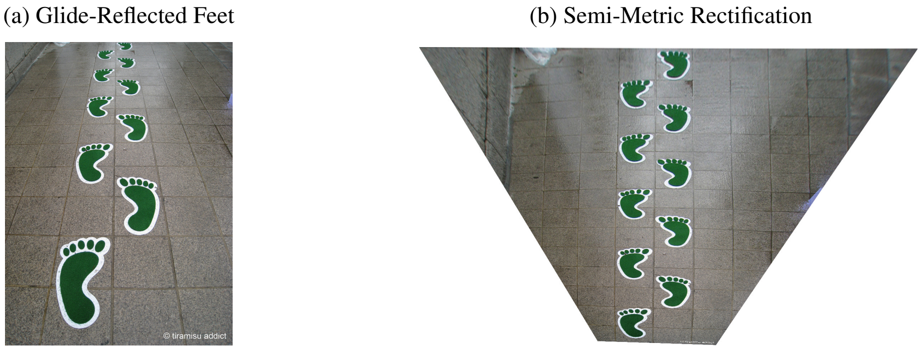
$$\alpha \mathbf{X} = \mathtt{A}\underline{\mathbf{x}} = \mathtt{AH}\mathbf{x}, \tag{2.25}$$

where $\mathtt{H}$ is an affine-rectifying homography.

For the derivations of the proposed upgrades, the use of the inhomogeneous representation of points will be more convenient. Let x be a inhomogeneous image point, $\underline{\mathsf{x}}$ be an affine-rectified inhomogeneous point and X be a metric-rectified or scene plane inhomogeneous point. Since transforming a line segment by a translation has no effect on its length, lengths of free vectors will be studied. Given a line segment $\overline{AB}$, the endpoint A is chosen as the origin of the affine frame and $\mathsf{x} = \big(x, y\big)^\top = B - A$ are the coordinates of the free vector defined by the line segment. This construction implies that the translation component of the unknown metric upgrade $\mathtt{A}$ need not be considered.

Axial Symmetry This paragraph examines the configuration of line segments that are reflected about an axial symmetry. Such a configuration frequently occurs on man-made objects, especially on building facades [97]. Let the coordinates of two free vectors constructed from glide-reflected line segments be denoted x and x'. By glide-reflection, we mean a symmetry operation that consists of a reflection over a line and then translated along that line. Without loss of generality, we assume that the axis of reflection is a vertical line on the scene plane. The geometry of computing a semi-metric upgrade from congruent glide-reflected line segments is shown in Figure 2.4. Thus, we are looking for an affine transformation $\mathtt{K} \in \mathbb{R}^{2\times 2}$ with rows $\mathbf{k}_1^\top$ and $\mathbf{k}_2^\top$ such that

$$\operatorname{diag}(-1, 1)\mathtt{K}\mathsf{x} = \mathtt{K}\mathsf{x}'. \tag{2.26}$$


(a) Glide-Reflected Feet

(b) Semi-Metric Rectification

Figure 2.5: *Semi-Metric Rectification from Glide-Reflections.* (a) Image containing a glide reflection. (b) Semi-metric rectification of floor from congruent line segments extracted from the feet. There is a scale ambiguity along the reflection axis.

This leads to a set of two homogeneous equations

$$\mathbf{k}_1^\top(\mathsf{x}+\mathsf{x}') \;=\; 0, \qquad \text{and} \tag{2.27}$$

$$\mathbf{k}_2^\top(\mathsf{x}-\mathsf{x}') \;=\; 0. \tag{2.28}$$

A single pair of points x, x' is enough to compute $\mathbf{k}_1$ and $\mathbf{k}_2$ up to a scalar factor. Any upgrade matrix $\mathtt{A}$ constructed such that

$$\mathtt{A} = \begin{bmatrix} \alpha_1\mathbf{k}_1^\top & 0 \\ \alpha_2\mathbf{k}_2^\top & 0 \\ 0 \quad 0 & 1 \end{bmatrix} \tag{2.29}$$

for non-zero scalars $\lambda_{1,2}$ will semi-metrically rectify the affine-rectified imaged scene plane. The rectification has an ambiguity of an anisotropic scaling along the direction of the axis of symmetry, the overall scale, and the rotation, which is fixed by aligning the axis of symmetry with a vertical line.

Rigidly Transformed We will assume that a scene plane has corresponded sets of rigidly-transformed line segments with each set having at least 2 lines. Let the sets be indexed by j, the length of a line in set j be r_j, and the coordinates of a free vector constructed from the rotated line segment i in set j on the scene plane be X_{ij}. Then the length constraint for scene or metric-rectified points can be written as

$$\mathsf{X}_{ij}^\top \mathsf{X}_{i'j} = r_j^2. \tag{2.30}$$

Substituting the metric upgrade of affine-rectified points for the scene points in (2.30) gives the constraint on the metric upgrade $\mathtt{K}$ as

$$\mathsf{X}_{ij}^\top \mathsf{X}_{i'j} = \underline{\mathsf{x}}_{ij}^\top \mathtt{K}^\top \mathtt{K} \underline{\mathsf{x}}_{i'j} = \underline{\mathsf{x}}_{ij}^\top \Sigma \underline{\mathsf{x}}_{i'j} = r_j^2, \tag{2.31}$$

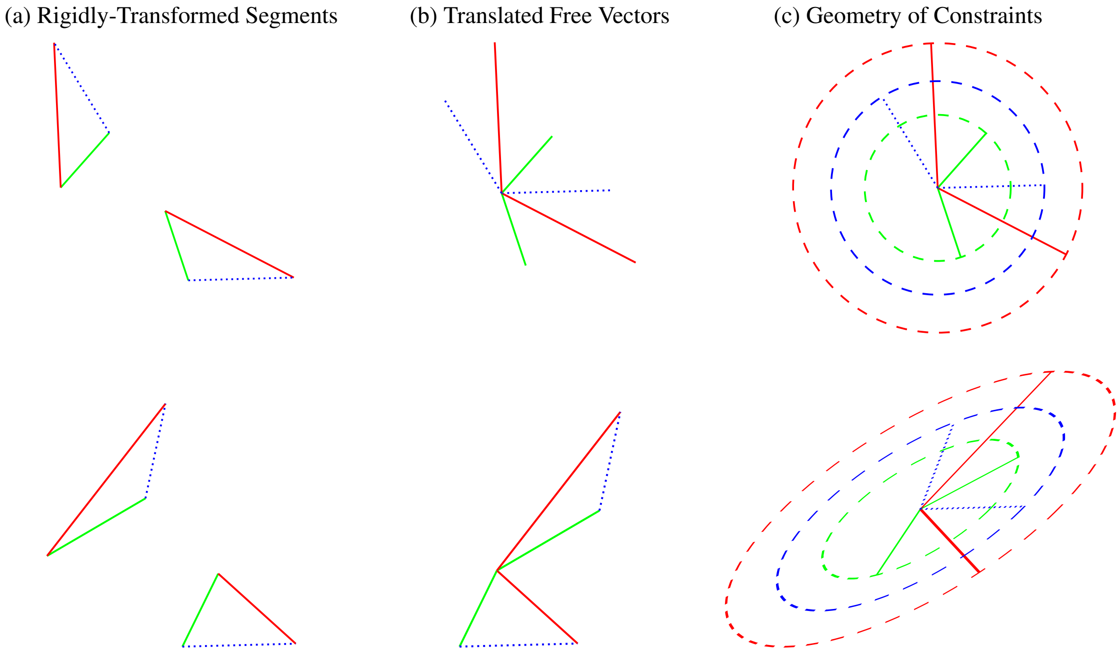

(a) Rigidly-Transformed Segments
(b) Translated Free Vectors
(c) Geometry of Constraints

Figure 2.6: *Rigidly-Transformed Congruent Line Segments.* The first row is in a Euclidean frame, and the second row is an affine frame. (a) Correspondences of line segments that are congruent in the scene are colored the same. (b) The line segments are interpreted as free vectors good. and are translated to the origin. (c) Corresponded congruent line segments are on the perimeter of the same circle in the Euclidean frame and on the perimeter of the same ellipse in the affine frame.

where $\Sigma = \mathtt{K}^T\mathtt{K}$. Solving for the symmetric matrix Σ instead of directly for $\mathtt{K}$ enables the problem to be formulated as algebraic least squares. The geometry of the problem formulation is depicted in Figure 2.6. The geometric constraint in the world space is that free vectors constructed from congruent line segments in the world space are on a circle with diameter r_j. In the affine-warped space, the free vectors are on an ellipse. Free vectors with the same length are color coded.

In equation 2.31, Σ is an ellipse (visualized in Fig. 2.6c), where

$$\Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \tag{2.32}$$

Equation (2.31) can be rewritten as

$$(x_{ij}^2 \quad 2x_{ij}y_{ij} \quad y_{ij}^2 \quad -1)(a \quad b \quad c \quad r_j^2)^\top = 0, \tag{2.33}$$

which gives a system of homogeneous linear equations. There are three unknowns for Σ, and each set of imaged congruent line segments adds one unknown r_j. Each line segment in general

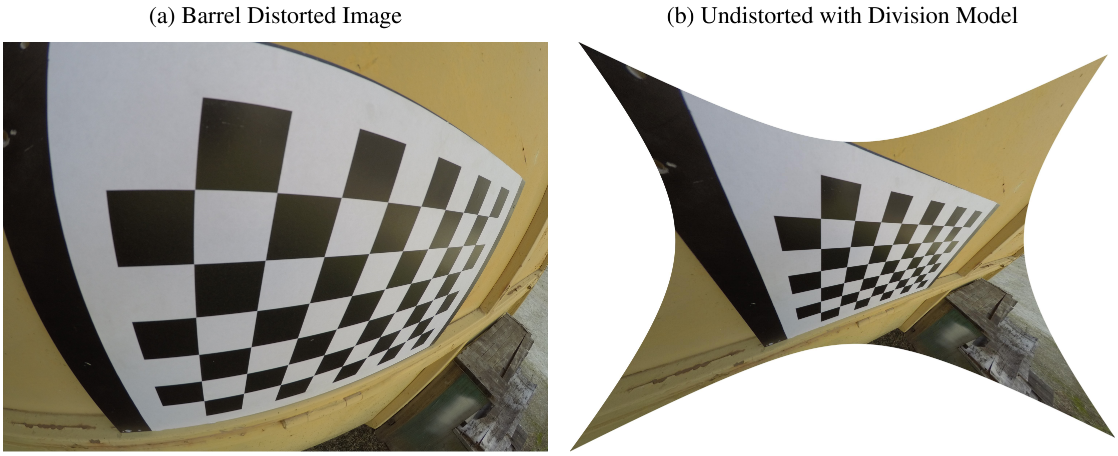

(a) Barrel Distorted Image
(b) Undistorted with Division Model

Figure 2.7: *Lens Undistortion.* (a) Chessboard captured with a GoPro Hero 4 (b) Image undistorted with the division model of lens undistortion. The one-parameter model is able to undistort the image such that the projective invariant that projected straight lines have to be straight is restored. [22].

position (rotation) adds one constraint. For two pairs of line segments, there are $3 + 2 = 5$ unknowns and four linear equations, giving a one-dimensional linear space of solutions. An alternative minimal solution is given by one triplet of reflected line segments, having $3 + 1 = 4$ unknowns and 3 linear equations. The affine transformation can be derived from the solution of the system of linear equations (2.33) up to a scale factor and a rotation. The unknown scale comes from the homogeneous nature of the system–both Σ and r_i^2s can be multiplied by a positive scalar. The unknown rotation comes from the ambiguity of the Cholesky decomposition $\Sigma = \mathtt{K}^\top\mathtt{K} = \mathtt{K}^\top\mathtt{R}^\top\mathtt{RK}$, where $\mathtt{R}$ is a rotation. The homogeneous metric upgrade $\mathtt{A}$ can be reconstructed from the upper triangular $\mathtt{K}$ factored from Σ as

$$\mathtt{A} = \begin{bmatrix} \multicolumn{2}{c}{\mathtt{K}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.34}$$

A rotation by 180 degrees (or an integer multiple) creates a special case: if the pattern is only rotated by integer multiplications of 180 degrees, then the matching vectors lie on parallel lines. Since affine transformations affect the lengths of vectors on parallel lines equally, the situation is similar to the pure translation case with full affine ambiguity.

2.11 Radial Lens Undistortion

Affine rectification as given in (2.10) 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

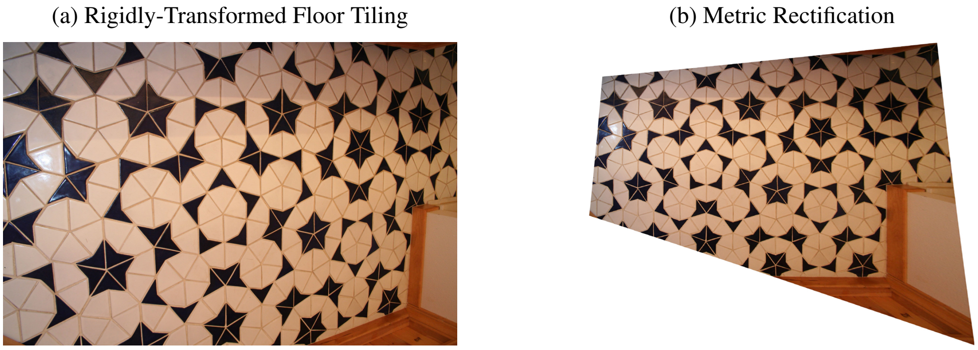


Figure 2.8: *Metric Rectification from Rigidly-Transformed Line Segments.* (a) Image containing rigidly-transformed line segments. (b) Metric rectification of floor from congruent line segments. Note that the room corner is restored to a right angle in the metric-rectified image.

undistortion function

$$\gamma \mathbf{x} = f(\tilde{\mathbf{x}}, \lambda) = \left(\tilde{x}, \tilde{y}, 1 + \lambda(\tilde{x}^2 + \tilde{y}^2)\right)^\top \tag{2.35}$$

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

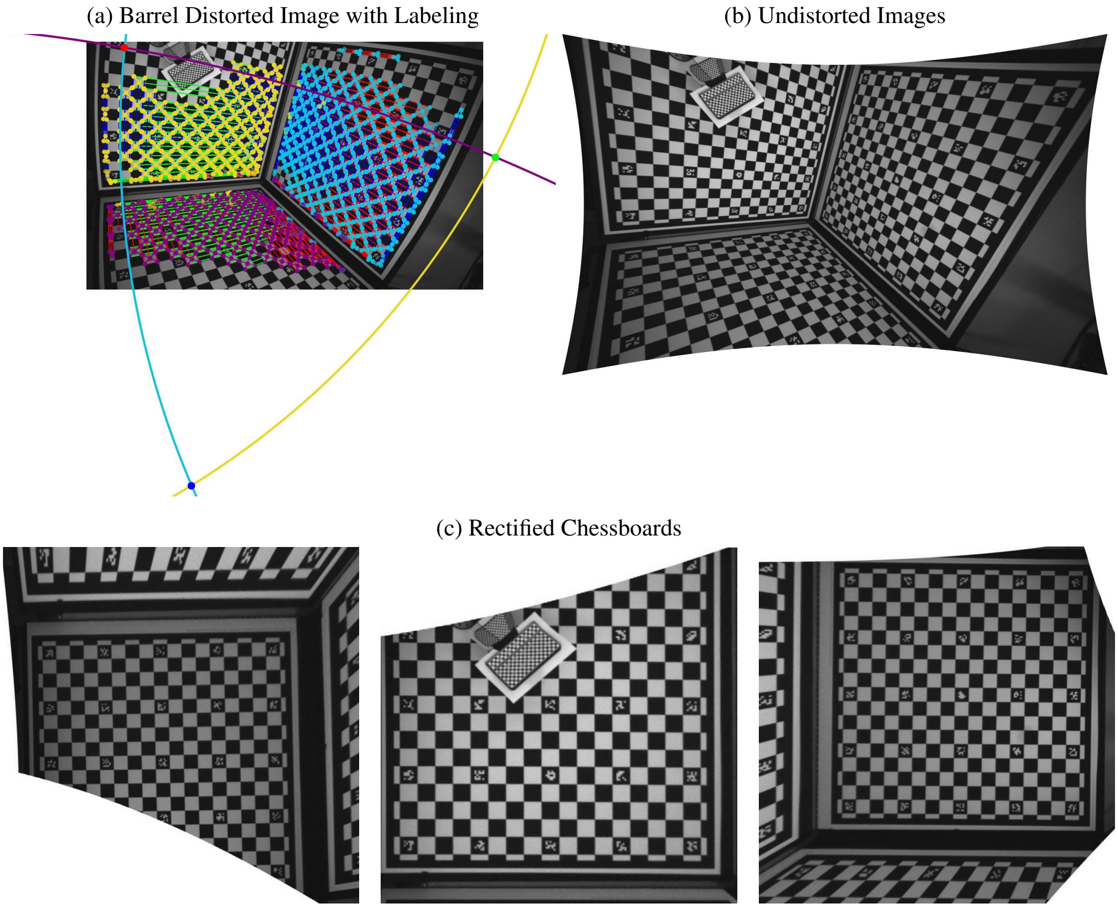
The strengths of this model were shown by Fitzgibbon [26] for the joint estimation of two-view geometry and non-linear lens distortion as given by (2.35). The division model is especially suited for minimal solvers since it is able to express a wide range of distortions (*e.g.*, see second row of Figure 5.3) 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. [95] and Fitzgibbon [26] showed that the precise positioning of the distortion center does not strongly affect image correction. Furthermore, we will see in the experiments of Chapter 5 and Chapter 6 that the proposed method is robust to deviations in the distortion center. Importantly, no constraints are placed on the location of the principal point of the camera by these assumptions, which is an influential calibration parameter [95]. However, the choice to fix the distortion center at the image center does make it difficult to remove a modeling degeneracy at the image center, which will be discussed in detail in Chapters 5.7 and 6.

2.11.1 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 (2.35) into (2.10), which gives

$$\begin{aligned} \alpha \underline{\mathbf{x}} = \left(\alpha \underline{x}, \alpha \underline{y}, \alpha\right)^\top = \mathtt{H}(\mathbf{l}) f(\tilde{\mathbf{x}}, \lambda) = \\ \left(\tilde{x}, \tilde{y}, l_1 \tilde{x} + l_2 \tilde{y} + l_3(1 + \lambda(\tilde{x}^2 + \tilde{y}^2))\right)^\top. \end{aligned} \tag{2.36}$$



(a) Barrel Distorted Image with Labeling

(b) Undistorted Images

(c) Rectified Chessboards

Figure 2.9: *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.

Interestingly, the rectifying function $\mathtt{H}(\mathbf{l})f(\tilde{\mathbf{x}},\lambda)$ in (2.36) also acts radially about the distortion center, but unlike the division model in (2.35), it is not rotationally symmetric.

Figures 5.3 and 5.7 render the distorted vanishing line in the source images, which affirm the accuracy of the rectifications by the proposed solvers.

2.11.2 Rectification From Distorted Parallel Scene Lines

Under the division model, the radially-distorted images of scene lines are circles [10, 26, 86, 92]. Antunes et al. [3] and Wildenauer et al. [94] are two methods that jointly undistort and rectify lens-distorted images using minimal solvers that admit circles fitted to the contours of imaged parallel scene lines.

The solver of Wildenauer et al. [94] requires five circular arcs, three of which are used to estimate the first vanishing point, which is formulated as a function of the division model parameters, and the remaining two arcs are undistorted to lines and used to compute the second vanishing point. The solver of Antunes et al. [3] estimates lens undistortion, distortion center and rectification from seven fitted circles, where the vanishing points are formulated as functions of the division model parameter and distortion center.

The requirement for sets of parallel scene lines is a strong scene content assumption. Chapters 6 and 5 introduce solvers that can jointly undistort and rectify from imaged coplanar repeated texture, which complements the arc based methods of Wildenauer et al. and Antunes et al. .

2.11.3 Radial-Distortion Homographies

The radial-distortion homography solvers jointly compute the full homography (eight degrees of freedom) with the lens undistortion parameters [26, 45]. These solvers are typically used to estimate the homography for panorama stitching or to estimate the change of basis between two lens-distorted cameras viewing the same scene plane, but they can also be used to estimate distorted imaged rigid transforms on the scene plane, which includes conjugate translations.

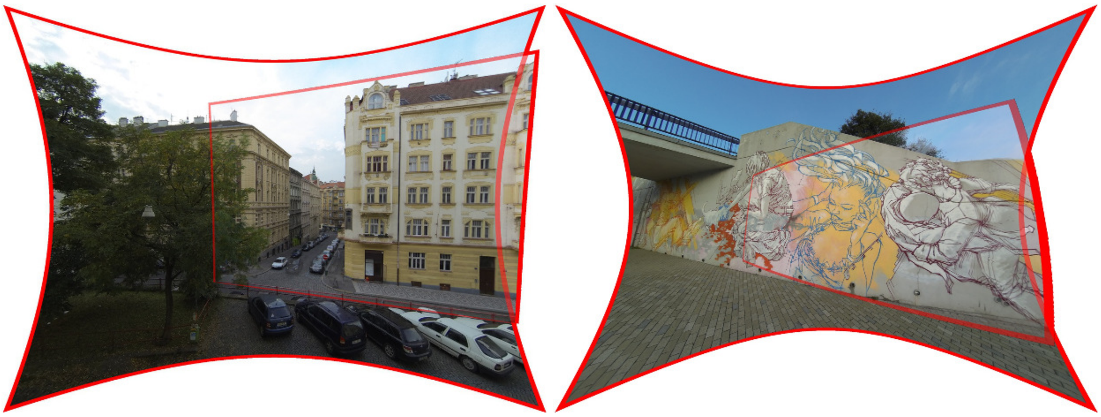


Figure 2.10: *Radial-Distortion Homography.* The images in each panorama were stitched using the radially-distorted homography estimated by the solvers proposed by Kukelova et al. [45]. This figure is taken from [45].

For estimating distorted imaged rigid transforms, the input to the radial-distortion homography solvers are five points that are consistent with the same rigid transform in the scene plane. The solver of Fitzgibbon et al. [26] returns one radial distortion parameter, while the solver of Kukelova et al. [45] returns two parameters (the assumption is that there can be two cameras with different distortions viewing the scene). In both these solvers radial distortion is modeled using the one parameter division model of (2.35).

The solver of Fitzgibbon et al. has nine degrees of freedom, while the solver of Kukelova et al. has ten. In contrast, a conjugate translation has only four degrees of freedom. If the division model is used for lens distortion then the distorted conjugate translation has five degrees of freedom. Thus the radial-distortion homography has an extra degree of freedom that can be used to fit noise in the measurements.

The radial-distortion homography solvers of Fitzgibbon et al. [26] and [45] are compared with the joint undistortion and affine-rectifying solvers proposed in Chapters 5 and 6 for estimating lens-distortion and radially-distorted conjugate translations.

2.12 Warp Error

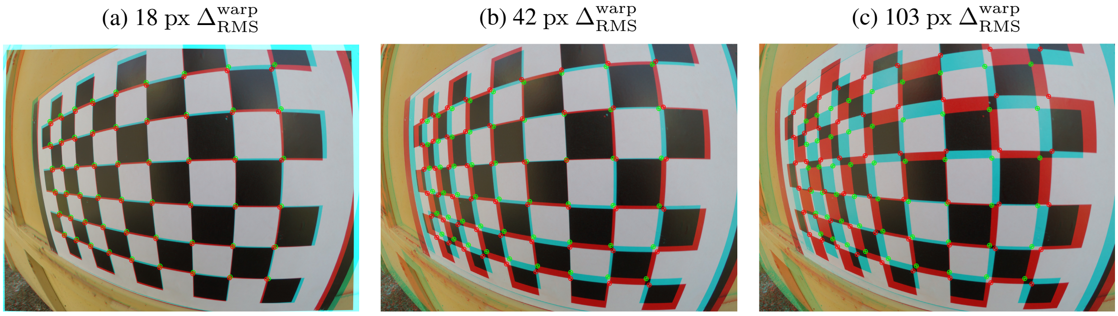
Since the accuracy of scene-plane rectification is a primary concern in this thesis, a warp error that jointly measures the accuracy of the estimated lens undistortion and rectifying homography is introduced. In synthetic experiments for virtual scenes, the scene plane is tessellated by a 10x10 square 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. For real images, extracted features on the segmented imaged scene plane are used to measure rectification accuracy.

A round trip between the image space and rectified space is made by affine-rectifying $\{\, \tilde{\mathbf{x}}_i \,\}_{i=1}^{100}$ using the estimated division model parameter $\hat{\lambda}$ and rectifying homography $\mathtt{H}(\hat{\mathbf{l}})$ (see (2.10)) and then imaging the rectified plane by the ground-truth camera $\mathtt{P}$. Ideally, the ground-truth camera $\mathtt{P}$ images the rectified points $\{\, \underline{\mathbf{x}}_i \,\}_i$ onto the distorted points $\{\, \tilde{\mathbf{x}}_i \,\}_i$. There is an affine ambiguity, denoted $\mathtt{A}$, between $\mathtt{H}(\hat{\mathbf{l}})$ and the ground-truth camera matrix $\mathtt{P}$. The ambiguity is estimated during computation of the warp error,

$$\Delta^{\text{warp}} = \min_{\mathtt{A}} \sum_i d^2(\tilde{\mathbf{x}}, f^d(\mathtt{PAH}(\hat{\mathbf{l}}) f(\tilde{\mathbf{x}}, \hat{\lambda})), \lambda), \tag{2.37}$$

where $d(\cdot,\cdot)$ is the Euclidean distance, f^d is the inverse of the division model (the inverse of (2.35)).

Rectification accuracy is reported as the RMS warp error computed over all grid points. The chessboard corners in Figure 2.11 provide the grid points from which $\Delta^{\text{warp}}_{\text{RMS}}$ is computed. Registration errors caused by inaccurate rectifications are shown as false colors in Figure 2.11. The warp error increases from left to right, which gives some geometric intuition regarding the magnitude of registration errors for different $\Delta^{\text{warp}}_{\text{RMS}}$ values on a 2250×3000 px pixel resolution image.



(a) 18 px $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ (b) 42 px $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ (c) 103 px $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$

Figure 2.11: *Warp Error Visualization.* Registration error of the ground-truth camera imaging the rectified scene plane is shown in false colors. The ground truth camera is estimated offline with a calibration toolbox that incorporates the division model. The warp error increases from left to right to give geometric intuition of how the registration error relates to $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$. The resolution of the image is 2250×3000 px.

3 The Correspondence Problem for Imaged Coplanar Repeats

In general, the goal of the correspondence problem is to match salient image regions that correspond to the same scene content. Region correspondences can be used to induce constraints on camera calibration and scene structure parameters and are used as inputs of estimators for single and multi-view geometry problems. The correspondence problem is challenging because of the effects of perspective warp, radial distortion and viewpoint change as well as scene variations caused by different lighting conditions and partial occlusions. In practice, correspondence is made between local regions that have a similar texture. Raw pixels are typically not used to measure texture; instead, a nonlinear transformation, usually called an embedding, transforms the region's texture into a high-dimensional vector (*e.g.*, 128-dimensional). An embedding is also called a descriptor in the computer-vision literature.

Region detectors are designed such that they are covariant to geometric transformations. Covariance with a geometric transformation ensures that the detection of a region warps with a warp of the region. Embeddings are designed or learned such that they are invariant to geometric and photometric transformations and discriminative, meaning that embeddings that describe distinctive scene content should be far away from each other in the embedded space. Invariance of the region embedding enables robust region matching across varying viewpoints and lighting conditions, while discriminability increases the likelihood that matched image regions correspond to repeated scene content.

These properties, covariance of detection and invariance and discriminability of description, are combined to robustly correspond coplanar repeats. The following sections detail these concepts and introduce the region detectors and embeddings that are used to identify and match imaged coplanar repeated scene texture.

In the context of imaged coplanar repeats, solving the correspondence problem means partitioning the set of region detections such that the preimages of the detections of a non-singleton subset are near copies of coplanar textures. See Figure 3.1 for an example. The singletons represent region detections of non-repeating texture. The correspondence problem for coplanar repeats differs distinctly from the two-view geometry setting in that two view correspondences are one-to-one whereas coplanar repeats have a many-to-many relationship. *E.g.*, for one set of n coplanar repeats, there are $\binom{n}{2}$ correspondences.

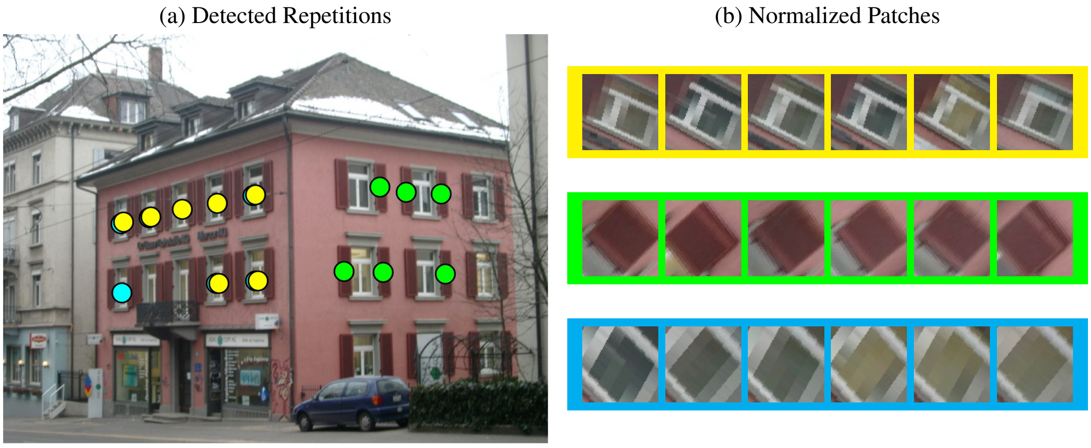


Figure 3.1: *Imaged Coplanar Repeats.* (a) The colored dots are the centroids of regions corresponded as imaged coplanar repeats. (b) The rows of same-colored boxes show corresponded sets of imaged coplanar repeats that are transformed to a normalized reference frame by an affinity. The nearly same appearances in the normalized frame show that coplanar repeats can be approximately registered by an affine transform. Note that the centroids of the blue and yellow-coded repeats are coincident, which is why the blue dots are occluded, but the region shape and extents are different.

3.1 What is a Coplanar Repeat?

Let $\{\,\mathbf{X}_i \leftrightarrow \mathbf{X}'_i\,\}_i$ be a set of coplanar point correspondences. Let $\mathtt{T}$ be a homogeneous rigid transform and define the rigid transformation of the point set $\{\,\mathbf{X}_i\,\}$ as

$$T_{\mathtt{T}}(\{\,\mathbf{X}_i\,\}) = \{\,\mathtt{T}\mathbf{X}_i \mid \mathbf{X}_i \in \{\,\mathbf{X}_i\,\}\,\}. \tag{3.1}$$

We say that $\{\,\mathbf{X}'_i\,\}$ is a *repeated point set* of $\{\,\mathbf{X}_i\,\}$ (and vice versa) if and only if there exists a homogeneous rigid transform matrix $\mathtt{T}$ such that $\{\,\mathbf{X}'_i\,\} = T_{\mathtt{T}}(\{\,\mathbf{X}_i\,\})$.

Let $\mathcal{R}_i$ and $\mathcal{R}_j$ denote coplanar regions that are connected and compact subsets of a scene plane Π. Then region $\mathcal{R}_j$ is a *coplanar repeat* if and only if it is a repeated point set (or simply a repeat) to region $\mathcal{R}_i$.

3.1.1 Imaged Rigid Transforms

Assume that the scene plane Π and a camera's image plane π are related point-wise by the homography $\mathtt{P}$, so that $\alpha\mathbf{x} = \mathtt{P}\mathbf{X}$, where $\alpha \neq 0$, $\mathbf{X} \in \Pi$ and $\mathbf{x} \in \pi$ (see Section 2.3 and Section 2.4). Let $\mathbf{X}$ and $\mathbf{X}'$ be two points on the scene plane Π such that $\mathbf{X}' = \mathtt{T}\mathbf{X}$, where $\mathtt{T}$ is a homogeneous rigid transform matrix.Then the points $\mathbf{X}$ and $\mathbf{X}'$ as imaged by camera $\mathtt{P}$ can be expressed as

$$\alpha\mathbf{x}' = \mathtt{P}\mathbf{X}' = \mathtt{P}\mathtt{T}\mathbf{X} = \beta\underbrace{\mathtt{P}\mathtt{T}\mathtt{P}^{-1}}_{\mathtt{H}_{\mathtt{T}}}\mathbf{x} \implies \gamma\mathbf{x}' = \mathtt{H}_{\mathtt{T}}\mathbf{x}, \tag{3.2}$$

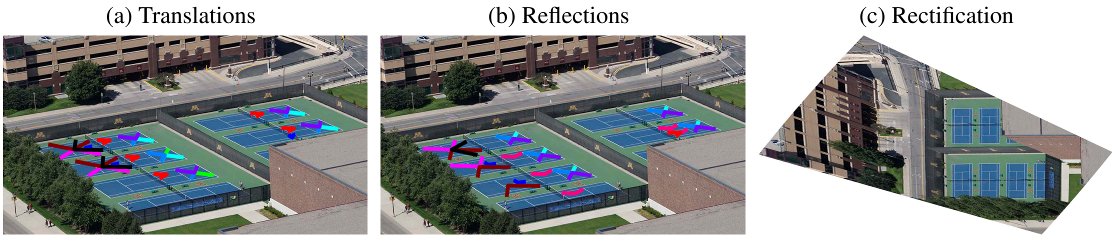


Figure 3.2: *Wallpaper Rectification.* The tennis court is a finite subset of a wallpaper pattern from which its symmetries are used to rectify the ground plane. (a) Translational symmetries are detected in the original image, (b) reflections are detected in the reflected image and then warped into the original image, and (c) The solvers of Chum et al. and Pritts et al. rectify the ground plane from symmetries [14, 74].

where $\beta \neq 0$ and $\gamma = \alpha/\beta$. We say that the points $\mathbf{x}$ and $\mathbf{x}'$ are in correspondence with respect to the *imaged rigid transform* defined by the homography $\mathtt{H}_\mathtt{T} = \mathtt{PTP}^{-1}$.

3.1.2 Radially-Distorted Imaged Rigid Transforms

Correspondence with an imaged rigid transform as derived in (3.2) is valid only if $\mathbf{x}$ is imaged by a rectilinear camera (see Section 2.3). Cameras always have some lens distortion, and the distortion can be significant for wide-angle lenses. In this text, the radial undistortion function is assumed to be the division model [26], which is defined in Section 2.11.

Incorporating radial undistortion into (3.2) gives the function of rigidly-transformed scene-plane points imaged by a lens-distorted camera

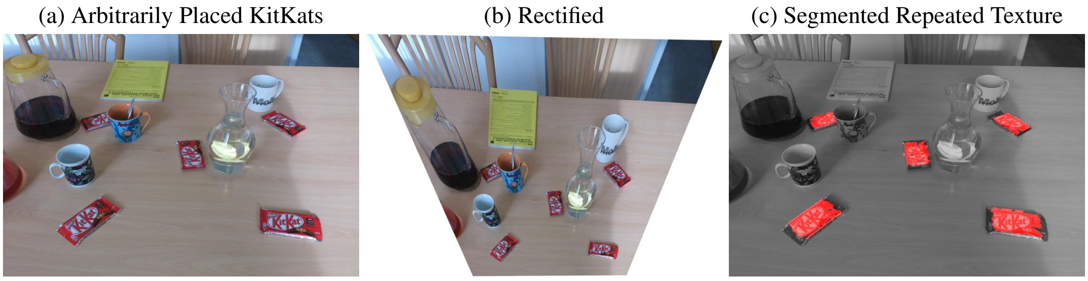
$$\gamma\tilde{\mathbf{x}}' = g(\tilde{\mathbf{x}}, \mathtt{H}_\mathtt{T}, \lambda) = f^d(\mathtt{H}_\mathtt{T} f(\tilde{\mathbf{x}}, \lambda), \lambda), \tag{3.3}$$

where f^d denotes the radial lens distortion function. In the following sections, $g(\tilde{\mathbf{x}}, \mathtt{H}_\mathtt{T}, \lambda)$ determines how scene points rigidly transformed by $\mathtt{T}$ and imaged by a radially-distorted camera are related.

3.1.3 Imaged Coplanar Repeats

Let $\tilde{\mathcal{R}}$ and $\tilde{\mathcal{R}}'$ denote regions in a lens distorted image. Thus if all $\tilde{\mathbf{x}} \in \tilde{\mathcal{R}}$ and $\tilde{\mathbf{x}}' \in \tilde{\mathcal{R}}'$ are corresponded by an imaged rigid transform $g(\cdot, \mathtt{H}_\mathtt{T}, \lambda)$, then the preimages of $\tilde{\mathcal{R}}$ and $\tilde{\mathcal{R}}'$, namely $\underline{\mathcal{R}}$ and $\underline{\mathcal{R}}'$, are related by the rigid transform $\mathtt{T}$, which is defined as a sufficient condition in Section 3.1 for the preimages $\underline{\mathcal{R}}$, $\underline{\mathcal{R}}'$ to be a coplanar repeat. This motivates the name *imaged coplanar repeat* for regions $\tilde{\mathcal{R}}$ and $\tilde{\mathcal{R}}'$ that are registered by the imaged rigid transform $g(\cdot, \mathtt{H}_\mathtt{T}, \lambda)$.

While the concept of a coplanar repeat is formalized, in practice, the definition will be relaxed to account for the limitations in the repeatability of the region detectors, measurement noise and simplifying modeling assumptions. The image patches in Figure 3.1b have been normalized to a



(a) Arbitrarily Placed KitKats

(b) Rectified

(c) Segmented Repeated Texture

Figure 3.3: *Rigidly Transformed Coplanar Repeats.* (a) Each pair of KitKats is registered by a unique rigid transform. (b) The solvers of Chum et al. and Pritts et al. admit inputs of rigidly-transformed coplanar repeats to metrically-rectify the scene plane [14, 74] (c) Segmented coplanar repeats are used to densely segment the KitKats from the background using the method of Cech et al. [13].

common reference plane to demonstrate how closely an approximation to $g(\cdot, \mathtt{H}_\mathtt{T}, \lambda)$ can register the coplanar repeats corresponded in Figure 3.1a.

3.1.4 Relating Planar Symmetries to Coplanar Repeats

In general, the definition of symmetry is relaxed by computer-vision practitioners. For example, a translational symmetry is a group law of a frieze or wallpaper group, which are unbounded sets, but the term is often applied to imaged translated coplanar repeats [28, 59, 58]. *E.g.*, a finite subset of a frieze or wallpaper pattern is not closed under the actions of a translational symmetry. The tennis courts in Figure 3.2 are a finite subset of a wallpaper group from which its translational symmetries in Figure 3.2a and reflections in Figure 3.2b are used to rectify the ground plane, which is shown in Figure 3.2c.

Coplanar repeats are closely related to symmetries. Let $\mathtt{T}$ be a planar symmetry of the set $\underline{\mathcal{S}}$ so that $\underline{\mathcal{S}} = T_\mathtt{T}(\underline{\mathcal{S}})$ and let $\underline{\mathcal{R}}_i \subset \underline{\mathcal{S}}$ and $\underline{\mathcal{R}}_j = \mathtt{T}\underline{\mathcal{R}}_i$. Clearly $\underline{\mathcal{R}}_j \subset \underline{\mathcal{S}}$ and $\underline{\mathcal{R}}_j$ is a coplanar repeat of $\underline{\mathcal{R}}_i$. This implies that the symmetry $\mathtt{T}$ can be recovered from coplanar repeats that are subsets of a symmetry group.

However, the images of coplanar repeats that are not members of a symmetry group also put strong constraints on undistortion and rectification and are useful for scene segmentation. Thus, the goal is to use the constraints induced by all imaged coplanar repeats. Suppose three coplanar repeats $\underline{\mathcal{R}}_i$,$\underline{\mathcal{R}}_j$ and $\underline{\mathcal{R}}_k$ are placed on a scene plane. Regions $\underline{\mathcal{R}}_i$,$\underline{\mathcal{R}}_j$ and $\underline{\mathcal{R}}_k$ are not necessarily in a symmetry group, but if their images $\mathcal{R}_i$, $\mathcal{R}_j$, and $\mathcal{R}_k$ are corresponded, then they can be inputted to a solver to compute the metric rectification of the imaged scene plane [14, 20, 71, 74, 82]. Figure 3.3 contains arbitrarily placed KitKats on a table. The imaged KitKats provide provided the necessary constraints as imaged coplanar repeats to metrically rectify (see Figure 3.3b).

3.1.5 Approximate Imaged Rigid Transforms

The Euclidean coordinates $\left(\tilde{x}', \tilde{y}'\right)^\top$ of an imaged rigidly-transformed point is given by the vector-valued nonlinear function

$$\tilde{\mathbf{x}}'(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda) = \left(\tilde{x}'(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda)/\gamma(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda), \tilde{y}'(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda)/\gamma(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda)\right)^\top, \tag{3.4}$$

where $\lambda\tilde{\mathbf{x}}' = \left(\tilde{x}'(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda), \tilde{y}'(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda), \gamma(\tilde{x}, \tilde{y}, \mathtt{H_T}, \lambda)\right) = g\left(\left(\tilde{x}, \tilde{y}, 1\right)^\top, \mathtt{H_T}, \lambda\right)^\top$. The function $\tilde{\mathbf{x}}'$, which returns the inhomogeneous coordinates of the imaged rigidly-transformed point $\left(\tilde{x}', \tilde{y}'\right)^\top$, can be linearized at $\left(\tilde{x}, \tilde{y}\right)^\top$ with the first-order Taylor expansion,

$$\tilde{\mathbf{x}}'(\tilde{x} + \delta_{\tilde{x}}, \tilde{y} + \delta_{\tilde{y}}, \mathtt{H_T}, \lambda) = \tilde{\mathbf{x}}'(\tilde{x}, \tilde{y}) + \mathtt{J}_{\tilde{\mathbf{x}}'}(\mathtt{H_T}, \lambda)\big|_{\left(\tilde{x}, \tilde{y}\right)} \cdot \left(\delta_{\tilde{x}}, \delta_{\tilde{y}}\right)^\top. \tag{3.5}$$

Let $\mathtt{A}$ denote the 2×2 affine matrix $\mathtt{J}_{\tilde{\mathbf{x}}'}(\mathtt{H_T}, \lambda)\big|_{\left(\tilde{x}, \tilde{y}\right)}$. Substituting $\mathtt{A}$ into (3.5) gives

$$\tilde{\mathbf{x}}'(\tilde{x} + \delta_{\tilde{x}}, \tilde{y} + \delta_{\tilde{y}}, \mathtt{H_T}, \lambda) = \tilde{\mathbf{x}}'(\tilde{x}, \tilde{y}) + \mathtt{A}\left(\delta_{\tilde{x}}, \delta_{\tilde{y}}\right)^\top. \tag{3.6}$$

Thus, if $\tilde{\mathbf{x}}$ in $\tilde{\mathcal{R}}$ and $\tilde{\mathbf{x}}'$ in $\tilde{\mathcal{R}}'$ are in correspondence with an imaged rigid transform $g(\tilde{\mathbf{x}}, \mathtt{H_T}, \lambda)$ and if the imaged coplanar repeats $\tilde{\mathcal{R}}$ and $\tilde{\mathcal{R}}'$ are sufficiently small, then all points in $\tilde{\mathcal{R}}$ and $\tilde{\mathcal{R}}'$ are approximately related by an affinity. This fact motivates the use of region detectors that covary with affine transformations of the image, which will be introduced in Section 3.2. Figure 3.1 demonstrates that an affine transformation is sufficient to approximately register imaged coplanar repeated textures, which empirically verifies (3.6). The image patches from the coplanar repeats detected in Figure 3.1a are affinely warped into a common reference frame in Figure 3.1b, where they have similar appearances. Another example is shown in the third row of Figure 3.6.

3.2 Covariant Regions

Suppose that $T(\cdot)$ is a warp that can be applied to image I. A region detector that is covariant with respect to T will extract regions $\{\, \mathcal{R}_1, \ldots \mathcal{R}_n \,\}$ from I and regions $\{\, T(\mathcal{R}_1), \ldots T(\mathcal{R}_n) \,\}$ from $T(I)$ (see Figure 3.4). The repeatability of a detector measures the amount the detector covaries with respect to a given transformation T [64]. A region is labeled repeatable with respect to transformation T if the Jaccard distance threshold between $T(\mathcal{R}_1)$ and $\mathcal{R}_2$ is less than some threshold,

$$1 - \frac{\mathcal{R}_1 \cap T(\mathcal{R}_2)}{\mathcal{R}_1 \cup T(\mathcal{R}_2)} < \epsilon, \tag{3.7}$$

The threshold ϵ controls the amount of tolerated overlap error with respect to a given transformation class. The repeatability of a detector is measured over many examples, and T is usually sampled from the same transformation class, *e.g.*, T is a homography. A visualization of contour extraction exhibiting good repeatability is shown in Figure 3.8b, while Figure 3.8c shows bad

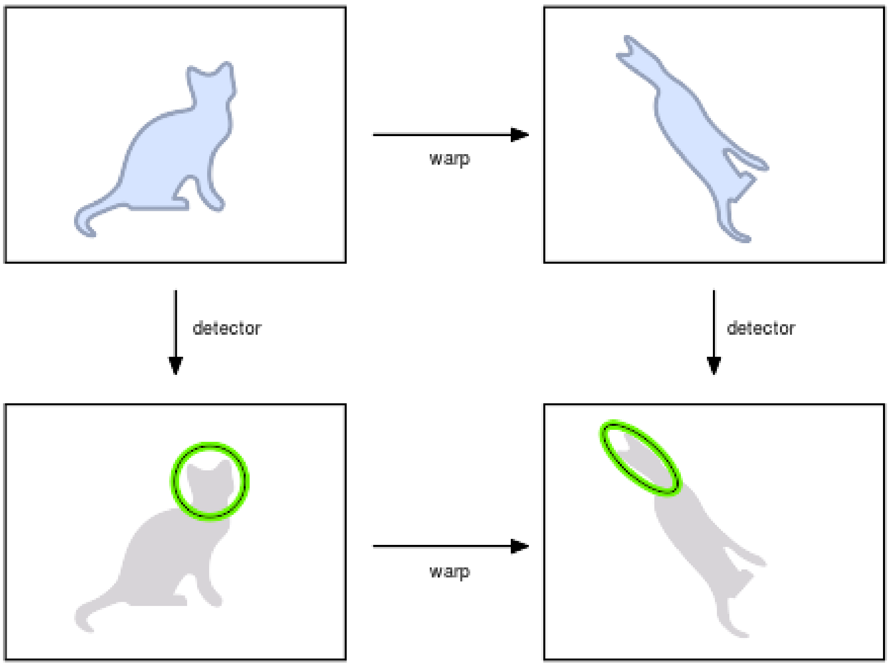


Figure 3.4: *Region Covariance.* An affine-covariant region detector is run on the original image I and its affine warp $T(I)$. The detected region $\mathcal{R}'$ in the warped image is the warp of the detection $T(\mathcal{R})$ in the original image. The figure is taken from [91].

repeatability.

The same region viewed from two cameras may appear radically different, and, indeed, the transformation that preimages $\mathcal{R}_1$ from the first camera and images it as $\mathcal{R}_2$ in the second camera is highly non-linear [23]. As shown in Section 3.1.5, imaging transformations can be approximated locally by linearizations, which typically have fewer degrees of freedom. Thus repeatability is expected to be higher for smaller regions, where the linearization has a low approximation error. The state-of-the art region detectors are covariant up to an affinity [64, 66].

The methods proposed in this thesis are agnostic to the specific type of covariant detector used, but all proposed methods have either similarity or affine-covariance as a necessary property. Section 3.2.3 discusses the parameterization used in this thesis to represent covariant regions. Ultimately, the covariant region parameterization is a compact representation for the local geometry of a coplanar repeated pattern.

Term	Description
$\underline{\mathcal{R}}$,$\mathcal{R}$,$\tilde{\mathcal{R}}$	affine-rectified, rectilinear and distorted covariant region detections

Table 3.1: *Notation for Covariant Regions.*

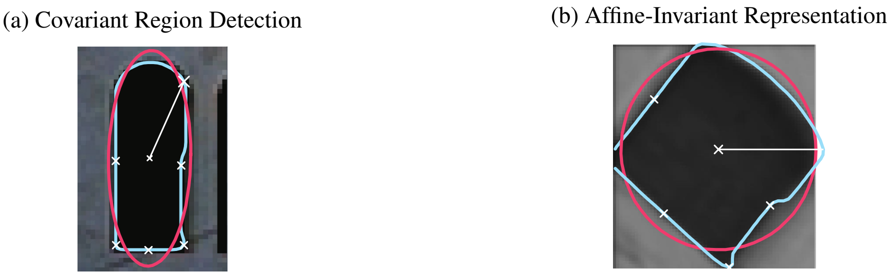

(a) Covariant Region Detection
(b) Affine-Invariant Representation

Figure 3.5: *Affine-Invariant Region Representation.* (a) The boundary of the low intensity region (cyan) is used to estimate the second moment matrix (red ellipse). The ellipse establishes the shape and extent of the covariant region $\mathcal{R}$. The most distant point from the center of gravity of the contour is used to fix the orientation. (b) The ellipse is warped to the unit circle and the orientation is aligned with the x-axis in the normalized frame. The figure is taken from [70].

3.2.1 Region Notation

Covariant regions are denoted $\mathcal{R}$ if an image is from a camera with a rectilinear lens, or $\tilde{\mathcal{R}}$ if an image is from a camera with a radially-distorted lens. In practice, the same detector is used in both cases, but a distinction is necessary to develop the theory in Chapters 5 and 6. The preimage and rectified image of an imaged covariant region are both denoted $\underline{\mathcal{R}}$. Going forward, region detections are assumed to have the similiarity or affince covariance property. The notation is summarized in Table 3.1.

3.2.2 Affine-Invariant Region Representation

Section 3.1.5 shows that imaged coplanar repeats can be accurately registered with affinities if they are sufficiently small. Thus coplanar repeats can be approximately generated as affine warps of a normalized image patch in a common reference frame, if photometric differences are ignored. This property simplifies the correspondence problem since the putative coplanar repeats will have approximately the same appearance in the common reference frame. To establish the common reference frame, an affine basis is constructed for each covariant region, which is mapped to the orthonormal basis at the origin, which is shown in Figure 3.5. Examples patches normalized from the imaged coplanar repeats are shown in Figures 3.1. Examples of different affine basis constructions from covariant regions are shown in Figure 3.6. More details about these constructions are discussed in Section 3.2.3.

3.2.3 Covariant Region Parameterization

An affine-covariant region $\mathcal{R}$ is defined by an affine basis in the image coordinate system, which is called a *local affine frame* (LAF) in the computer-vision literature. The local affine frame can be minimally parameterized by three points $\{\, \mathbf{x}_1,\, \mathbf{x}_2,\, \mathbf{x}_3 \,\}$, where $\mathbf{x}_2$ is designated as the origin of the affine basis. The three points are obtained from affine-covariant constructions, which are

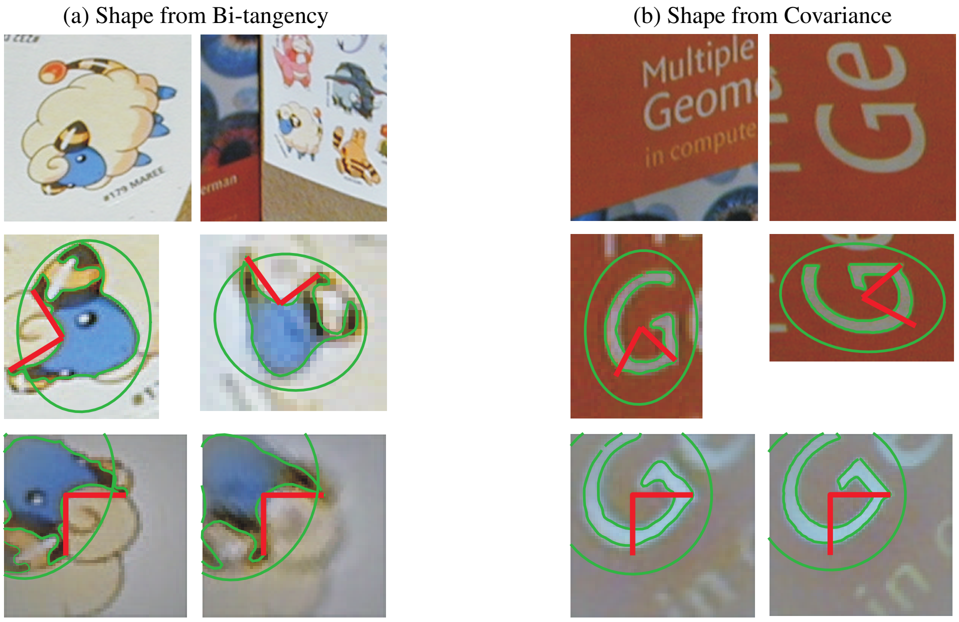


Figure 3.6: *Local Affine Frame Construction Type.* (top row) Original image, (middle row) MSER detection and local affine frame (LAF) construction, (bottom row) normalized patches computed from local affine frames. (a) LAFs are constructed from bi-tangents of the MSER detections (green contour). The LAF origin is the point in the concavity most distant from the bi-tangent. (b) The shape and extent of the LAF is computed from the second central moments (or covariance matrix) of the MSER detection and the orientation is fixed by a curvature extrema of the contour. Figure taken from [70].

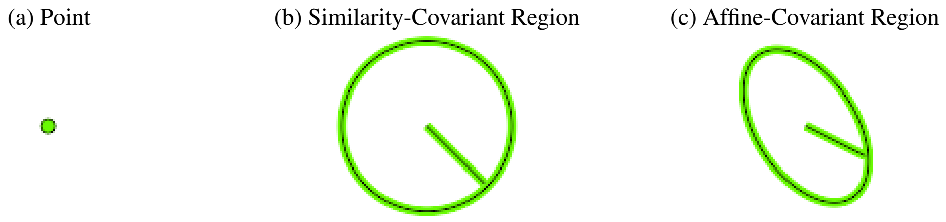
differential geometric constructions like bi-tangents (see Figure 3.6a) or curvature extrema of extracted contours, or moments of image features (see Figure 3.6a) [63, 65, 69, 70].

For a similarity-covariant region such as a Difference-of-Gaussian feature with its orientation set by the dominant gradient of the region, the local affine frame parameterization has the constraint that $\mathbf{x}_1 - \mathbf{x}_2 \perp \mathbf{x}_3 - \mathbf{x}_2$ and $d(\mathbf{x}_1, \mathbf{x}_2) = d(\mathbf{x}_2, \mathbf{x}_3)$ [60, 91]. Equivalently, similarity-covariant regions are minimally parameterized by two points.

The local affine frame defines a change of basis given by the orientation-preserving homogeneous transformation $\mathtt{A}$ that maps from the orthonormal affine basis at the orgin to the image space as

$$\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{bmatrix} = \mathtt{A}\begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 1 & 1 & 1\end{bmatrix}, \tag{3.8}$$

where $\mathbf{x}_2$ is the origin of the linear basis defined by vectors $\mathbf{x}_1 - \mathbf{x}_2$ and $\mathbf{x}_3 - \mathbf{x}_2$ in the image coordinate system [65, 91].



(a) Point

(b) Similarity-Covariant Region

(c) Affine-Covariant Region

Figure 3.7: *Feature Types.* There are three feature types used to represent imaged coplanar repeats: (a) Points are extracted from local affine frames that are constructed from region detections and have two degrees of freedom. (b) Similarity-covariant regions are returned by the Difference-of-Gaussians (DoG) detector and have four degrees of freedom. [60] (c) Affine-covariant regions are given by the MSER detector with LAF upgrade or by the Hessian Affine detector with Baumberg iteration and have six degrees of freedom [7, 62, 63, 69]. The ambiguity of rotation of the transformation taking the ellipse to the unit circle is fixed by the point on the ellipse (or circle). The figure is taken from [91].

Geometric Interpretations of Covariant Regions

Alternately, the geometry of an *affine-covariant region* $\mathcal{R}$ can be given by an ellipse with a point on its contour, where the ellipse fixes the translation, anisotropic scaling and skew, and the point fixes the rotation of the region with respect to the orthonormal affine basis at the origin. The ellipse is the simplest geometric primitive that can be used as an affine-covariant representation since the set of ellipses is closed with respect to an affine transformation, where, *e.g.*, the set of circles is not. An example of an affine-covariant region is illustrated in Figure 3.7c.

The geometry of a *similarity-covariant region* is given by a circle with a point on its circumference, where the circle fixes the translation and isotropic scaling, and the point fixes the rotation of the region with respect to the orthonormal affine basis at the origin. The circle is the simplest geometric primitive that can be used as a similarity-covariant representation since the set of circles is closed with respect to similarity transformation.Figure 3.7b illustrates the geometry of a similarity covariant region.

The matrix of the quadratic form $\mathtt{C}$ defining the ellipse of the affine-covariant region $\mathcal{R}$ can be expressed in terms of the change-of-basis matrix $\mathtt{A}$ defined in (3.8) as

$$\mathtt{C} = \mathtt{A}^{-\top} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathtt{A}^{-1}, \tag{3.9}$$

where the ellipse is given by the locus of $\mathbf{x}^T \mathtt{C} \mathbf{x} = 0$. Thus (3.9) relates the local affine frame parameterization of affine-covariant region $\mathcal{R}$ to the ellipse that gives the extent and shape of $\mathcal{R}$, which defines the patch of the image that can be approximately registered with a coplanar repeat (see Section 3.1.5) and Figure 3.1.

In the case where second central moments of image features are used (equivalently covariance) to compute the shape and extent of covariant region $\mathcal{R}$, *e.g.*, as shown in Figure 3.6b or with Hessian Affine features with Baumberg iteration [65, 7], $\mathtt{C}$ is the homogeneous form of an ellipse, where the upper 2×2 sub matrix is the covariance matrix. In the case where differential geometric constructions are used, as in Figure 3.6a, $\mathtt{C}$ is meaningless with respect to the method used to construct the covariant region. In both cases, two of the three points of the LAF lie on the ellipse defined by $\mathtt{C}$.

Note that the matrix quadratic form $\mathtt{C}$ is insufficient to define all degrees of freedom of the affine-covariant region since

$$\mathtt{C} = \mathtt{A}^{-\top} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathtt{A}^{-1} = \mathtt{A}^{-\top} \mathtt{R}_z^{\top} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathtt{R}_z \mathtt{A}^{-1} \tag{3.10}$$

for any rotation matrix $\mathtt{R}_z$ about the z-axis such that

$$\mathtt{R}_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.11}$$

Thus, there is a set of transformations mapping the ellipse $\mathtt{C}$ to the unit circle, and the rotation of the ellipse with respect to the orthonormal frame is undefined if only $\mathtt{C}$ is provided. Typically, either the affine basis $\{\, \mathbf{x}_1,\, \mathbf{x}_2,\, \mathbf{x}_3 \,\}$ or the change of basis $\mathtt{A}$ is used to define the local affine frame [91].

In Chapters 5 and 6 the affine basis parameterization is used in several derivations of solvers. The matrix $\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$ constructed from the affine basis vectors of a local affine frame is used as a parameterization of affine-covariant region $\mathcal{R}$, which we call its *point-parameterization*. It's properties will be exploited in Chapter 6.

3.2.4 Maximally Stable Extremal Regions (MSERs)

A maximally stable extremal region (MSER) [62] is bordered by a high-contrast boundary in the intensity image. MSER level sets can have an arbitrary shape and can include interior boundaries; however, an affine-covariant construction is typically estimated from the region, which is a straightforward low-parameter representation of localized geometry. Example affine-covariant constructions include representing the region with an ellipse calculated from the first and second moments of the region (see Figure 3.6b for an example), which was used in the seminal papers of [62, 65], or extracting affine-covariant points that correspond to differential geometric properties along the region boundary [63, 69] (see Figure 3.6a for an example).

The region detector thresholds the intensity image at all values of its sampled discretized range, *e.g.*, $t \in \{\, 0 \ldots 255 \,\}$, where t is the intensity threshold. Thresholding creates a nested sequence of contiguous regions $\{\, Q_1 \ldots Q_i \,\}$, where $Q_i \subseteq Q_{i+1}$. The extremal region Q_{i^*} is maximally stable only if

$$i^* = \operatorname*{argmin}_i \frac{|Q_{i+\delta} \setminus Q_{i-\delta}|}{|Q_i|}, \tag{3.12}$$

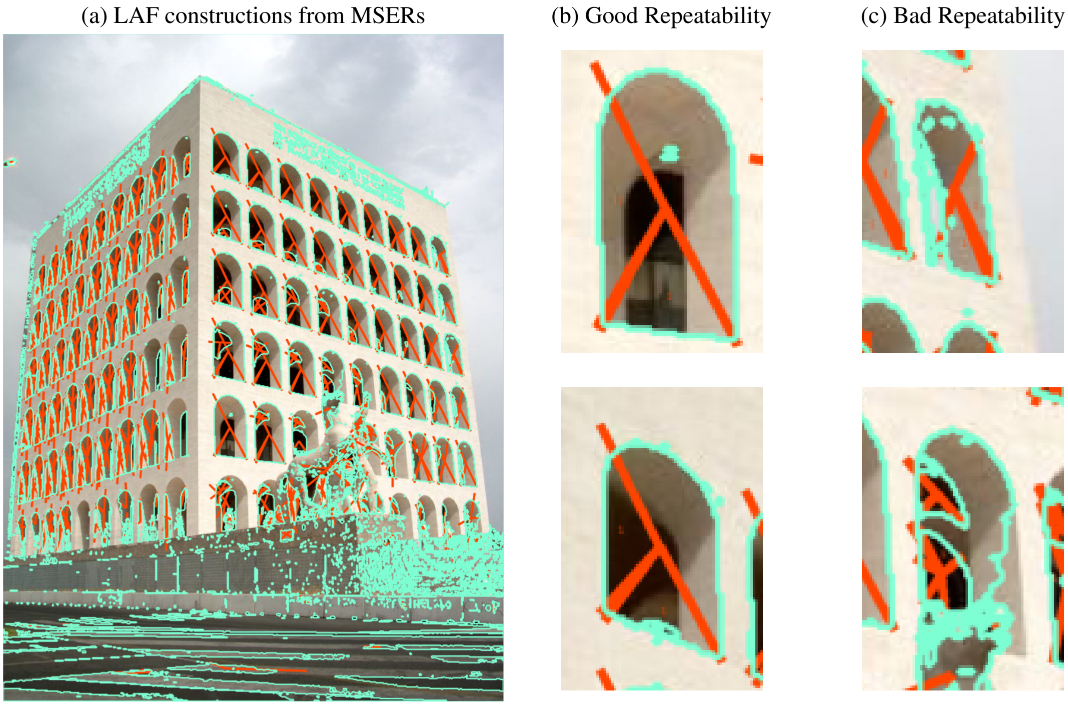


Figure 3.8: *MSERs, LAFs and Repeatability.* (a) LAFs (orange) constructed from MSER detections (cyan contours). (b) Two patches extracted from the front building facade where the MSER detector gives good repeatability. (c) Two patches from the front building facade where the MSER detector gives bad repeatability.

where δ is a user-supplied parameter, and $|\cdot|$ is set cardinality, which corresponds to the area of the extremal region. (3.12) selects regions that have stationary area with respect to varying threshold t.

The first and second central moment of the extremal region are computed as

$$\boldsymbol{\mu}_i = \frac{1}{|Q_i|}\sum_{\mathbf{x}\in Q_i} x, \quad \Sigma_i = \frac{1}{|Q_i|}\sum_{\mathbf{x}\in Q_i}(\mathbf{x}-\boldsymbol{\mu}_i)^\top(\mathbf{x}-\boldsymbol{\mu}_i), \tag{3.13}$$

where $\mathbf{x} \in Q_i$ are the coordinates of the pixels in the extremal region. This sets the shape and extent of the affine-covariant region construction $\mathcal{R}$ for Q_i. The orientation is set either by differential geometric property extracted from the boundary (see Figure 3.6b) or by the orientation of the dominant gradient in the patch local to Q_i.

3.2.5 Local Affine Frame upgrade from an MSER

Matas et al. and Obdrzalek et al. [63, 69] show that several affine-covariant constructions can be extracted from the differential geometry of the boundary of an extremal region. Some examples of affine-covariant constructions from points extracted from the boundary include (i) extremal points with respect to the center of gravity after region normalization, (ii) stable bi-tangents and points maximally distant from the bi-tangents, (iii) and points of extremal curvature. Importantly, these additional affine-covariant region constructions can be corresponded and used as constraints on geometry.

3.2.6 Hessian-Affine Regions

The Hessian-Affine detector applies the Hessian operator to the scale-space representation of the image, which is a sequence of images generated by convolving the intensity image with Gaussian kernels of increasing variance [57]. The blurring enables the selection of structures at their characteristic scales in the scale-space representation. The characteristic scale of an image structure (*e.g.* a blob) is the scale at which the convolved kernel achieves its maximum response. Lindberg [57] showed that the selected scale covaries with the relative scale of the same structure in the image. In practice, the operations of blurring and differentiation are combined using Gaussian derivatives with the operator

$$H(x,y,\sigma) = \begin{bmatrix} \frac{\partial^2 G(x,y,\sigma)}{\partial x^2} & \frac{\partial^2 G(x,y,\sigma)}{\partial x \partial y} \\ \frac{\partial^2 G(x,y,\sigma)}{\partial x \partial y} & \frac{\partial^2 G(x,y,\sigma)}{\partial y^2} \end{bmatrix}, \quad \text{where} \quad G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{3.14}$$

At each image in the scale space representation, the hessian affine detector returns points that are simultaneously extrema of both $\sigma^2 \det(H(x,y,\sigma))$ and $\sigma \operatorname{Tr}(H(x,y,\sigma))$ for x, y in the image-coordinate system. Jointly requiring extrema of both the determinant and trace prevents the detection of elongated blobs in the image. An iterative search is used to refine the spatial localization of the detections.

Baumberg Iteration

The scale-covariant detections can be upgraded to affine-covariant constructions through an iterative shape-adaptation process known as Baumberg iteration [7]. The scale-covariant Hessian affine detection (equivalently a circle) is iteratively adapted to an ellipse such that the second central moment matrix of image gradients in the neighborhood including and surrounding the ellipse is isotropic after the neighborhood is warped to the normalized coordinate system defined by a transformation taking the adapted ellipse to the unit circle (note that there is a rotational ambiguity, which gives a set of sufficient transformations). The second moment matrix $\mathtt{M}_I$ is given by the covariance of gradients in the neighborhood of $\Omega \in \mathbb{R}^2$ surrounding the center of a hessian affine detection

$$\mathtt{M}_I = \frac{1}{\Omega} \int_{\Omega} \nabla I \nabla I^{\top} dx dy, \tag{3.15}$$

where $\nabla I = \left(\frac{\partial G(x,y,\sigma)}{\partial x}, \frac{\partial G(x,y,\sigma)}{\partial y}\right)^{\top}$ and σ is chosen to filter image noise, which strongly affects gradient calculations. Let $\mathtt{M}_I$ be the shape-adapted ellipse with its origin at its midpoint, and suppose that $\mathtt{A}$ is the affine matrix mapping the locus of $\mathtt{M}_I$ to the unit circle $\mathtt{M}_J$. Then $\mathtt{M}_I$ transforms to the unit circle $\mathtt{M}_J$ with respect to the affine transformation as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathtt{A}^{-\top}\mathtt{M}_I\mathtt{A}^{-1}, \tag{3.16}$$

which implies that $\mathtt{M}_I = \mathtt{A}^{\top}\mathtt{A}$.

Any rotation $\mathtt{R} \in \mathbb{R}^{2\times 2}$ can be chosen since $\mathtt{M}_I = \mathtt{A}^{\top}\mathtt{R}^{\top}\mathtt{R}\mathtt{A}$. The rotation is fixed to map the dominant gradient orientation of the patch local to the shape-adapted hessian-affine detection to the x axis in the normalized frame.

3.2.7 Sparse Representation of Coplanar Repeated Patterns

The covariant regions given by the MSER and Hessian-Affine regions (see Sections 3.2.4 and 3.2.6) are detected on image texture with structure that is common to repetitive patterns, *i.e.*, blobs, corners, and salient differential geometric properties of contours, such as curvature extrema. Furthermore, these detectors have high repeatability on the same imaged texture, as defined in (3.7), with respect to significant changes of viewpoint and illumination [64, 66]. Their proven robustness in the multi-view matching task makes them good candidates for representing the local geometry of repeated textures. Figure 3.9 illustrates the geometric representation of the scene that is used as input to the undistorting and rectifying solvers proposed in Sections 5 and 6.

3.3 Corresponding Coplanar Repeats

The covariant region detections are corresponded by appearance. The patch local to each covariant region is transformed to a normalized frame that provides and invariant representation of the patch. The process of normalizing a covariant region to a patch that can be embedded to a descriptor is illustrated in Figures 3.6 and 3.11.

The transformed patches are embedded into high-dimensional feature vectors and are used to compute pairwise distances, where distance in the feature space is proportional to how close in appearance the texture local to two covariant regions are. The feature vectors are clustered, which partitions the covariant regions into sets of repeats and non-repeated content. The results is tentatively corresponded coplanar repeats based on appearance. The details of this process are provided in the next sections.

3.3.1 The Scale Invariant Feature Transform (SIFT)

The Scale Invariant Feature Transform (SIFT) characterizes the appearance of an image patch by embedding a spatial histogram of gradients of the patch texture into a feature vector [60]. A *spatial histogram* divides the image patch into a grid, where a histogram is computed on the image

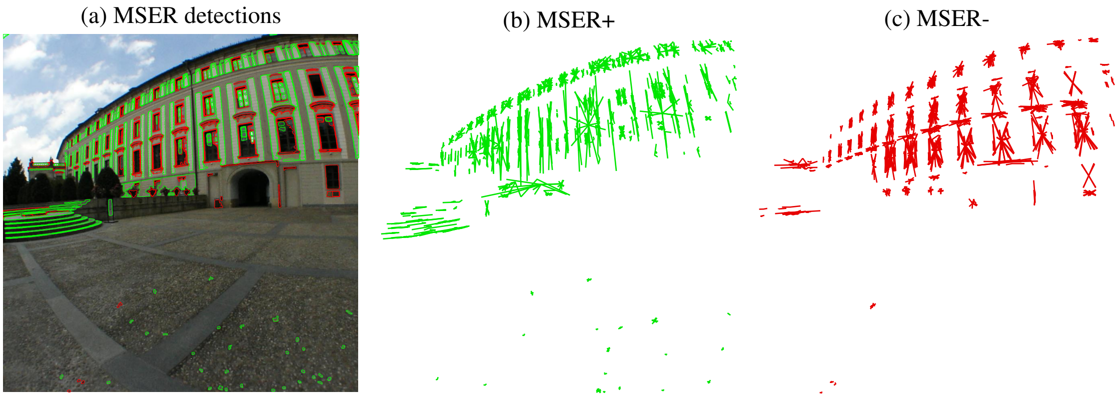


Figure 3.9: *Sparse Representation.* (a) MSERs are detected. Local Affine Frames (LAFs) are constructed from contours on the boundaries of high-contrast regions with (b) higher intensities than the surrounding pixels, called MSER+ and (c) lower intensities than the surrounding pixels, called MSER-. Note that salient geometric structure is retained by the set of LAF constructions and the pattern is easily discernible in its sparse representation.

features contained in each grid cell (see Figure 3.10). The image features for SIFT are gradient orientation and position of the gradient orientation at for each pixel in the cell. The histograms are concatenated to construct the spatial histogram. The gradient orientations are weighed by the gradient magnitude and are accumulated in each smaller patch's histogram, which up to normalization and clamping is the embedding for the smaller patch. The gradients of the patch are weighted by a Gaussian with its mean at the patch center, which gives more importance to gradients at the center of the covariant region. The spatial histogram is constructed in the normalized frame defined by the covariant region detection, which is shown in Figure 3.11. The construction of the spatial histogram is shown in Figure 3.10. The feature vector is constructed from the spatial histogram by stacking each orientation bin from each cell along the columns of the grid. The SIFT embedding of a covariant region $\mathcal{R}$ is denoted $\mathbf{r}^{\mathrm{S}}(\mathcal{R})$. The feature vector is L2-normalized, $\|\mathbf{r}^{\mathrm{S}}(\mathcal{R})\| = 1$.

RootSIFT

Arendelovic et al. [4] noted that empirical results from texture classification methods showed that superior performance resulted from using measures on histograms, such as χ^2 or Hellinger distance, rather than the Euclidean distance between histograms. In particular, the Hellinger distance is desirable since a simple transformation of the SIFT embeddings as defined in Section 3.3.1 allows the Hellinger distance to be computed from the Euclidean distance operator. This is a pragmatic consideration that allows the Hellinger distance to be used in black-box feature matching frameworks, where Euclidean distance is hard coded.

In particular, if $\mathbf{r}^{\mathrm{S}}(\mathcal{R})$ denotes a SIFT embedding of a covariant region $\mathcal{R}$ and $\|\mathbf{r}^{\mathrm{S}}(\mathcal{R})\| =$

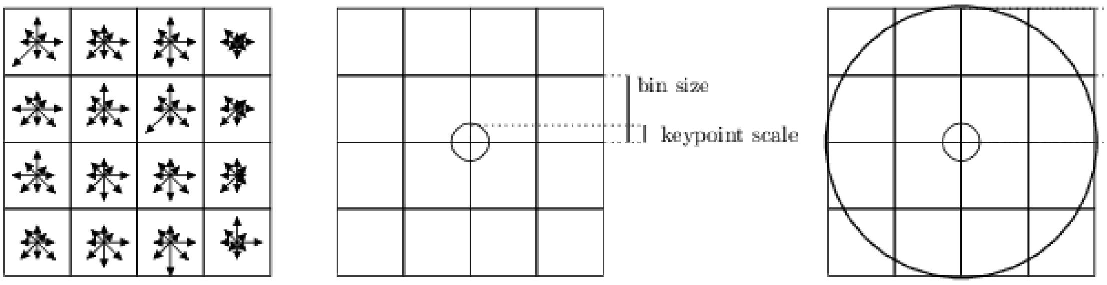


Figure 3.10: *SIFT Spatial Histogram.* The spatial histogram used by SIFT. (left) The patch is divided into a grid, where the gradient orientations in each cell are accumulated and discretized into an eight bin circular histogram. (center) The patch in the normalized frame is centered around the covariant region detection and its extents are scaled. (right) The gradients are weighted with an isotropic Gaussian kernel with its mean at center of the covariant region detection, equivalently, the patch. The figure is taken from [91].

1, then the Euclidean distance $d(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j))$ is related to their similarity kernel, namely $S(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j))$, as

$$\begin{aligned} d(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)) &= \|\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i) - \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)\|_2^2 \\ &= \|\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i)\|_2^2 + \|\mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)\|_2^2 + 2\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i)^\top \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j) \\ &= S(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_i)) + S(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_j), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j) + 2S(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)) \\ &= 2 - 2S(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)), \end{aligned} \tag{3.17}$$

where $S(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)) = \mathbf{r}^{\mathrm{S}}(\mathcal{R}_i)^\top \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)$.

The Hellinger kernel of two N-dimensional embeddings $\mathbf{e}_1 = \big(x_1, \ldots, x_N\big)^\top$ and $\mathbf{e}_2 = \big(y_1, \ldots, y_N\big)^\top$ is given as

$$H(\mathbf{e}_1, \mathbf{e}_2) = \sum_{k=1}^{N} \sqrt{x_k y_k}, \tag{3.18}$$

where $x_k, y_k > 0$ and $\sum_k x_k = 1, \sum_k y_k = 1$.

Then the Hellinger distance (3.18) between two SIFT embeddings can be computed with the Euclidean similarity kernel $S(\cdot,\cdot)$ by enforcing $\sum \mathbf{r}_k(\mathcal{R}) = 1$ and taking the element-wise square root,

$$S\left(\sqrt{\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i)}, \sqrt{\mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)}\right) = \sqrt{\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i)}^\top \sqrt{\mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)} = H(\mathbf{r}^{\mathrm{S}}(\mathcal{R}_i), \mathbf{r}^{\mathrm{S}}(\mathcal{R}_j)) \tag{3.19}$$

The RootSIFT method explicitly performs this transformation on the feature vector. Let $\mathbf{r}^{\mathrm{S}}(\mathcal{R}) = \big(x_1, \ldots, x_N\big)$ be a SIFT vector. Then the RootSIFT embedding $\mathbf{r}(\mathcal{R}) = \big(y_1, \ldots, y_N\big)^\top$

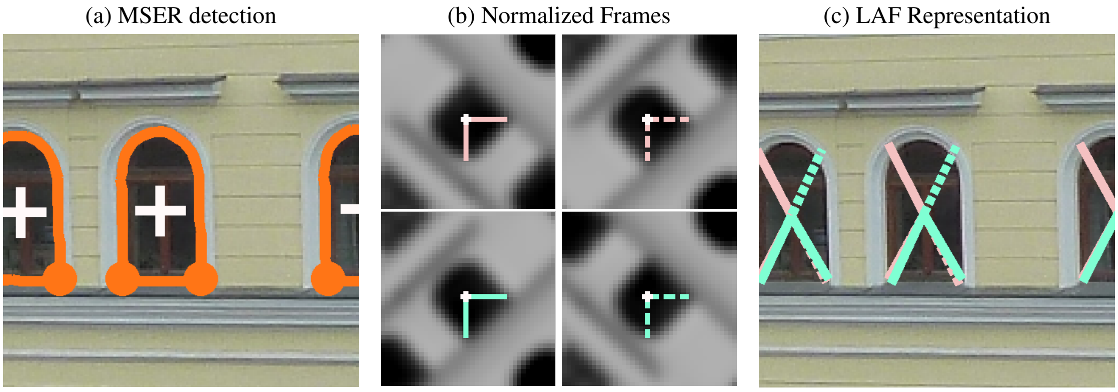


Figure 3.11: *Region Detection and Description.* (a) Center of gravity (white cross) and curvature extrema (orange circles) of a detected MSER (orange contour [62]). Patches are normalized to a square and oriented to define an affine frame as in [63], (b) Bases are reflected for detecting axial symmetries. The RootSIFT transform embeds the local texture [4, 60]. (c) Affine frames are mapped back into image.

can be computed from $\mathbf{r}^{\mathrm{S}}(\mathcal{R}) = \left(x_1, \ldots, x_N\right)^\top$ by

$$y_k = \sqrt{x_k / \sum_k x_k}. \tag{3.20}$$

The RootSIFT transformation has the effect of reducing the ratio between the largest and smallest bin values in the histogram. If these ratios are large, then the Euclidean distance between L2-normalized SIFT embeddings will be dominated by these extremal bins. Arendelovic et al. [4] showed that the RootSIFT transformation gave dramatic improvements to the mean average precision of bag-of-words image retrieval systems. In the methods presented in this thesis, the appearance of patches local to covariant regions are embedded using the RootSIFT transformation.

3.3.2 Establishing Tentative Coplanar Repeats

Affine frames are tentatively labeled as repeated texture by their appearance. Figure 3.12 shows an example tentative labeling of coplanar repeats. The appearance of an affine frame is given by the RootSIFT embedding 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 such that their orientation is flipped (handedness).

The RootSIFT descriptors are agglomeratively clustered, which establishes pair-wise tentative correspondences amongst connected components. Denote the RootSIFT embedding of an affine-covariant region as $\mathbf{r}(\mathcal{R})$. The tentative clustering is given by single-link hierarchical

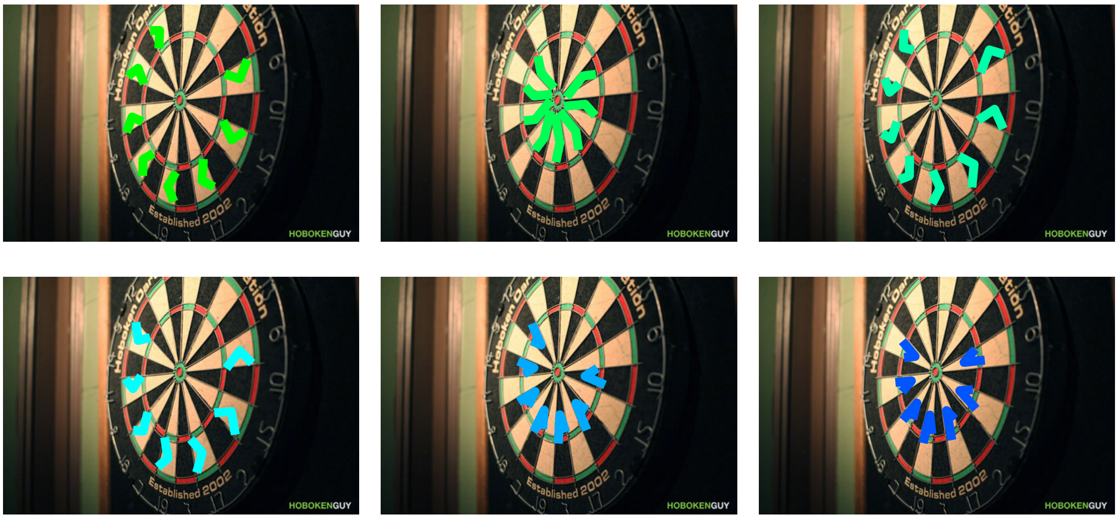


Figure 3.12: *Tentative Coplanar Repeats.* Descriptors of patches local to covariant region detectors are clustered. The local affine frames constructed from the detected covariant regions are color coded by their cluster label. Covariant regions with the same color (same cluster label) are tentatively corresponded as coplanar repeats.

agglomerative clustering, which merges two sets of RootSIFT embeddings, denoted $C_j, C_{j'}$, if

$$\min_{i \in C_j, i' \in C_{j'}} \|\mathbf{r}(\mathcal{R}_i) - \mathbf{r}(\mathcal{R}_j)\|_2^2 < t_{app}^2,$$

where t_{app} is conservatively set to favor over-segmentation and smaller but more precise clusters. Each appearance cluster has some proportion of its indices corresponding to affine frames that represent the same coplanar repeated scene content, which are the *inliers* of that appearance cluster. The remaining affine frames are the *outliers*.

Denote the tentative clustering as the collection of K appearance clusters $\mathcal{C} = \{ C_1, \dots, C_K \}$, where each appearance cluster is a subset of the indices of the affine frames, and an index occurs in exactly one appearance cluster, $C_j \cap C_{j'} = \varnothing$, where $j \neq j'$.

3.3.3 Spatial Verification of Tentative Coplanar Repeats

Let $\{ \tilde{\mathbf{x}}_{i,k} \}$ be the affine basis of a covariant region $\tilde{\mathcal{R}}_i$ detected in a radially-distorted image, where $k \in \{ 1 \dots m \}$ and m is either 2 or 3 for a similarity-covariant or affine-covariant detection. Then the point-wise symmetric transfer error between two imaged coplanar rigidly-transformed covariant regions $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ is

$$\begin{aligned}\epsilon(\mathtt{H}_\mathtt{T}, \lambda, \tilde{\mathcal{R}}_i, \tilde{\mathcal{R}}_j) = \sum_{k=1}^{m} d(\tilde{\mathbf{x}}_{i,k}, g(\tilde{\mathbf{x}}_{j,k}, \mathtt{H}_\mathtt{T}, \lambda))^2 + \dots \\ d(\tilde{\mathbf{x}}_{j,k}, g(\tilde{\mathbf{x}}_{i,k}, \mathtt{H}_\mathtt{T}, \lambda))^2,\end{aligned} \tag{3.21}$$

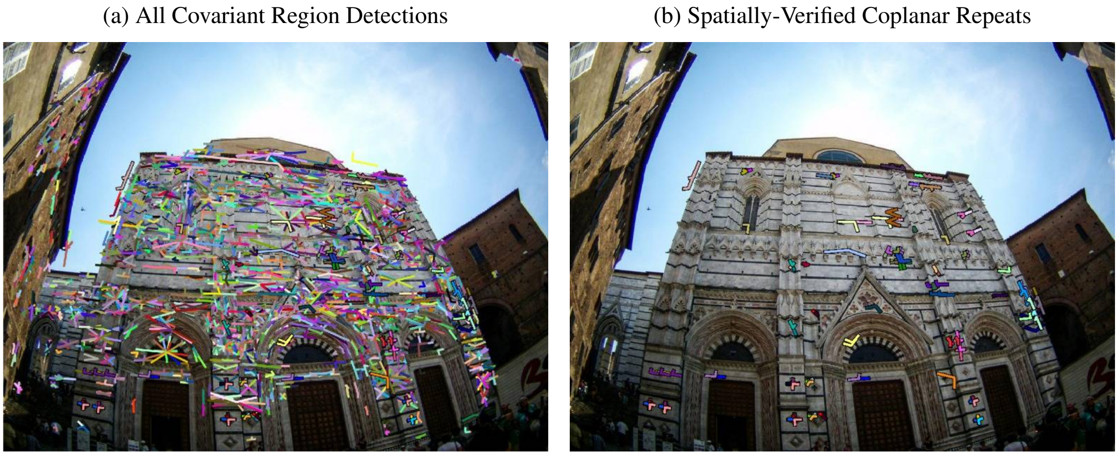

(a) All Covariant Region Detections
(b) Spatially-Verified Coplanar Repeats

Figure 3.13: *Spatially-Verified Coplanar Repeats.* (a) LAFs constructed from covariant region detections. (b) LAFs spatially verified as coplanar repeats. The proportion of regions that are tentatively corresponded and spatially verified, *i.e.*, inliers, is small. Even after spatial verification there are some errors: *E.g.*, two coplanar repeats remain that are artefacts of image compression along building edge.

where $d(\cdot,\cdot)$ is the Euclidean distance. The thresholded symmetric transfer error is used as a sufficient condition for labeling clustered covariant regions as coplanar repeats. In other words, if there is $\mathtt{H}_{\mathtt{T}}, \lambda$ such that $\epsilon(\mathtt{H}_{\mathtt{T}}, \lambda, \tilde{\mathcal{R}}_i, \tilde{\mathcal{R}}_j) < t$, where $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ are in a covariant region cluster (as defined in Section 3.3.2), then $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ are coplanar repeats. Figure 3.13 shows what remains the sparsity of spatial verified coplanar repeats Figure 3.13b) with respect to all detected regions (see Figure 3.13a). The ratio of good-to-bad regions is quite small, which motivates the design of the minimal solvers in Chapters 5 and 6 for this problem.

4 Solving Systems of Polynomial Equations

The polynomial systems of equations encoding the rectifying constraints for a subset of the proposed solvers in Chapter 5 and all of the proposed solvers in Chapter 6 are solved using an algebraic method based on Gröbner bases. Automated solver generators using the Gröbner basis method [44, 48] have been used to generate solvers for several camera geometry estimation problems [44, 45, 48, 49], see also Chapters 5 and 6.

4.1 The Gröbner basis Method

Camera geometry estimation problems frequently lead to a formulation as a system of multivariate polynomial equations. Estimating geometry as a *minimal problem* means that the minimal number of measurements (usually point or feature correspondences) together with all geometric constraints are used to eliminate extra degrees of freedom from the constraint equations. However, minimal problems often result in complicated systems of polynomial equations. Since minimal solvers require the fewest number of measurements, they are key parts of robust estimation scheme like RANSAC [25]. Since robust schemes draw many measurement samples, the solvers need to be fast.

State-of-the-art polynomial solver generators can create specific polynomial solvers [85, 42] to efficiently solve a given minimal problem. Solver generators are based on methods from algebraic geometry such as the Gröbner bases and action/multiplication matrices [19, 42], or the method of resultants [19, 46]. The generated specific solvers, unlike general algebraic methods, cannot solve general systems of polynomial equations. They can efficiently solve only systems of polynomial equations of a given form, i.e. systems consisting of the same unknowns and monomials and differing only in non-degenerate coefficients. However, these specific solvers are usually more efficient than the general methods. For example, modern specific solvers for camera geometry problems usually execute in mere microseconds.

In the next paragrahs an introduction to basic concepts of algebraic methods for solving systems of polynomial equations is provided. In this thesis we use the notation and basic concepts from the algebraic geometry book of Cox et al. [19]. Consider a system of m polynomial equations,

$$F = \{f_1(x_1, ..., x_n) = 0, ..., f_m(x_1, ..., x_n) = 0\} \tag{4.1}$$

in n unknowns $X = \{x_1, ..., x_n\}$. The goal is to solve this system. Let $\mathbb{C}[X]$ denote the set of all polynomials in unknowns X with coefficients in $\mathbb{C}$. The ideal

$$I = \langle f_1, \ldots, f_m \rangle \subset \mathbb{C}[X] \tag{4.2}$$

is the set of all polynomial combinations of our generators $f_1, \ldots, f_m$ (4.1). An affine variety is the set of all solutions to the system (4.1), *i.e.* the set

$$V(F) = \{\mathbf{x} \in \mathbb{C} | f_i(\mathbf{x}) = 0, i = 1, \ldots, m\} \tag{4.3}$$

Each polynomial $f \in I$ vanishes on the solutions to the input system of equations (4.1). Here we assume that the ideal I generates a zero dimensional ideal, in other words, that the system (4.1) has a finite number of solutions.

The ideal I can be used to define the quotient ring $A = \mathbb{C}[X]/I$, which is the set of equivalence classes over $\mathbb{C}[X]$ defined by the relation $a \sim b \iff a = b \mod I \iff a - b \in I$. We will denote these equivalence classes using brackets, i.e. $a \sim b \iff [a] = [b]$. If I is a zero-dimensional ideal, equivalently, our system (4.1) has a finitely many solutions, then the quotient ring $A = \mathbb{C}[X]/I$ is a finite-dimensional vector space over $\mathbb{C}$.

For an ideal I there exist special sets of generators called Gröbner bases, which have the property that the remainder after division is unique. Gröbner bases can be used to define a basis B for the quotient ring $A = \mathbb{C}[X]/I$ and they can be used to solve systems of polynomial equations (4.1).

The action matrix method [19, 5] (also called the multiplication matrix method) is a frequently used approach for solving systems of equations with Gröbner bases. *E.g.*, the method has been used to generate efficient solvers for many minimal computer-vision problems [42, 44, 51, 55]. The strategy of the action-matrix method is to transform the problem of finding solutions to (4.1) to a problem of eigendecomposition of a special multiplication matrix [17].

Let us consider the mapping $T_f : A \to A$ of the multiplication by a polynomial $f \in \mathbb{C}[X]$ in $A = \mathbb{C}[X]/I$ as

$$T_f([g]) = [f].[g] = [fg] \in A. \tag{4.4}$$

T_f is a linear mapping for which $T_f = T_g \iff f - g \in I$. In our case A is a finite-dimensional vector space over $\mathbb{C}$ and therefore we can represent $T_f : A \to A$ by its matrix with respect to some linear basis B of A.

Without loss of generality, let the basis B be a monomial basis consisting of K monomials $B = ([b_1], \ldots, [b_K])$ then T_f can be represented by $K \times K$ multiplication (action) matrix $\mathtt{M}_f := (m_{ij})$ such that

$$T_f([b_j]) = [fb_j] = \sum_{i=1}^{K} m_{ij} [b_i]. \tag{4.5}$$

It can be easily shown [17] that $\lambda \in \mathbb{C}$ is an eigenvalue of the matrix $\mathtt{M}_f$ iff λ is a value of the function f on the variety V (4.3). In other words, if f is e.g. x_n then the eigenvalues of $\mathtt{M}_f$ are the x_n-coordinates of the solutions of (4.1) and the solutions to the remaining variables can be obtained from the eigenvectors of $\mathtt{M}_f$. This means that the multiplication matrix $\mathtt{M}_f$ can be used to recover the solutions by solving the eigendecomposition of $\mathtt{M}_f$ for which efficient algorithms exist. Moreover, if the ideal I is a radical ideal, *i.e.* $I = \sqrt{I}$, [17], which is usually the case of camera geometry problems, then K is equal to the number of solutions to the system (4.1). This means that we are solving an eigenvalue problem of size that is equivalent to the number of solutions of the considered problem. For more details and proofs we refer the reader to Cox et al. [19]

The coefficients of the multiplication matrix $\mathtt{M}_f$ are polynomial combinations of coefficients of the input polynomials (4.1). For computer vision problems these polynomial combinations are often found “offline“ in a pre-processing step. In this pre-processing step an expanded set of equations constructed by multiplying original equations with different monomials [42] is generated, which is called an *elimination template*.

After filling the template matrix with coefficients from the input equations and performing Gauss-Jordan (G-J) elimination or QR decomposition of this template matrix, the coefficients of the the multiplication matrix $\mathtt{M}_f$ can be obtained from this eliminated template matrix.

The first automated method for generating elimination templates and Gröbner basis solvers was presented in [44]. Larsson et al. [51] proposed and inprovement to the automatic generator of Kukelova et al. [44]. The proposed method uses the inherent relations between the input polynomial equations and it generates more efficient solvers than [44]. Further extensions to the method of Larsson et al. [51] include handling saturated ideals [52] and symmetry detection in polynomial systems [50].

Several approaches for optimizing Gröbner basis solvers with respect to numerical stability and efficiency have been proposed recently. In [40, 44] and [68] the authors presented methods for optimizing the size of elimination templates. Methods for improving numerical stability based on QR and SVD decomposition of template matrices were proposed in [11]. In [43] authors transformed elimination template matrices into a block diagonal form and in this way they sped up several solvers. A method for extracting univariate characteristic polynomial of the action matrix was proposed in [53]. The roots of the characteristic polynomial were found efficiently using Sturm-sequences [34], instead of computing the full eigendecomposition of the action matrix.

In general it is difficult to find the smallest elimination template for a given problem. In [55] the authors proposed two methods for generating small elimination templates. The first enumerates and tests all Gröbner bases in an efficient way and generates solvers w.r.t. all different Gröbner bases and standard monomial bases B of $A = \mathbb{C}[X]/I$. While there are (uncountably) infinitely many different monomial orderings for a given ideal I, there are only finitely many different reduced Gröbner bases [67, 27]. The set of all reduced Gröbner bases of an ideal is computed [27, 36] using the Gröbner fan of the ideal [67, 87]. The second method presented in [55] uses a heuristic sampling scheme for generating “non-standard" monomial bases B of $A = \mathbb{C}[X]/I$, and it leads to more efficient solvers than the Gröbner fan method for many problems. In [55] the heuristic sampling scheme was used to generate 1000 feasible candidate “non-standard" monomial bases B. From these 1000 bases a basis that was minimizing the size of the elimination template matrix was used to generate the final solver. In minimal solvers for rectifying from radially-distorted scales presented in Chapter 6 we use the heuristic sampling scheme to optimize the numerical stability of the solvers.

The Gröbner basis method is applicable for generating solvers for systems of polynomial equations with few unknowns. Therefore, it is important to eliminate some unknowns and simplify the input equations before applying the Gröbner basis method. There are many different ways to eliminate unknowns from the input equations. The applicability of the method depends on the structure of input equations. Kukelova et al. [47] propose a semi-automatic method for eliminating unknowns from input equations, which results in simplified solvers being generated. This method is based on elimination ideal theory [19] and was applied to several minimal problems from computer vision. The main limitation of this method is that it can be applied only to systems which contain equations that are independent on input measurements.

Another elimination approach that has been applied to several minimal problems [42, 33] in computer vision is based on the *hidden variable trick*. The hidden variable trick uses hidden variable resultants and is mostly used to eliminate all variables except one and transform the system of polynomial equations to a univariate polynomial. In the problems presented in this thesis we apply the hidden variable trick to eliminate a subset of unknowns. Next we describe the main idea of the hidden variable trick.

4.2 The Hidden Variable Trick

The proposed solvers in Chapter 5 uses the hidden variable trick to transform its polynomial constraint equations into a tractable form. The hidden variable trick is a resultant-based technique in algebraic geometry that is used to eliminate subsets of variables from multivariate polynomial systems of equations [19].

Suppose that a multivariate polynomial system of m equations in n unknowns (4.1) is given. The hidden variable trick works by assuming that a set $Y = \{x_j\}_{j \in I}$, $I \subset \{1, \dots n\}$ of $k < n$ unknowns are parameters belonging in the coefficient field, i.e. assuming that input n polynomials are polynomials in $n - k$ variables. Without loss of generality let us assume that $Y = \{x_1, \dots, x_k\}$. Then the input polynomials (4.1) are considered as polynomials in variables $X \setminus Y = \{x_{k+1}, \dots, x_n\}$, *i.e.*

$$f_1, \dots, f_m \in (\mathbb{C}[Y])[x_{k+1}, \dots, x_n]. \tag{4.6}$$

We sometimes say that we "hide" the variables $Y = \{x_1, \dots, x_k\}$ in the coefficient field, which gives also the name of the method. With this assumption the system can be rewritten in the matrix form as

$$\mathtt{M}(x_1, \dots, x_k)\mathbf{y} = \mathbf{0}, \tag{4.7}$$

where $\mathtt{M}(x_1, \dots, x_k)$ is $m \times l$ matrix containing polynomials in variables $Y = \{x_1, \dots, x_k\}$ and $\mathbf{y}$ is a $l \times 1$ vector of l monomials in the remaining $n - k$ variables (*i.e.*, monomials in $X \setminus Y = \{x_{k+1}, \dots, x_n\}$ including 1).

If a nontrivial solution to the system (4.1) exists then the matrix $\mathtt{M}(x_1, \dots, x_k)$ in (4.7) is rank-deficient. Therefore all the $l \times l$ minors of the matrix $\mathtt{M}(x_1, \dots, x_k)$ vanish. This generates a system of $\binom{m}{l}$ polynomial equations in k unknowns $\{x_1, \dots, x_k\}$. In this way the problem is simplified by eliminating $n - k$ unknowns $\{x_{k+1}, \dots, x_n\}$.

Unfortunately in this way we may introduce false solutions, i.e. the new system of $\binom{m}{l}$ equations in k unknowns $\{x_1, \dots, x_k\}$ may have solutions that are not solutions to the original system (4.1). These false solutions correspond to solutions where a $l - 1 \times l - 1$ submatrix of $\mathtt{M}(x_1, \dots, x_k)$ with columns corresponding to monomials other than 1 is rank-deficient. Then the full matrix $\mathtt{M}(x_1, \dots, x_k)$ has a right nullspace vector with zero coordinate where $\mathbf{y}$ in (4.7) has 1. In this way, we may even introduce a one-dimensional family of false solutions. Since the Gröbner basis method and the automatic generator [51, 44] assumes zero-dimensional ideals, *i.e.* systems with a finite number of solutions, such false solutions have to be removed. This can be done using the saturation trick presented in [52].

5 Minimal Solvers for Rectifying from Radially-Distorted Conjugate Translations

This chapter introduces minimal solvers that jointly solve for affine-rectification and radial lens undistortion from the images of translated and reflected coplanar textures (*e.g.*, see Figures 5.1, 5.2, and 5.3). In addition, the solvers estimate the vanishing point of the translation direction of the inputted point or region correspondences. The proposed solvers use the invariant that the affine-rectified image of the meet of the joins of radially-distorted conjugately-translated point correspondences is on the line at infinity. The hidden-variable trick from algebraic geometry is used to reformulate and simplify the constraints so that the generated solvers are stable, small and fast. Multiple solvers are proposed to accommodate various local feature types and sampling strategies, and, remarkably, three of the proposed solvers can recover rectification and lens undistortion from only one radially-distorted conjugately-translated affine-covariant region correspondence. Synthetic and real-image experiments confirm that the proposed solvers demonstrate superior robustness to noise compared to the state of the art. Accurate rectifications on imagery taken with narrow to fisheye field-of-view lenses demonstrate the wide applicability of the proposed method. The method is fully automatic.

5.1 Introduction

Each of the proposed minimal solvers exploits the following properties of radially-distorted conjugate translations: (i) The affine-rectified image of the meet of the joins of conjugately–translated point correspondences is on the line at infinity (see Section 5.2), and (ii) a conjugate translation is a homography with only four degrees of freedom (see Section 5.3).

The proposed minimal solvers are differentiated by the choice to eliminate either the unknown vanishing point or vanishing line from the polynomial systems that arise from constraints induced by radially-distorted conjugately translated local features. The group of Eliminated Vanishing Point (EVP) solvers provide flexible sampling in a RANSAC-based estimator: they can jointly recover undistortion and rectification from radially-distorted conjugate translations in one or two directions, where some of the point correspondences can translate with arbitrary distance. In addition, there is an EVP variant that admits reflections. The one-direction variants require one affine-covariant region correspondences, while the two-direction variants require two similarity-covariant region correspondences.

The Eliminated Vanishing Line (EVL) solver jointly recovers undistortion and rectification from one radially-distorted conjugately-translated affine-covariant region correspondence. The geometry of this configuration enables the elimination of the vanishing line, which results in a

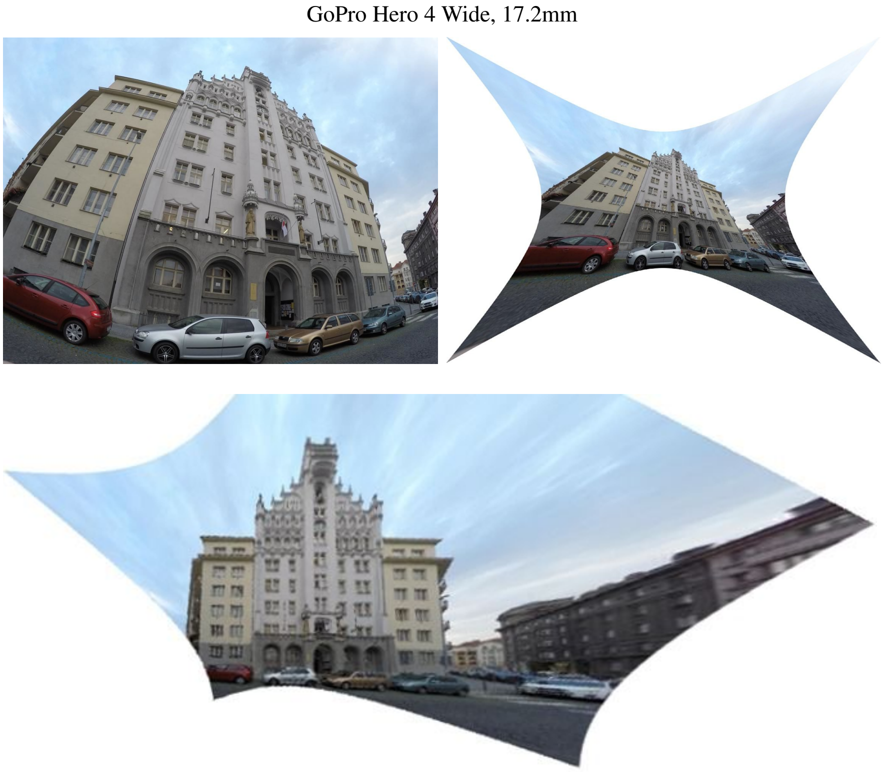


Figure 5.1: *Inputs and Outputs.* Input (top left) is a distorted view of a scene plane with translational symmetries and reflections, and the outputs (top right, bottom) are the radially undistorted image and the rectified scene plane. The method is fully automatic.

solver that is very stable, fast and robust to feature noise.

Covariant region detections reduce the number of required correspondences to as few as one for the proposed solvers, but corners or combinations of corners and covariant regions can also be used as input. Since the proposed solvers are derived from constraints induced by point correspondences, points are extracted from the region correspondences as input to the proposed solvers.

With one or two-correspondence region sampling, an accurate undistortion and rectification is quickly recovered, even for difficult scenes (see Figure 5.7). The proposed solvers are ideally suited for RANSAC, where the minimal sample size reduces the required trials, the fast time to solution ensures fast trials, and the noise robustness ensures an accurate rectification is recovered when inlying correspondences are sampled [24].

Examples of both frame constructions are shown in Figures 5.2, 5.5, and 5.6.

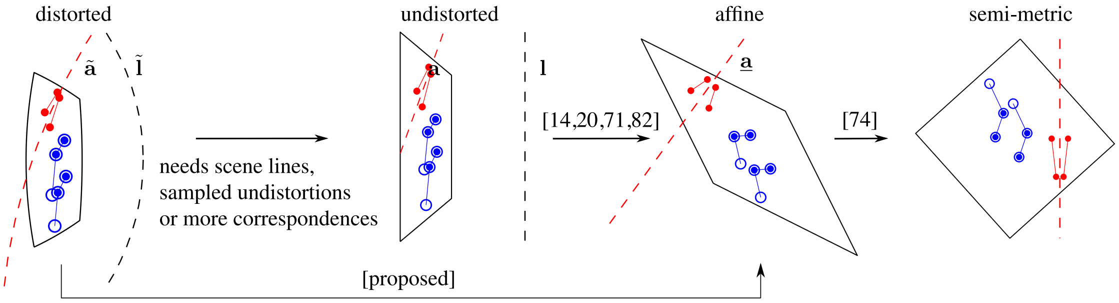


Figure 5.2: *Direct Affine Rectification.* The hierarchy of rectifications from distorted to metric space is ascended from the left. Color denotes the transformation: blue is conjugate translation and red is imaged reflection. Marker type denotes the correspondence configurations that the proposed solvers admit: circles for three conjugately-translated point correspondences and filled circles for two pairs of two point correspondences, where one pair is consistent with a radially-distorted conjugate translation and the other pair is consistent with either a distorted conjugate translation or distorted reflection (shown here as $\tilde{\mathbf{a}}$). The scene plane's vanishing line is shown in the original and undistorted image ($\tilde{\mathbf{l}}$ and $\mathbf{l}$, respectively), as well as the reflection axis of the red features ($\tilde{\mathbf{a}}$, $\mathbf{a}$, respectively, where $\underline{\mathbf{a}}$ is the rectified reflection axis). Point correspondences (circles) are extracted from scale or affine-covariant region correspondences (solid polylines), which can reduce the number of required correspondences to one. The state-of-the art requires sampled undistortions, scene lines [3, 94], or three affine-covariant region correspondences (see Chapter 6). Affine-rectified images are metrically upgraded with the method of [74] for presentation (see Section 5.6.4).

5.1.1 Previous Work

Chapter 6 introduces minimal solvers that can rectify from the image of rigidly-transformed coplanar repeats, but these solvers are over 2000 times slower than the fastest of the proposed solvers (see Table 5.3) in this chapter and require three affine-covariant region correspondences for the most commonly used configuration. In contrast, the proposed solvers include three variants requiring only one region correspondence, which, in addition to the very fast time to solution of the proposed solvers, results in a massive speedups of the RANSAC-based estimator used in this chapter (from [74]) compared to the solvers in Chapter 6 (see Table 5.1). Furthermore, the solvers in Chapter 6 admit only region correspondences since those solvers place constraints on the rectified scales of corresponded coplanar regions, whereas the proposed solvers also admit radially-distorted conjugately-translated point correspondences.

An exhaustive list of minimal solvers that are capable of jointly estimating lens undistortion and affine-rectification with local feature extracted from radially-distorted conjugately-translated textures is included in the survey of solvers listed in Table 5.2.

	Wildenauer et al. [94]	Antunes et al. [3]	Chapter 6	Proposed
Feature Type	fitted circles	fitted circles	covariant regions	points, covariant regions
Number	set of 2 and 3 lines	set of 3 and 4 lines	3 region correspondences	1 region correspondence
Assumption	parallelism	parallelism	rigidly transformed	translated, reflected
Rectification	multi-model	multi-model	direct	direct

Table 5.1: *Scene Assumptions.* Rectifying solvers from [94, 3] require distinct sets of parallel scene lines as input and multi-model estimation. Solvers in Chapter 6 admit region correspondences extracted from rigidly-transformed coplanar repeated scene texture, but require 3 correspondences for the most common solver variant and cannot admit points correspondences. The proposed solvers rectify from just 1 radially-distorted conjugately-translated region correspondence and also admit point correspondences (see Figures 5.2 and 5.5).

5.1.2 Solving Systems of Polynomial Equations

The polynomial systems of equations encoding the rectifying constraints for the Eliminated Vanishing Point Solvers (EVP) are solved using an algebraic method based on Gröbner bases. Automated solver generators using the Gröbner basis method [44, 48] have been used to generate solvers for several camera geometry estimation problems [44, 45, 48, 49], see also Chapter 6. However, the straightforward application of automated solver generators to the proposed constraints resulted in unstable solvers (see Section 5.7 and Figure 5.8a). Larsson et al. [49] introduced a method called ideal saturation for generating polynomial solvers for problems where unwanted solutions arise because of simplifications during modeling. The hidden variable trick with ideal saturation is used to eliminate unknowns from the polynomial system of equations arising in the formulations of the Eliminated Vanishing Point solvers (see Section 5.4.1), which results in significantly more numerically stable solvers (see Figure 5.8a) than solvers generated from the original constraint equations.

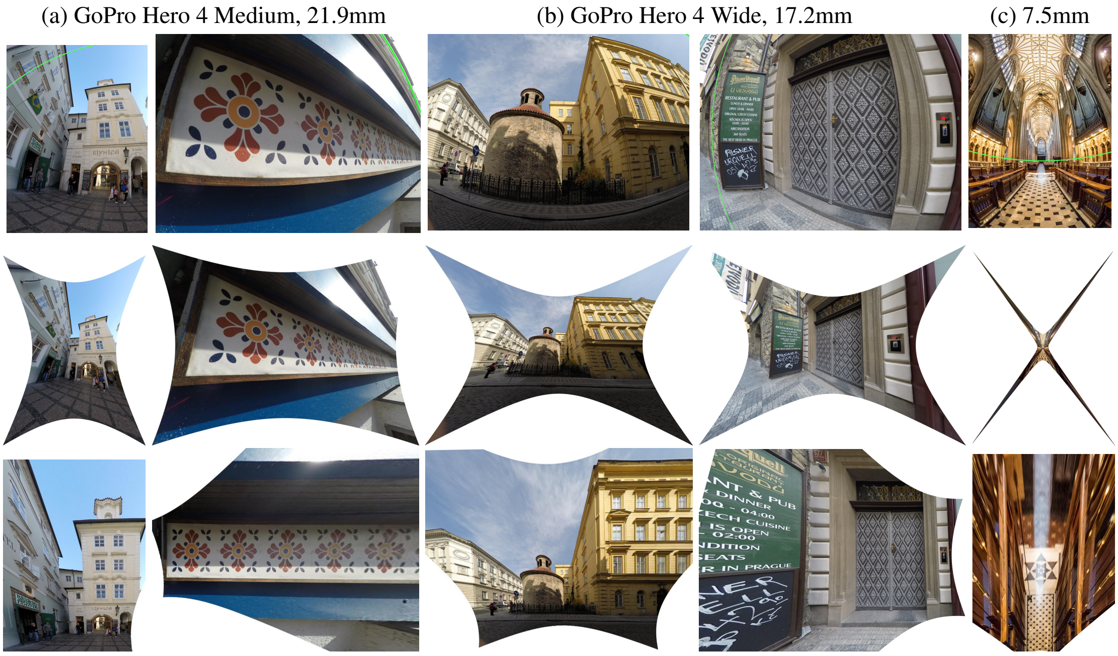
5.2 Meets of Joins

Let $\mathbf{m}_i$ be the join of the conjugately translated point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. Then $\mathbf{m}_i$ can be expressed in terms of the camera matrix $\mathtt{P}$, joined scene point correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$, and scene translation direction $\mathbf{U}$ as

$$\alpha \mathbf{m}_i = \alpha\left(\mathbf{x}_i \times \mathbf{x}'_i\right) = (\mathtt{P}\mathbf{X}_i \times \mathtt{P}\mathbf{X}'_i)/|\mathtt{P}| = (\mathtt{P}\mathbf{X}_i \times \mathtt{P}(\mathbf{X}_i + \mathbf{U}))/|\mathtt{P}| = \mathtt{P}^{-\top}(\mathbf{X}_i + \mathbf{U}), \quad (5.1)$$

where $\alpha \neq 0$ and $|\mathtt{P}| = \det \mathtt{P}$.

Using (5.1) to express the meet of joins $\mathbf{m}_i$ and $\mathbf{m}_j$ in terms of the camera $\mathtt{P}$ and joined scene



(a) GoPro Hero 4 Medium, 21.9mm

(b) GoPro Hero 4 Wide, 17.2mm

(c) 7.5mm

Figure 5.3: *Field-of-View Study.* The proposed solvers give accurate undistortions and rectifications across all fields-of-view. The distorted image of the vanishing line is rendered in green. Left-to-right with increasing levels of distortion: (a) GoPro Hero 4 at the medium-FOV setting, (b) GoPro Hero 4 at the wide-FOV setting, (c) and a Samyang 7.5mm fisheye lens. The outputs are the undistorted (middle row) and rectified images (bottom row). Note the stability of the undistortion estimates for the GoPro images. The rotunda image is rectified from features extracted mostly from the wrought iron fence below the rotunda. Focal lengths are 35mm equivalents.

point correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ and $\mathbf{X}_j \leftrightarrow \mathbf{X}'_j$ gives

$$\begin{aligned} \alpha_i \mathbf{m}_i \times \alpha_j \mathbf{m}_j &= \left(\mathtt{P}^{-\top}(\mathbf{X}_i + \mathbf{U})\right) \times \left(\mathtt{P}^{-\top}(\mathbf{X}_j + \mathbf{U})\right) = \\ \mathtt{P}((\mathbf{X}_i + \mathbf{U}) \times (\mathbf{X}_j + \mathbf{U}))/|\mathtt{P}| &= \mathtt{P}\left(\mathbf{U}^\top(\mathbf{X}_i \times \mathbf{X}_j)\right)\mathbf{U}/|\mathtt{P}| = \beta \mathtt{P}\mathbf{U} = \eta \mathbf{u}, \end{aligned} \tag{5.2}$$

where $\beta = \mathbf{U}^\top(\mathbf{X}_i \times \mathbf{X}_j)/|\mathtt{P}|$, η is non-zero and $\mathbf{U}^\top(\mathbf{X}_i \times \mathbf{X}_j)$ is non-zero for non-degenerate point configurations (see Figure 2.3). In general (5.2) shows that the image of all joined scene point correspondences translating in the same direction meet at the vanishing point of their translation direction, *i.e.* $\eta\mathbf{u} = \beta\mathtt{P}\mathbf{U}$. Note that if correspondence $\mathbf{x}_k \leftrightarrow \mathbf{x}'_k$ from Figure 2.3 were used in lieu of $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$ in (5.2), then $\mathbf{U}^\top(\mathbf{X}_i \times \mathbf{X}_k) = 0$, which implies that $\eta = 0$. This is a degenerate configuration of the solvers and is discussed in detail in Section 5.5.

Since $\mathbf{U}$ is coincident with $\mathbf{l}_\infty$ by construction (see Figure 2.3) and point-line incidence is

	Reference	Rectifies	Undistorts	Motion	# Correspondences Regions	 Points	# Solutions	Size
$\mathtt{H}_2\mathbf{l}$	[82]	✓		translation	1	2	1	closed form
$\mathtt{H}_2\mathbf{l}\lambda$		✓	✓	translation	1	3	4	closed form
$\mathtt{H}_2\mathbf{lu}\lambda$		✓	✓	translation	1	3	4	14×18
$\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$		✓	✓	translation	1	3	2	24×26
$\mathtt{H}_{22}\mathbf{luv}\lambda$		✓	✓	translation	2	4	6	54×60
$\mathtt{H}_{22}\mathbf{luv}s_{\cdot}$		✓	✓	translation	2	4	4	76×80
$\mathtt{H}_{22}\lambda$	[26]		✓	rigid[1]	2	5	18	18×18
$\mathtt{H}_{22}\lambda_1\lambda_2$	[45]		✓	rigid[1]	2	5	5	16×21
$\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$	Chapter 6	✓	✓	rigid	3	9	54	133×187

[1] The preimages of both region correspondences must be related by the same rigid transform in the scene plane.

Table 5.2: *Proposed Solvers (shaded in grey) vs. State of the Art.* The proposed solvers require a few as 1 region correspondence instead of three and are significantly simpler than the undistorting and rectifying solver $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ of Chapter 6. The homography solvers of [26, 45] do not directly recover the vanishing line and require two affine-covariant region correspondences or five points, all of which have the same relative orientation, which restricts sampling.

invariant under projection by $\mathtt{P}$ [32], $\mathbf{u}$ and $\mathbf{l}$ are also coincident,

$$\mathbf{l}^{\top}\mathbf{u} = 0. \tag{5.3}$$

The EVL solver introduced in 5.4.2 uses the relation between conjugately-translated points and vanishing points derived in (5.1) and (5.2) and the vanishing point-vanishing line incidence equation of (5.3) to place constraints on $\mathbf{l}$.

5.3 Radially-Distorted Conjugate Translations

Conjugate translations as defined in (2.23) can be written in terms of radially-distorted conjugately-translated point correspondences undistorted by (2.35) as

$$\alpha f(\tilde{\mathbf{x}}', \lambda) = \mathtt{H}_{\mathbf{u}} f(\tilde{\mathbf{x}}, \lambda) = [\mathtt{I}_3 + s^{\mathbf{u}}\mathbf{u}\mathbf{l}^{\top}] f(\tilde{\mathbf{x}}, \lambda), \tag{5.4}$$

where $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ is a radially-distorted point correspondence that is consistent with the conjugate translation $\mathtt{H}_{\mathbf{u}}$. We call $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ a *radially-distorted conjugately-translated point correspondence* going forward.

Each of the EVP solvers introduced in Section 5.4.1 uses the relation defined in (5.4) and the vanishing point-vanishing line incidence equation of (5.3) to place constraints on $\mathbf{l}$ and λ.

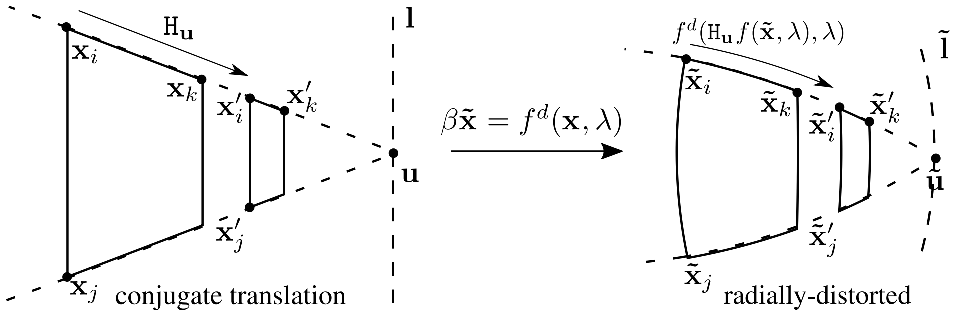


Figure 5.4: *The Geometry of a Radially-Distorted Conjugate Translation.* A translation of coplanar scene points $\{\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k\}$ by $\mathbf{U}$ induces a conjugate translation $\mathtt{H}_{\mathbf{u}}$ in the undistorted image as viewed by camera $\mathtt{P}$, as shown in Figure 2.3. Joined conjugately-translated point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$ and $\mathbf{x}_k \leftrightarrow \mathbf{x}'_k$ must meet at the vanishing point $\mathbf{u}$. Vanishing line $\mathbf{l}$ is the set of all vanishing points of translation directions. The division model images lines as circles, thus the distorted vanishing point $\tilde{\mathbf{u}}$ is given by the intersection of three circles, two of which are coincident with the radially-distorted conjugately-translated point correspondences $\tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{x}}'_i$,$\tilde{\mathbf{x}}_j \leftrightarrow \tilde{\mathbf{x}}'_j$ and $\tilde{\mathbf{x}}_k \leftrightarrow \tilde{\mathbf{x}}'_k$, and the third is given by the distorted vanishing line $\tilde{\mathbf{l}}$. Radially-distorted conjugately-translated points are related by $f^d(\mathtt{H}_{\mathbf{u}} f(\tilde{\mathbf{x}}, \lambda), \lambda)$, where $f^d(\cdot, \lambda)$ is the division-model distortion function.

5.4 Solvers

This chapter proposes five different minimal solvers for different geometric configurations of radially-distorted conjugate translations, which are distinguished by the number of directions and magnitudes of translations that the proposed solver variants admit. These variants are motivated by the types of covariant feature detectors used to extract point correspondences, which give the constraints needed to jointly solve for the division model parameter, vanishing line and the vanishing point of the translation direction(s) [60, 62, 65, 66, 91].

The proposed solvers can be differentiated by the choice to use the hidden variable trick (see Section 4.2 and [18]) to either eliminate the unknown parameters of the vanishing point of the imaged translation direction or the imaged scene plane's vanishing line from the solver's polynomial system of equations. The solvers are eponymously named after their eliminated unknowns: (i) the *eliminated vanishing point* (EVP) solvers (see Section 5.4.1) hide the lens undistortion parameter and vanishing line parameters and have the vanishing point eliminated, and (ii) the *eliminated vanishing line* (EVL) solver (see Section 5.4.2) hides the lens undistortion parameter and eliminates the vanishing line parameters (the vanishing points are recovered by construction). It is interesting to compare the significant differences in solver complexity, time to solution (see Table 5.3), stability (see Figure 5.8) and noise sensitivity (see 5.9) that differs by the elimination choice. Sections 5.4.1 and 5.4.2 detail how either the vanishing point of the translation direction or the vanishing line is eliminated to simplify the systems of polynomial equations that arise from constraints induced by radially-distorted conjugately-translated local features.

The EVP solvers introduced in Section 5.4.2 are grouped by whether they admit one or two directions of radially-distorted conjugate translations. The EVL solver introduced in Section 5.4.2 is a one-direction variant. While it doesn't admit all the various configurations of the EVP solvers, it is the fastest and most robust of the proposed solvers.

5.4.1 The Eliminated Vanishing Point (EVP) Solvers

The model for radially-distorted conjugate translations in (5.4) defines the unknown geometric quantities: (i) division-model parameter λ, (ii) imaged scene-plane vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$, (iii) vanishing point of the translation direction $\mathbf{u} = \left(u_1, u_2, u_3\right)^\top$ (see Section 5.4.1 for the two-direction extensions), (iv) scale of translation $s^{\mathbf{u}}$ for correspondence $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$, (v) and the homogeneous scale parameter α.

The solution for the vanishing line $\mathbf{l}$ is constrained to the affine subspace $l_3 = 1$ of the real-projective plane, which makes it unique. This inhomogeneous choice of $\mathbf{l}$ is unable to represent the pencil of lines that pass through the image origin; however, the degeneracy remains even with a homogeneous representation of $\mathbf{l}$. See Section 5.5 for a more detailed discussion of the degeneracies.

The vanishing direction $\mathbf{u}$ must meet the vanishing line $\mathbf{l}$, which defines a subspace of solutions for $\mathbf{u}$. The magnitude of $\mathbf{u}$ is set to the magnitude of conjugate translation $s^{\mathbf{u}}_1$ of the first correspondence $\tilde{\mathbf{x}}_1 \leftrightarrow \tilde{\mathbf{x}}'_1$, which defines a unique solution

$$\mathbf{l}^\top \mathbf{u} = l_1 u_1 + l_2 u_2 + u_3 = 0 \quad \wedge \quad \|\mathbf{u}\| = s^{\mathbf{u}}_1. \tag{5.5}$$

The relative scale of translation $\bar{s}^{\mathbf{u}}_i$ for each correspondence $\tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{x}}'_i$ with respect to the magnitude of $\|\mathbf{u}\|$ is defined so that $\bar{s}^{\mathbf{u}}_i = s^{\mathbf{u}}_i / \|\mathbf{u}\|$. Note that $\bar{s}^{\mathbf{u}}_1 = 1$. The relationship between magnitude of translation in the scene plane and the magnitude of conjugate translation is derived in the Appendix in the supplemental materials.

Two *one-direction solvers* are proposed, which require 3 radially-distorted conjugately-translated point correspondences. A radially-distorted conjugately-translated affine-covariant region correspondence provides the necessary 3 point correspondences (see Section 3.2.3). Solver $\mathtt{H}_2\mathbf{lu}\lambda$ assumes that all point correspondences have the same relative scales of translation, i.e. $\bar{s}^{\mathbf{u}}_1 = \bar{s}^{\mathbf{u}}_2 = \bar{s}^{\mathbf{u}}_3 = 1$. Solver $\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$ relaxes the equal translation scale assumption of the $\mathtt{H}_2\mathbf{lu}\lambda$ solver. In particular, solver $\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$ assumes that two of the point correspondences have the same magnitude of conjugate translation (i.e. $\bar{s}^{\mathbf{u}}_1 = \bar{s}^{\mathbf{u}}_2 = 1$), and the third point correspondence has an unknown relative scale of the translation $\bar{s}^{\mathbf{u}}_3$. The $\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$ admits combinations of similarity-covariant regions (defining 2 point correspondences) and corner detections for flexible sampling of complementary features.

In addition, two *two-direction solvers* are proposed that require 4 coplanar point correspondences, 2 of which have the vanishing point of translation direction $\mathbf{u}$ and the remaining 2 a different vanishing point $\mathbf{v}$. Two similarity-covariant region correspondences consistent with two radially-distorted conjugate translations provide 2 pairs of 2 point correspondences (see Section 3.2.3) provide the necessary 4 point correspondences.

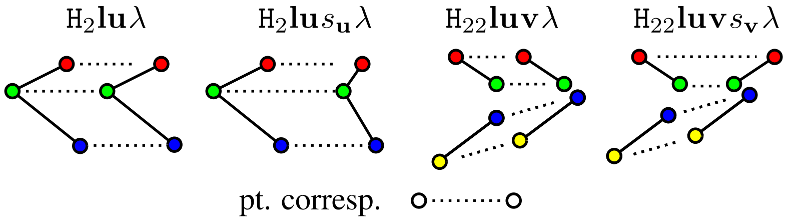

$\texttt{H}_2\mathbf{lu}\lambda$
$\texttt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$
$\texttt{H}_{22}\mathbf{luv}\lambda$
$\texttt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$
pt. corresp.

Figure 5.5: *Input Configurations for the EVP Solvers.* Each of the one-direction solvers—$\texttt{H}_2\mathbf{lu}\lambda$ and $\texttt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$—requires 3 points, which can be obtained from only 1 affine-covariant region correspondence. The $\texttt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$ admits a point correspondence with a unique magnitude of conjugate translation, which provides flexibility when sampling complementary feature correspondences. The two-direction solvers—$\texttt{H}_{22}\mathbf{luv}\lambda$,$\texttt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$–require 4 points, which can be obtained from 2 similarity-covariant feature correspondences. Solver $\texttt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$ admits reflections of similarity-covariant features since $s_{\mathbf{v}}$ allows a point correspondence to move along the line of the imaged translation going through the vanishing point.

Solver $\texttt{H}_{22}\mathbf{luv}\lambda$ requires four points and assumes equal relative scales of conjugate translation in both directions, namely $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = 1$ with respect to $\|\mathbf{u}\| = s_1^{\mathbf{u}}$ and $\bar{s}_3^{\mathbf{v}} = \bar{s}_4^{\mathbf{v}} = 1$ with respect to $\|\mathbf{v}\| = s_3^{\mathbf{v}}$.

Solver $\texttt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$ requires four point correspondences (equivalently, two similarity covariant region correspondences—see Section 3.2.3) and relaxes the assumption of the $\texttt{H}_{22}\mathbf{luv}\lambda$ solver that both point correspondences in the $\mathbf{v}$ direction have the same magnitudes of conjugate translation. In particular, $\texttt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$ assumes that the first two point correspondences translate in the direction $\mathbf{u}$ with the same relative scale of translation, i.e., $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = 1$. The remaining two point correspondences translate in the direction $\mathbf{v}$ with arbitrary translation magnitudes, i.e., the relative scales of translations of these two correspondences with respect to $\|\mathbf{v}\| = s_3^{\mathbf{v}}$ are $\bar{s}_3^{\mathbf{v}} = 1$ and an unknown relative scale $\bar{s}_4^{\mathbf{v}}$. In the case that similarity-covariant regions are extracted from the image and its reflection, reflected covariant regions can be used for jointly solving for undistortion and rectification (see Figure 5.5).

In all of the proposed solvers the scalar values α_i are eliminated from (5.4). This is done by multiplying (5.4) by the skew-symmetric matrix $[f(\tilde{\mathbf{x}}', \lambda)]_\times$. The fact that the join of a point $\mathbf{x}$ with itself $[\mathbf{x}]_\times\mathbf{x}$ is $\mathbf{0}$ gives,

$$\begin{bmatrix} 0 & -\tilde{w}_i' & \tilde{y}_i' \\ \tilde{w}_i' & 0 & -\tilde{x}_i' \\ -\tilde{y}_i' & \tilde{x}_i' & 0 \end{bmatrix} \times \begin{bmatrix} 1+\bar{s}_i^{\mathbf{u}}u_1l_1 & \bar{s}_i^{\mathbf{u}}u_1l_2 & \bar{s}_i^{\mathbf{u}}u_1 \\ \bar{s}_i^{\mathbf{u}}u_2l_1 & 1+\bar{s}_i^{\mathbf{u}}u_2l_2 & \bar{s}_i^{\mathbf{u}}u_2 \\ \bar{s}_i^{\mathbf{u}}u_3l_1 & \bar{s}_i^{\mathbf{u}}u_3l_2 & 1+\bar{s}_i^{\mathbf{u}}u_3 \end{bmatrix} \begin{pmatrix} \tilde{x}_i \\ \tilde{y}_i \\ \tilde{w}_i \end{pmatrix} = \mathbf{0}, \tag{5.6}$$

where $\tilde{w}_i = 1+\lambda(\tilde{x}_i^2+\tilde{y}_i^2)$ and $\tilde{w}_i' = 1+\lambda(\tilde{x}_i'^2+\tilde{y}_i'^2)$. The matrix equation in (5.6) contains three polynomial equations from which only two are linearly independent since the skew-symmetric matrix $[f(\tilde{\mathbf{x}}', \lambda)]_\times$ is rank two.

To solve the systems of polynomial equations resulting from the presented problems, we use the Gröbner basis method [18]. In particular, we used the automatic generators proposed in [44, 48]; however, for our problems the coefficients of the input equations are not fully independent. This means that using the default settings for the automatic generator [44, 48], which initialize the coefficients of equations by random values from $\mathbb{Z}_p$, does not lead to correct solvers. Correct problems instances with values from $\mathbb{Z}_p$ are needed to initialize the automatic generator to obtain working Gröbner basis solvers.

The straightforward application of the automatic generator [44, 48] to the needed constraints with correct coefficients from $\mathbb{Z}_p$ resulted in large templates and unstable solvers, especially for the two-direction problems. The Gröbner basis solvers generated for the original constraints have template matrices with sizes 80×84, 74×76, 348×354, and 730×734 for the $\mathtt{H}_2\mathbf{lu}\lambda$, $\mathtt{H}_2\mathbf{lu}s_\mathbf{u}\lambda$, $\mathtt{H}_{22}\mathbf{luv}\lambda$ and $\mathtt{H}_{22}\mathbf{luv}s_\mathbf{v}\lambda$ problems, respectively. Therefore, we use the hidden-variable trick (see Section 4.2 and [18]) to eliminate the vanishing translation directions together with ideal saturation [49] to eliminate parasitic solutions. The reformulated constraints are simpler systems in only 3 or 4 unknowns, and the solvers generated by the Gröbner basis method are smaller and more stable. The reduced elimination template sizes for the simplified solvers are summarized in Table 5.2, and wall clock timings for the simplified solvers are reported in Section 5.7.2. Optimized C++ implementations for all the proposed solvers are provided.

Next, we describe the solvers based on the hidden-variable trick in more detail.

One-Direction EVP Solvers

For the one-direction $\mathtt{H}_2\mathbf{lu}s_\mathbf{u}\lambda$ solver we have $\bar{s}_1^\mathbf{u} = \bar{s}_2^\mathbf{u} = 1$. Therefore the constraints (5.6) result in two pairs of linearly independent equations without the scale parameter $\bar{s}_i^\mathbf{u}$ for $i = 1, 2$, and two linearly independent equations with an unknown relative scale $\bar{s}_3^\mathbf{u}$ for the third point correspondence, *i.e.*, $i = 3$. Additionally, we have the orthogonality constraint in (5.5). All together we have seven equations in seven unknowns $(l_1, l_2, u_1, u_2, u_3, \bar{s}_3^\mathbf{u}, \lambda)$.

Note, that these equations are linear with respect to the vanishing translation direction $\mathbf{u}$. Therefore, we can rewrite the seven equations as

$$\mathtt{M}(l_1, l_2, \bar{s}_3^\mathbf{u}, \lambda) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{pmatrix} = \mathbf{0}, \tag{5.7}$$

where $\mathtt{M}(l_1, l_2, \bar{s}_3^\mathbf{u}, \lambda)$ is a 7×4 matrix whose elements are polynomials in $(l_1, l_2, \bar{s}_3^\mathbf{u}, \lambda)$.

Since $\mathtt{M}(l_1, l_2, \bar{s}_3^\mathbf{u}, \lambda)$ has a null vector, it must be rank deficient. Therefore, all the 4×4 cofactors of $\mathtt{M}(l_1, l_2, \bar{s}_3^\mathbf{u}, \lambda)$ must equal zero. This results in $\binom{7}{4} = 35$ polynomial equations which only involve four unknowns.

Unfortunately, the formulation (5.7) introduces a one-dimensional family of false solutions. These are not present in the original system and corresponds to solutions where the first three columns of $\mathtt{M}$ become rank deficient. In this case there exist null vectors to $\mathtt{M}$ such that the last element of the vector is zero, *i.e.*, not on the same form as in (5.7).

These false solutions can be removed by saturating [49] any of the 3×3 cofactors from the first three columns of $\mathtt{M}$. The matrix $\mathtt{M}$ has the following form,

$$\mathtt{M}(l_1, l_2, \bar{s}_3^{\mathbf{u}}, \lambda) = \begin{bmatrix} m_{11} & m_{12} & 0 & m_{14} \\ m_{21} & m_{22} & 0 & m_{24} \\ m_{31} & 0 & m_{33} & m_{34} \\ m_{41} & 0 & m_{43} & m_{44} \\ m_{51} & m_{52} & 0 & m_{54} \\ m_{61} & 0 & m_{63} & m_{64} \\ l_1 & l_2 & 1 & 0 \end{bmatrix}, \tag{5.8}$$

where m_{ij} are polynomials in $l_1, l_2, \bar{s}_3^{\mathbf{u}}$ and λ. We choose to saturate the 3×3 cofactor corresponding to the first, second and last row since it reduces to only the top-left 2×2 cofactor, *i.e.*, $m_{11}m_{22} - m_{12}m_{21}$, which is only a quadratic polynomial in the unknowns. The other 3×3 determinants are more complicated and leads to larger polynomial solvers. Using the saturation technique from Larsson et al. [49], we were able to create a polynomial solver for this saturated ideal. The size of the elimination template is 24×26. Note that without using the hidden-variable trick the elimination template was 74×76. The number of solutions is two.

For the $\mathtt{H}_2\mathbf{lu}\lambda$ solver we can use the same hidden-variable trick. In this case $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = \bar{s}_3^{\mathbf{u}} = 1$; therefore, the matrix $\mathtt{M}$ in (5.7) contains only three unknowns l_1, l_2 and λ. This problem is over-constrained, and one of the two constraints from a point correspondence goes unused. Thus, for this problem we can drop one of the equations from (5.6), e.g., for $i = 3$, and the matrix $\mathtt{M}$ in (5.7) has size 6×4. In this case all 4×4 cofactors of $\mathtt{M}$ result in 15 equations in 3 unknowns. Similar to the 3 point case, this introduces a one-dimensional family of false solutions. The matrix $\mathtt{M}$ has a similar structure as in (5.8) and again it is sufficient to saturate the top-left 2×2 cofactor. For this formulation we were able to create a solver with template size 14×18 (compared with 80×84 without using hidden-variable trick). The number of solutions is four.

Two-Direction EVP Solvers

In the case of the two-direction $\mathtt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$ solver, the input equations for two vanishing translation directions $\mathbf{u} = \left(u_1, u_2, u_3\right)^\top$ and $\mathbf{v} = \left(v_1, v_2, v_3\right)^\top$ can be separated into two sets of equations, *i.e.*, the equations containing $\mathbf{u}$ and the equations containing $\mathbf{v}$. Note that in this case we have two equations of the form (5.5), *i.e.*, the equation for the direction $\mathbf{u}$ and the equation for the direction $\mathbf{v}$ and we have an unknown relative scale $\bar{s}_4^{\mathbf{v}}$. Therefore, the final system of 10 equations in 10 unknowns can be rewritten using two matrix equations as

$$\mathtt{M}_1(l_1, l_2, \lambda) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{pmatrix} = \mathbf{0}, \quad \mathtt{M}_2(l_1, l_2, \bar{s}_4^{\mathbf{v}}, \lambda) \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ 1 \end{pmatrix} = \mathbf{0}, \tag{5.9}$$

where $\mathtt{M}_1$ and $\mathtt{M}_2$ are 5×4 matrices such that the elements are polynomials in (l_1, l_2, λ) and $(l_1, l_2, \bar{s}_4^{\mathbf{v}}, \lambda)$, respectively.

Again all 4×4 cofactors of $\mathtt{M}_1$ and $\mathtt{M}_2$ must concurrently equal zero. This results in $5 + 5 = 10$

polynomial equations in four unknowns $(l_1, l_2, \bar{s}_4^{\mathbf{v}}, \lambda)$. In this case, only 39 additional false solutions arise from the hidden-variable trick. The matrices $\mathtt{M}_1$ and $\mathtt{M}_2$ have a similar structure as in (5.8) and again it is sufficient to saturate the top-left 2×2 cofactors to remove the extra solutions. By saturating these determinants we were able to create a solver with template size 76×80 (previously 730×734). The number of solutions is four.

Finally, for the $\mathtt{H}_{22}\mathbf{luv}\lambda$ two-direction solver, $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = 1$ and $\bar{s}_3^{\mathbf{v}} = \bar{s}_4^{\mathbf{v}} = 1$. This problem is over-constrained, so we can drop one of the equations from constraint (5.6), e.g., for $i = 4$. Therefore, the matrix $\mathtt{M}_2$ from (5.9) has size 4×4, and it contains only three unknowns (l_1, l_2, λ). All 4×4 cofactors of $\mathtt{M}_1$ and $\mathtt{M}_2$ result in $5 + 1 = 6$ polynomial equations in three unknowns (l_1, l_2, λ).

For this case we get 18 additional false solutions. Investigations in Macaulay2 [29] revealed that for this particular formulation, it is sufficient to only saturate the top-left 2×2 cofactor of $\mathtt{M}_1$ and the top-left element of $\mathtt{M}_2$. Generating the polynomial solver with saturation resulted in a template size of 54×60 (previously 348×354). The number of solutions is six.

5.4.2 The Eliminated Vanishing Line (EVL) Solver

Suppose $\{\, \tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{x}}'_i \,\}_{i=1}^3$ are point correspondences extracted from a radially-distorted conjugately-translated affine-covariant region correspondence as shown in Figure 5.6. Then their preimages $\{\, \mathbf{X}_i \leftrightarrow \mathbf{X}'_i \,\}_{i=1}^3$ on the scene plane Π are in correspondence with a translation, denote it $\mathbf{U}$, which is color coded cyan in Figure 5.6. This point configuration has three additional translation directions $\mathbf{V}_1$,$\mathbf{V}_2$ and $\mathbf{V}_3$, (colored red, green and blue, respectively), where each of the four imaged translation directions induces four radially-distorted conjugate translations in the distorted image.

A vanishing point, *i.e.*, $\mathbf{u}$, $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, can be recovered from each meet of joins (see Section 5.2) of pairs of conjugate translations that share the same translation direction in the scene plane, *e.g.*,

$$\gamma \mathbf{v}_1 = (\mathbf{x}_1 \times \mathbf{x}_3) \times (\mathbf{x}'_1 \times \mathbf{x}'_3). \tag{5.10}$$

There are six such pairs to choose from, one for each of $\mathbf{v}_1$,$\mathbf{v}_2$ and $\mathbf{v}_3$ and three for $\mathbf{u}$, which is the vanishing point of the translation direction for the undistorted point correspondences $\{\, \mathbf{x}_i \leftrightarrow \mathbf{x}'_i \,\}_{i=1}^3$.

As proved in Section 5.2, each meet of joins puts a constraint on the vanishing line $\mathbf{l}$. It will be shown that only three of the six vanishing point constructions are necessary to solve for the undistortion parameter λ and vanishing line $\mathbf{l}$. It will also be shown that exactly one of any of the three meets of joins of conjugate translations from $\{\, \mathbf{x}_i \leftrightarrow \mathbf{x}'_i \,\}_{i=1}^3$ can be used to constrain $\mathbf{l}$ (see Section 5.4.2).

Without loss of generality, we use the joins of pairs of conjugate translations meeting at $\mathbf{v}_1$,$\mathbf{v}_2$, and $\mathbf{v}_3$, which are substituted into the vanishing point-vanishing line incident constraint of (5.3)

$$\mathbf{v}_i^\top \mathbf{l} = \left((\mathbf{x}_i \times \mathbf{x}_j) \times (\mathbf{x}'_i \times \mathbf{x}'_j) \right)^\top \mathbf{l} = 0, \tag{5.11}$$

where $i < j$ and $i, j \in \{\, 1 \ldots 3 \,\}$. The homogeneity of (5.11) is used to eliminate any non-zero

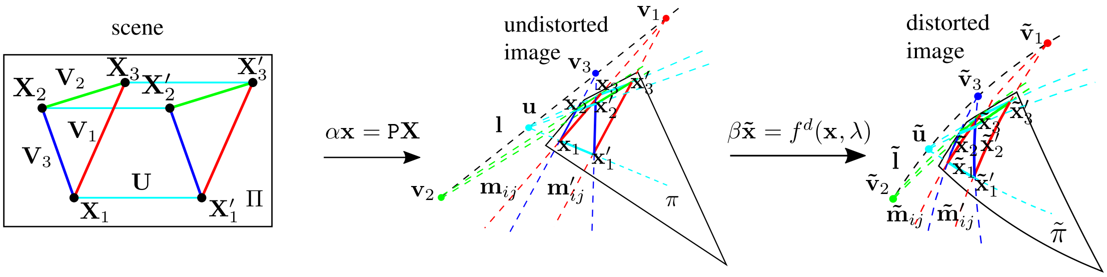


Figure 5.6: *The Geometry of the EVL Constraints.* The scene plane Π contains the preimage of radially-distorted conjugately-translated affine-covariant regions, equivalently, 3 translated points in the direction $\mathbf{U}$. This configuration had 3 additional translation directions $\mathbf{V}_1$,$\mathbf{V}_2$,$\mathbf{V}_3$ that can be used to design a solver. In the image plane π, the joins of each of the images of the 3 pairs of parallel lines (colored red, green and blue) meet at the imaged scene plane's vanishing line $\mathbf{l}$. Each incidence of a vanishing point $\mathbf{u}$,$\mathbf{v}_1$,$\mathbf{v}_2$ and $\mathbf{v}_3$ with $\mathbf{l}$ generates a scalar constraint equation. Two equations are needed to estimate $\mathbf{l}$ and three are necessary to jointly estimate $\mathbf{l}$ and λ. Note that $\mathbf{u}$ can be estimated from one of 3 meets of distinct joins of undistorted point correspondences, but only 1 such meet can be used as a Constraint to estimate the rectification (see Section 5.4.2 for details).

scalars. Substituting radially-distorted points for undistorted points in (5.11) using (2.35) gives

$$(f(\tilde{\mathbf{x}}_i,\lambda) \times f(\tilde{\mathbf{x}}_j,\lambda)) \times \left(f(\tilde{\mathbf{x}}_i',\lambda) \times f(\tilde{\mathbf{x}}_j',\lambda)\right)^{\top} \mathbf{l} = 0. \tag{5.12}$$

The skew-symmetric operator, denoted $[\cdot]_\times$, is used to transform (5.12) into the homogeneous matrix-vector equation

$$\left([[f(\tilde{\mathbf{x}}_i,\lambda)]_\times f(\tilde{\mathbf{x}}_j,\lambda)]_\times [f(\tilde{\mathbf{x}}_i',\lambda)]_\times f(\tilde{\mathbf{x}}_j',\lambda)\right)^{\top} \mathbf{l} = 0, \tag{5.13}$$

where where $i < j$ and $i,j \in \{1\ldots3\}$. Independent scalar constraint equations of the form (5.13) can be stacked to add the necessary number of constraints for jointly estimating $\mathbf{l}$ and λ.

Creating the Solver

Each vanishing point $\mathbf{u}$,$\mathbf{v}_1$,$\mathbf{v}_2$ and $\mathbf{v}_3$ generates one scalar constraint on the vanishing line $\mathbf{l}$. There are four unknowns in constraint (5.13), namely $\mathbf{l} = \left(l_1,l_2,l_3\right)^{\top}$ and the division model parameter λ (see Section 2.11). The vanishing line $\mathbf{l}$ is homogeneous, so it has only two degrees of freedom. Thus 3 scalar constraint equations of the form (5.13) generated by 3 vanishing points from the set $\{\mathbf{u},\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ are needed, which, as shown in (5.13), can be concisely

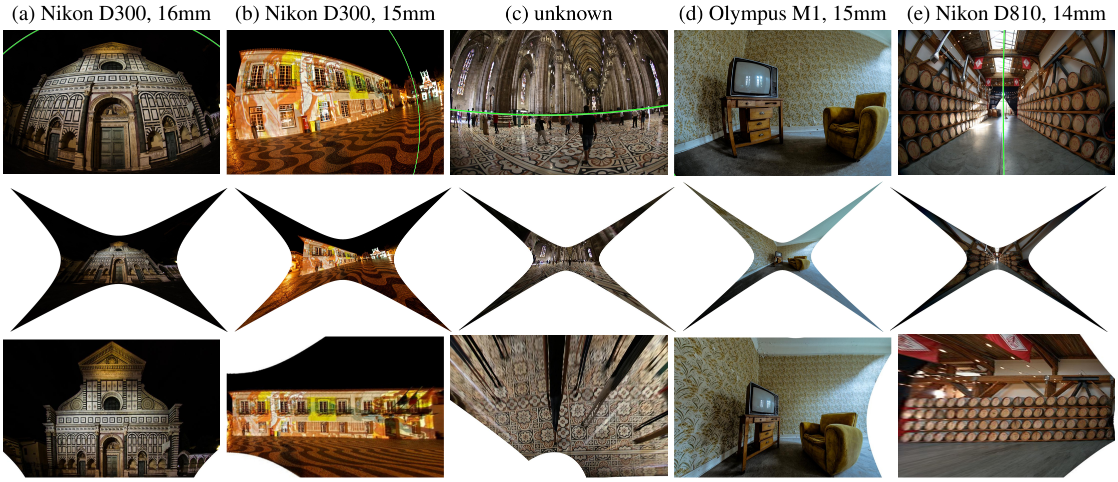

(a) Nikon D300, 16mm (b) Nikon D300, 15mm (c) unknown (d) Olympus M1, 15mm (e) Nikon D810, 14mm

Figure 5.7: *EVL Solver Results on Fisheye Images.* The distorted image of the vanishing line is rendered in green in the input images on the top row. Results were produced using the $\mathtt{H}_2\mathbf{l}\lambda$ with 1-correspondence sampling in a RANSAC framework. The $\mathtt{H}_2\mathbf{l}\lambda$ solver runs in 0.5 µs. Surprisingly, reasonable rectifications are possible using the 1-parameter division model for the extreme distortions of fisheye lenses. Focal lengths are reported as 35mm equivalent.

encoded in the matrix $\mathtt{M}(\lambda) \in \mathbb{R}^{3\times3}$ as

$$\mathtt{M}(\lambda)\begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \mathbf{0}. \tag{5.14}$$

Note that only 1 of the 3 meets of joins of conjugately-translated point correspondences from $\{\,\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\,\}_{i=1}^{3}$ can be used since there is no constraint included that enforces

$$((\mathbf{x}_i \times \mathbf{x}'_i) \times (\mathbf{x}_j \times \mathbf{x}'_j)) \times ((\mathbf{x}_i \times \mathbf{x}'_i) \times (\mathbf{x}_k \times \mathbf{x}'_k)) = \mathbf{0},$$

where $i, j, k \in \{\,1\ldots3\,\}$ and $i \neq j$. Therefore, at least two of $\mathbf{v}_1$,$\mathbf{v}_2$, and $\mathbf{v}_3$ must be used, and the two chosen meets can be combined with exactly one of the meets the can be constructed from $\{\,\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\,\}_{i=1}^{3}$. Including the case where each of $\mathbf{v}_1$,$\mathbf{v}_2$, and $\mathbf{v}_3$ is used gives $3\binom{3}{2} + 1 = 10$ possible combinations of meets. Selecting the optimal meets for the most accurate rectification is addressed in Section 5.4.2.

The division model parameter λ is hidden in (5.14) using the hidden-variable trick (see Section 4.2 and [18]) in the entries of coefficient matrix $\mathtt{M}$, which are polynomials only in λ. Thus $\mathbf{l}$ has been eliminated, which motivates the EVL name.

Matrix $\mathtt{M}(\lambda)$ is rank deficient since it has a null vector, which implies that $\det\mathtt{M}(\lambda) = 0$. The determinant constraint defines a univariate quartic with unknown λ, which can be solved in closed form. After λ has been recovered, the vanishing line $\mathbf{l}$ is obtained by solving for the null space of $\mathtt{M}$. The EVL solver is denoted $\mathtt{H}_2\mathbf{l}\lambda$.

Best Minimal Solution Selection

The EVL geometry of Figure 5.6 has 10 meets that can be used to generate scalar constraint equations in (5.13). However, only 3 meets are needed to jointly estimate $\mathbf{l}$ and λ. Since the time to solution for the $\mathtt{H}_2\mathbf{l}\lambda$ is only $\mathbf{0.5}$ $\mathbf{\mu s}$, the solutions for all minimal subsets of meets can be verified against the unused constraints, *e.g.*, if the meets of joins of the radially-distorted conjugately-translated correspondences associated with $\mathbf{v}_1$,$\mathbf{v}_2$, and $\mathbf{v}_3$ are used, then the correspondences associated with $\mathbf{u}$ (cyan direction) can be used for verification. The minimal subset of meets is chosen that minimizes the sum of symmetric transfer errors

$$\sum_i d(\tilde{\mathbf{x}}_i, f^d(\mathtt{H}^{-1}f(\tilde{\mathbf{x}}'_i,\lambda),\lambda))^2 + d(f^d(\mathtt{H}f(\tilde{\mathbf{x}}_i,\lambda),\lambda),\tilde{\mathbf{x}}'_i)^2, \tag{5.15}$$

where $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ are radially-distorted conjugately-translated point correspondences that are not included in a minimal configuration for estimating rectification. We call this approach *best minimal solution selection*.

Evaluating the quality of the minimal solution on (5.15) has several benefits: 1. Near degenerate correspondence configurations can be rejected (see Section 5.5.1), 2. Correspondences with geometric properties that are more robust to noise will be preferred, *e.g.*, regions that are further apart, 3. and expensive RANSAC consensus set construction can be preempted, if there is no minimal solution that has sufficiently small symmetric transfer error as defined in (5.15).

Best minimal solution selection is evaluated in the sensitivity studies in Section 5.7. The solver incorporating best minimal solution selection is denoted in the standard way, $\mathtt{H}_2\mathbf{l}\lambda$. For comparison we introduce a baseline solver, denoted $\mathtt{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$, which randomly selects from the 10 possible constraint configurations associated with the EVL geometry (see Figure 5.6). As expected, the $\mathtt{H}_2\mathbf{l}\lambda$ performs better than $\mathtt{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$ on all sensitivity measures. See Section 5.7.1 for the details.

Optimal Estimate of the Vanishing Point

Unlike the EVP solvers in Section 5.4.1, which jointly estimate the vanishing point $\mathbf{u}$ (shown in Figure 5.6) using all constraints from the set of conjugate translations $\{\,\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\,\}_{i=1}^3$ (see (5.6)), the $\mathtt{H}_2\mathbf{l}\lambda$ solver maximally uses two joins from $\{\,\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\,\}_{i=1}^3$ and possibly none if only the red, green and blue translation directions in Figure 5.6 are selected as the best minimal solution.

The vanishing point $\mathbf{u}$ of the cyan translation direction can be recovered after the vanishing line $\mathbf{l}$ and division model parameter λ are estimated (*e.g.*, by $\mathtt{H}_2\mathbf{l}\lambda$) by solving a constrained least squares system that includes all constraints induced by $\{\,\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\,\}_{i=1}^3$ (see Figure 5.6). The incidence of $\mathbf{u}$ with $\mathbf{l}$ is explicitly enforced by including (5.3) into the constraints. Define $\mathbf{h}_{\mathbf{u}}^{1\top}$, $\mathbf{h}_{\mathbf{u}}^{2\top}$, and $\mathbf{h}_{\mathbf{u}}^{3\top}$ to be the rows of a conjugate translation,

$$\alpha\mathbf{x}' = \mathtt{H}_{\mathbf{u}}\mathbf{x} = \begin{bmatrix}\mathbf{h}_{\mathbf{u}}^1 & \mathbf{h}_{\mathbf{u}}^2 & \mathbf{h}_{\mathbf{u}}^3\end{bmatrix}^\top \mathbf{x} = \begin{bmatrix}\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top\end{bmatrix}\mathbf{x}. \tag{5.16}$$

The homogeneous scale in (5.16) can be eliminated by substituting $\mathbf{h}_{\mathbf{u}}^{3\top}\mathbf{x}$ for α, and the system can be rearranged such that

$$\begin{aligned} \mathbf{x}^{\top}\mathbf{h}_{\mathbf{u}}^{1} &= (x'\mathbf{x}^{\top})\mathbf{h}_{\mathbf{u}}^{3} \\ \mathbf{x}^{\top}\mathbf{h}_{\mathbf{u}}^{2} &= (y'\mathbf{x}^{\top})\mathbf{h}_{\mathbf{u}}^{3}. \end{aligned} \tag{5.17}$$

Collecting the terms of vanishing point after expanding the dot products in (5.17) for each pair of $\{\, \mathbf{x}_i \leftrightarrow \mathbf{x}_i' \,\}_{i=1}^{3}$ along with an incidence constraint $\mathbf{l}^{\top}\mathbf{u} = 0$ gives the constrained least squares problem

$$\begin{aligned} &\underset{\mathbf{u}}{\text{minimize}} \;\; \|\mathtt{M}\mathbf{u} - \mathbf{y}\|^2 \\ &\text{subject to} \;\; \mathbf{l}^{\top}\mathbf{u} = 0, \end{aligned}$$

$$\text{where} \quad \mathtt{M} = \begin{bmatrix} & \vdots & \\ -\mathbf{l}^{\top}\mathbf{x}_i & 0 & x'(\mathbf{l}^{T}\mathbf{x}_i) \\ 0 & \mathbf{l}^{\top}\mathbf{x}_i & y'(\mathbf{l}^{\top}\mathbf{x}_i) \\ & \vdots & \end{bmatrix}, \quad \mathbf{y} = \begin{pmatrix} \vdots \\ x_i - x' \\ y_i - y' \\ \vdots \end{pmatrix}$$

Since the matrix $\begin{bmatrix}\mathtt{M}^{\top} & \mathbf{l}\end{bmatrix}^{\top}$ has linearly independent columns, and $\mathbf{l}^{\top}$ is trivially row independent, $\mathbf{u}$ is recovered by solving

$$\begin{bmatrix} \mathtt{M}^{\top}\mathtt{M} & \mathbf{l} \\ \mathbf{l}^{\top} & 0 \end{bmatrix} \begin{pmatrix} \mathbf{u} \\ z \end{pmatrix} = \begin{pmatrix} \mathtt{M}^{\top}\mathbf{y} \\ 0 \end{pmatrix}, \tag{5.18}$$

where z is a nuisance variable [8]. Surprisingly, a superior estimation of the vanishing point $\mathbf{u}$ is given by using (5.18) after rectifying with the EVL $\mathtt{H}_2\mathbf{l}\lambda$ solver than by jointly solving for the rectification, vanishing point, and division model parameter as done with the EVP group of solvers (see the transfer error sensitivity study Figure 5.9a).

5.5 Degeneracies

We identified three important degeneracies for the solvers: Section 5.5.1 describes two geometric configurations of features such that there exists either a subspace of rectifications or no valid solution, and Section 5.5.2 details the modeling degeneracy introduced from using the representation of (2.10) for the affine-rectifying homography, which requires $\mathbf{l} = \left(l_1, l_2, l_3\right)^{\top}$ such that $l_3 \neq 0$ [32]. The proposed solvers and the state-of-the-art solvers of Chapter 6 all suffer from this modeling degeneracy. It is shown that addressing this degeneracy requires increasing the complexity of the solvers. There are likely additional degeneracies between the EVL and EVP solver, but an exhaustive analysis is a difficult theoretical problem.

5.5.1 Degenerate Feature Configurations

Suppose that: (i) $\mathtt{H}$ is a rectifying homography other than the identity matrix, (ii) that the image has no radial distortion, (iii) and that all corresponding points from repeated affine-covariant regions fall on a single circle centered at the image center. Applying the division model (see Section 2.11) uniformly scales the points about the image center. Given $\lambda \neq 0$, for a transformation by $f(\cdot, \lambda)$ defined in (2.35) of the points lying on the circle there is a scaling matrix $\mathtt{S}(\lambda) = \operatorname{diag}(1/\lambda, 1/\lambda, 1)$ that maps the points back to their original positions. Thus there is a 1D family of rectifying homographies given by $\mathtt{HS}(\lambda)$ for the corresponding set of undistorted images given by $f(\cdot, \lambda)$.

Secondly, suppose that the conjugately-translated point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ and $\mathbf{x}_k \leftrightarrow \mathbf{x}'_k$ are collinear as shown in Figure 2.3. Let $\mathbf{m}_i = \mathbf{x}_i \times \mathbf{x}'_i$ and $\mathbf{m}_k = \mathbf{x}_k \times \mathbf{x}'_k$. Then $\mathbf{m}_i \times \mathbf{m}_k = \mathbf{0}$, which is not a point in the real-projective plane $\mathbb{RP}^2$, and cannot be used to place a constraint on $\mathbf{l}$. Unfortunately, this point configuration is common, *e.g.*, consider a row of windows on a facade. It is possible that the feature extraction pipeline will establish collinear correspondences. However, affine frames constructed from covariant region detections are typically not in this degenerate configuration since the origin is defined by blob's center of mass or peak response in scale space and one of the extents is constructed as a right angle to the first linear basis vector (see Figure 3.11). Regardless, the degeneracy can be avoided by using different meets.

5.5.2 The Pencil of Vanishing Lines Through the Distortion Center

If the vanishing line passes through the image origin, *i.e.* $\mathbf{l} = \left(l_1, l_2, 0\right)^\top$, then the radial term in the homogeneous coordinate of (2.36) is canceled. In this case, it is not possible to recover the division model parameter λ from the systems of equations (5.8), (5.9) or (5.14) solved by any of the proposed solvers. However, the degeneracy does not arise from the problem formulation. An affine transform can be applied to the undistorted image such that the vanishing line $\mathbf{l}$ in the affine-transformed space has $l_3 \neq 0$.

The division model requires the image origin to be the distortion center [26]. The derivations in this chapter assume that image center, distortion center and image origin are coincident (see Section 2.11). The proposed solvers and the state-of-the-art solvers of Chapter 6 formulate joint undistortion and rectification in terms of (2.36), which leaves the distortion center stationary.

Directional cameras see only points in front of the camera [30], so the vanishing line cannot intersect the convex hull of measurements. Therefore, changing basis in the undistorted space such that any point in the convex hull of the undistorted feature points (*i.e.*, affine covariant region detections) is the image origin guarantees that vanishing line will not pass through the origin. Furthermore, if a point is in the convex hull of measurements in the distorted space, then it is also in the convex hull of undistorted measurements. However, the change of basis (*i.e.*, a translation) is a function of the undistorted point, and thus a function of the unknown division model parameter λ, so applying the coordinate transform increases the complexity of the solvers. Empirically we did not find this degeneracy to be a problem. *E.g.*, Figures 5.3c, 5.7c, and 5.7e show good undistortions of images and rectifications of imaged scene planes that have vanishing lines passing close to the center of distortion, which suggest that in these near-degenerate cases the division-model parameter is sufficiently observable. Thus we choose to preserve the simplicity of the solvers (see Table 5.3). A new origin in the undistorted space can be defined by a distorted measurement in the convex hull of measurements, which will reduce the chance of encountering the degeneracy, but not eliminate it.

5.6 Robust Estimation for Radially-Distorted Conjugate Translations

The solvers are used in a LO-RANSAC-based robust-estimation framework [15, 74]. Affine rectifications and undistortions are jointly hypothesized by one of the proposed solvers. A metric upgrade is attempted and models with maximal consensus sets are locally optimized by an extension of the method introduced in [74]. The metric-rectifications are presented in the results.

5.6.1 Local Features and Descriptors

We use the Maximally-Stable Extremal Region and Hessian-Affine detectors as detailed in Sections 3.2.4 and 3.2.6 [62, 65]. The affine-covariant regions are given by an affine basis (see Section 3.2.3), equivalently three distinct points, in the image space [69]. The image patch local to the affine frame is embedded into a descriptor vector by the RootSIFT transform [4, 60] (see Section 3.3.1). See Figure 3.11 for a visualization.

5.6.2 Detection, Description, and Clustering

Affine-covariant regions are clustered by appearance as described in Section 3.3.2. Since the proposed $\mathtt{H}_2\mathbf{lu}\lambda$, $\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$, $\mathtt{H}_2\mathbf{l}\lambda$, and $\mathtt{H}_{22}\mathbf{luv}\lambda$ solvers do not admit reflections, the appearance-clusters are partitioned based on the handedness of the affine frames associated with the clustered embedded regions. Reflection partitioning is not necessary for the $\mathtt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$, which admits reflections of similarity-covariant regions.

5.6.3 Sampling

Sample configurations for the proposed minimal solvers are illustrated in Figures 5.1, 5.5, and 5.6 as well as detailed in Sections 5.4.1 and 5.4.2. For each RANSAC trial, appearance clusters are selected with the probability given by its relative cardinality to the other appearance clusters, and the required number of correspondences are drawn from the selected clusters.

5.6.4 Metric Upgrade and Local Optimization

The affine-covariant regions that are members of the minimal sample are affine rectified by each feasible model returned by the solver; typically there is only 1. Correspondences for the selected solver are sampled as detailed in Section 5.6.3. The affine rectification estimated by the minimal solver is used to build an affine-rectified scale consensus set. The scale consensus set is

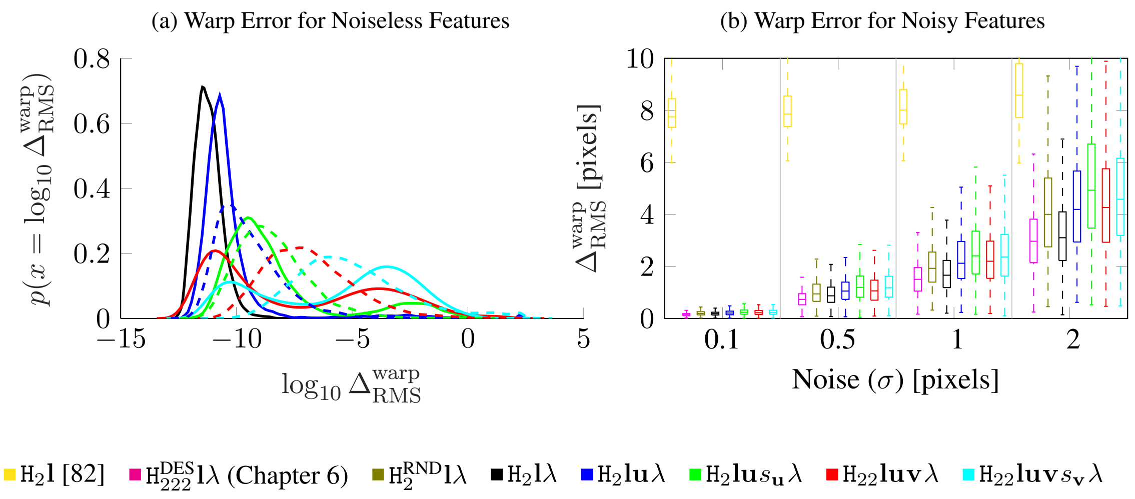


Figure 5.8: *Warp Error Stability and Sensitivity Studies.* (a) Hidden-variable trick solvers are solid; solvers generated without simplified constraints equations are dashed. The $\log_{10}$ RMS warp error $\Delta^{\mathrm{warp}}_{\mathrm{RMS}}$ is reported for noiseless scenes generated as described in Sections 2.12 and 5.7.1. The hidden-variable trick increases stability. The EVL $\texttt{H}_2\mathbf{l}\lambda$ solver is the most stable since it does not require solving a complicated polynomial system of equations. (b) Reports the RMS error $\Delta^{\mathrm{warp}}_{\mathrm{RMS}}$ (see Section 2.12) after 25 iterations of a simple RANSAC for the bench of solvers with increasing levels of white noise added to the affine-covariant region correspondences, where the normalized division model parameter is set to -4 (see Section 2.11), which is similar to the distortion of a GoPro Hero 4. Results are for radial-distorted conjugate translations. The proposed solvers demonstrate excellent robustness to noise, and the EVL solver $\texttt{H}_2\mathbf{l}\lambda$ is competitive with $\texttt{H}^{\mathrm{DES}}_{222}\mathbf{l}\lambda$, which requires two more correspondences. The $\texttt{H}_2\mathbf{l}\lambda$ solver uses best minimal solution selection (see Section 5.4.2), which improves its performance compared to $\texttt{H}^{\mathrm{RND}}_{2}\mathbf{l}\lambda$, which randomly selects a solution.

built by using the scale constraint of affine-rectified space: two instances of rigidly-transformed coplanar repeats occupy identical areas in the scene plane and in the affine rectified image of the scene plane [20, 14, 32], see also Chapter 6. Note that if clustered left and right-handed regions were partitioned for sampling with the $\texttt{H}_2\mathbf{lu}\lambda$, $\texttt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$, $\texttt{H}_2\mathbf{l}\lambda$, and $\texttt{H}_{22}\mathbf{luv}\lambda$ solvers, then they are merged so they are jointly verified for scale consistency. Absolute scales are calculated to account for handedness. The log-scale ratio of the each region in a cluster is computed with respect to the median affine-rectified scale. Note that covariant regions extracted from imaged rigidly-transformed coplanar texture can enter the scale consensus set since they will be equi-scalar after affine rectification, too. This admits the possibility of a full-metric upgrade. Regions with near 0 log-scale ratio with respect to the median scale of their cluster are considered tentatively inlying, and are used as inputs to the metric upgrade of Pritts et al. [74], which restores congruence.

The congruence consensus set is measured in the metric-rectified space by verifying the congruence of the linear basis vectors of the corresponded affine frames. Congruence is an invariant of metric rectified space and is a stronger constraint than, *e.g.*, the equal-scale invariant of affine-rectified space that was used to derive the solvers proposed in [14] and in Chapter 6. The metric upgrade essentially comes for free by inputting the covariant regions that are members of the scale consensus set to the linear metric-upgrade solver proposed in [74]. By using the metric upgrade, the verification step of RANSAC can enforce the congruence of corresponding covariant region extents (equivalently, the lengths of the linear basis vectors) to estimate an accurate consensus set. A model with the maximal congruence consensus set at the current RANSAC iteration is locally optimized in a method similar to [74].

5.7 Experiments

The stabilities and noise sensitivities of the proposed solvers are evaluated on synthetic data. We compare the proposed solvers to a bench of the four state-of-the-art solvers (see Table 5.2). We apply the denotations for the solvers introduced in Section 2.2 to all the solvers in the benchmark; *e.g.*, a solver requiring two correspondences of two affine-covariant regions will be prefixed by $\mathtt{H}_{22}$.

Included is the state-of-the-art joint undistorting and rectifying solver $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ of Chapter 6, which requires 3 correspondences of affine-covariant regions extracted from the image of rigidly-transformed coplanar repeated scene textures. While 6 variants of undistorting and rectifying solvers are proposed in Chapter 6, we test only the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver since all variants are reported to have similar noise sensitivities.

The bench includes the $\mathtt{H}_{2}\mathbf{l}$ solver of Schaffalitzky et al. [82], which incorporates similar constraints from conjugate translations that are used to derive the proposed solvers. Also included are two full-homography and radial-undistortion solvers, the $\mathtt{H}_{22}\lambda$ solver of Fitzgibbon et al. [26] and the $\mathtt{H}_{22}\lambda_1\lambda_2$ solver of Kukelova et al. [45], which are used to assess the benefits of jointly solving for radially-distorted conjugate translations (and lens undistortion) from the minimal problem, as done with the proposed solvers, versus the over-parameterized problem as in [26, 45]. The bench of state-of-the-art solvers is summarized in Table 5.2.

The sensitivity studies evaluate the solvers on noisy measurements over 3 task-related performance metrics: 1. the transfer error (see Section 5.7.1 and Figure 5.9a), which measures the accuracy of radially-distorted conjugate translation estimation 2. the root mean square warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ (see Figure 5.8b and Section 2.12), which measures rectification accuracy, and 3. the relative error Figure 5.9b of the division-model parameter estimate, which reports the accuracy of the lens undistortion estimate.

The stability study Figure 5.8a evaluates the proposed solvers by the warp error on noiseless measurements. The study demonstrates the benefit of constraint simplification by the hidden-variable trick (see Section 4.2 and [18]), which is used to derive both the EVP solvers and EVL solver, and shows that it improves the stability of all solvers, and, in fact, it is sometimes necessary to generate usable solvers.

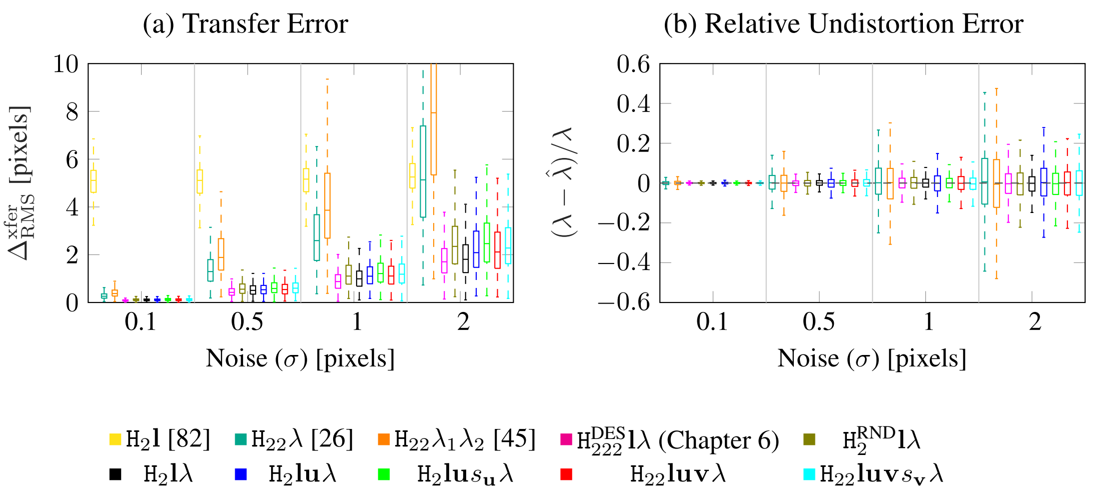


Figure 5.9: *Transfer Error and Undistortion Error For Noisy Features.* Comparison of two error measures after 25 iterations of a simple RANSAC for different solvers with increasing levels of white noise added to the affine covariant region correspondences, where the normalized division model parameter is set to -4 (see Sec. 3.1), which is similar to the distortion of a GoPro Hero 4. Results are for translated coplanar repeats. (a) Reports the root mean square transfer error Section 5.7.1) With the exception of the $\mathtt{H}^{\mathrm{DES}}_{222}\mathbf{l}\lambda$ solver, the proposed solvers are significantly more robust for both types of repeats on both error measures; however $\mathtt{H}^{\mathrm{DES}}_{222}\mathbf{l}\lambda$ requires the most correspondences, and (b) reports the relative error of the estimated division model parameter. The $\mathtt{H}_2\mathbf{l}\lambda$ solver uses best minimal solution selection (see Section 5.4.2), which improves its performance compared to $\mathtt{H}^{\mathrm{RND}}_{2}\mathbf{l}\lambda$, which randomly selects a solution.

5.7.1 Synthetic Data

The performance of the proposed solvers on 1000 synthetic images of 3D scenes with known ground-truth parameters is evaluated. A camera with a random but realistic focal length is randomly placed with respect to a scene plane such that it is mostly in the camera's field-of-view. The image resolution is set to 1000x1000 pixels. The noise sensitivities of the solvers are evaluated on conjugately-translated coplanar repeats (see Figure 5.9a, 5.8b, and 5.9b). Affine frames (see Section 3.2.3) are generated on the scene plane such that their scale with respect to the scene plane is realistic. The modeling choice reflects the use of affine-covariant region detectors on real images. The image is distorted according to the division model. For the sensitivity experiments, isotropic white noise is added to the distorted affine frames at increasing levels.

Transfer Error

The geometric transfer error of Figure 5.9a measures the accuracy of the estimated radially-distorted conjugate translation (see Section 5.3). The scene plane is tessellated by a 10x10 grid of points with a 1 unit spacing between adjacent points. The tessellation ensures that the imaged scene plane is uniformly covered by features. In this way, the accuracy of the estimated radially-distorted conjugate translation can be measured across most of the image. Denote the tessellation as $\{\, \mathbf{X}_i \,\}_{i=1}^{100}$. Suppose that $\mathbf{x} \leftrightarrow \mathbf{x}'$ are conjugately-translated points consistent with $\mathtt{H}_{\mathbf{u}} = [\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top]$. Points $\{\, \mathbf{X}_i \,\}_{i=1}^{100}$ are translated to $\{\, \mathbf{X}'_i \,\}_{i=1}^{100}$ by 1 unit on the scene plane in the direction given by the translation direction $\mathbf{U}$. The conjugate translation $\mathtt{H}_{\mathbf{u}}$ is not used directly because its translation magnitude may span the extent of the scene plane, so applying it to the tessellation would transform the grid out of the field of view.

The preimage of the translation direction is $\beta\mathbf{U} = \mathtt{P}^{-1}\mathbf{u} = \beta\left(u_x, u_y, 0\right)^\top$. Then $\|\mathbf{U}\|$ is the magnitude of translation between the repeated scene elements in the scene-plane coordinate system. Define the homogeneous translation matrix defined by $\mathbf{U}$ to be

$$\mathtt{T}\left(\left(t_x, t_y, \alpha\right)^\top\right) = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.19}$$

The translation of the grid points by unit distance in the scene plane coordinate system is given by $\mathbf{X}' = \mathtt{T}(\mathbf{U}/\|\mathbf{U}\|)\mathbf{X}$. Recall from (2.22) that a conjugate translation has the form $\mathtt{PT}(\cdot)\mathtt{P}^{-1}$. Using the decomposition in (2.23), the conjugate translation of unit distance in the direction of point correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ is

$$\mathtt{H}_{\mathbf{u}/\|\mathbf{U}\|} = \mathtt{P}\mathtt{I}_3\mathtt{P}^{-1} + \mathtt{P}\begin{pmatrix} u_x/\|\mathbf{U}\| \\ u_y/\|\mathbf{U}\| \\ 0 \end{pmatrix} \left[\mathtt{P}^{-\top}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right]^\top = [\mathtt{I}_3 + \frac{\mathbf{u}}{\|\mathbf{U}\|}\mathbf{l}^\top]. \tag{5.20}$$

The unit conjugate translation $\mathtt{H}_{\mathbf{u}/\|\mathbf{U}\|}$ can be written in terms of the conjugate translation $\mathtt{H}_{\mathbf{u}}$ induced by the undistorted point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ as

$$\mathtt{I}_3 + \frac{\mathbf{u}}{\|\mathbf{U}\|}\mathbf{l}^\top = \mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}[\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top - \mathtt{I}_3] = \mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}[\mathtt{H}_{\mathbf{u}} - \mathtt{I}_3]. \tag{5.21}$$

Thus the radial distorted conjugate translation of unit distance is given by

$$\gamma\tilde{\mathbf{x}}' = f^d([\mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}(\hat{\mathtt{H}}_{\mathbf{u}} - \mathtt{I}_3)]f(\tilde{\mathbf{x}}, \lambda), \lambda), \tag{5.22}$$

where f^d is the function that transforms from pinhole points to radially-distorted points.

The imaged grid is given by $\tilde{\mathbf{x}}_i = f^d(\mathtt{P}\mathbf{X}_i, \lambda)$ and the translated grid by $\tilde{\mathbf{x}}'_i = f^d(\mathtt{P}\mathbf{X}'_i, \lambda)$. Then the geometric transfer error is defined as

$$\Delta^{\text{xfer}} = d(f^d([\mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}(\hat{\mathtt{H}}_{\mathbf{u}} - \mathtt{I}_3)]f(\tilde{\mathbf{x}}, \hat{\lambda}_1), \hat{\lambda}_2), \tilde{\mathbf{x}}'), \tag{5.23}$$

where $d(\cdot,\cdot)$ is the Euclidean distance.

All solvers except $\mathtt{H}_{22}\lambda_1\lambda_2$ have the constraint that $\hat{\lambda}_1 = \hat{\lambda}_2$ [45]. The root mean square

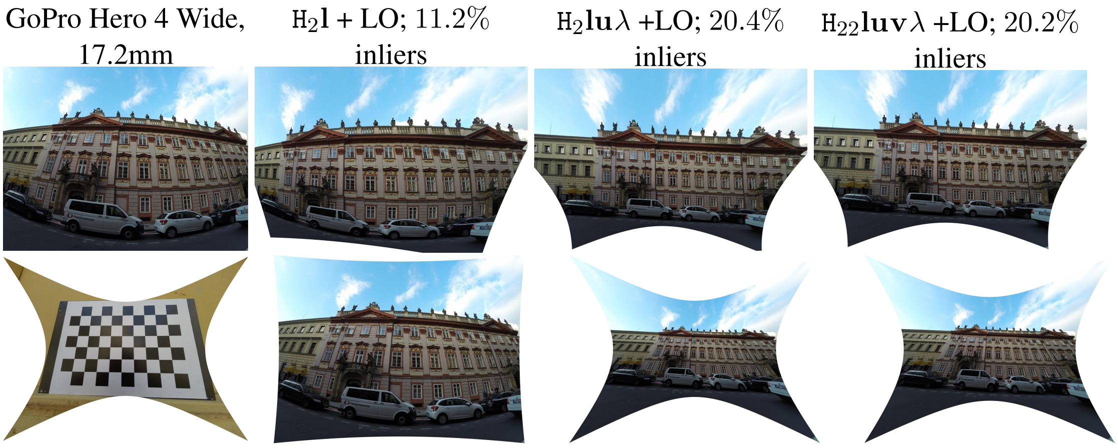

GoPro Hero 4 Wide, 17.2mm
$\mathtt{H}_2\mathbf{l}$ + LO; 11.2% inliers
$\mathtt{H}_2\mathbf{lu}\lambda$ +LO; 20.4% inliers
$\mathtt{H}_{22}\mathbf{luv}\lambda$ +LO; 20.2% inliers

Figure 5.10: *GoPro Hero 4 at the wide setting for different solvers.* Results from LO-RANSAC (see Section 5.6) for $\mathtt{H}_2\mathbf{l}$, which omits distortion, and the proposed solvers $\mathtt{H}_2\mathbf{lu}\lambda$ and $\mathtt{H}_{22}\mathbf{luv}\lambda$. The top row has rectifications after local optimization (LO); The bottom row has undistortions estimated from the best *minimal* sample. LO-RANSAC cannot recover from the poor initializations by $\mathtt{H}_2\mathbf{l}$ (column 2). The proposed solvers in columns 3 and 4 give a correct rectification. The bottom left has a chessboard undistorted using the division parameter estimated from the building facade by $\mathtt{H}_2\mathbf{lu}\lambda$ +LO.

transfer error $\Delta^{\text{xfer}}_{\text{RMS}}$ for radially-distorted conjugately-translated correspondences $\tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{x}}'_i$ is reported. For two-direction solvers, the transfer error in the second direction is included in $\Delta^{\text{xfer}}_{\text{RMS}}$. The transfer error is used in the sensitivity study, where the solvers are tested over varying noise levels with a fixed division model parameter.

Numerical Stability

The stability study Figure 5.8a measures the RMS warp error $\Delta^{\text{warp}}_{\text{RMS}}$ of solvers (see Section 2.12) for noiseless radially-distorted conjugately-translated affine frame correspondences across realistic scene and camera configurations generated as described in the introduction to this section. The normalized ground-truth division-model parameterλ is drawn uniformly at random from the interval $[-6, 0]$. For a reference, the division parameter of $\lambda = -4$ is typical for wide field-of-view cameras like the GoPro Hero 4, where the image is normalized by $1/(\text{width} + \text{height})$. Figure 5.8a reports the histogram of $\log_{10}$ warp errors $\Delta^{\text{warp}}_{\text{RMS}}$.

For the proposed EVP solvers we evaluate a solver generated from constraints derived with (solid histogram) and without (dashed histogram) the hidden-variable trick (see Section 4.2). The hidden-variable trick significantly improves the stability of the proposed solvers. The increased stabilities of the hidden-variable solvers (see Figure 5.8a) most likely result from the reduced size of the Gauss-Jordan elimination problems needed by these solvers. The hidden-variable EVP solvers are used for the remainder of the experiments. The proposed EVL solver

$\mathtt{H}_2\mathbf{l}\lambda$ is derived with the hidden-variable trick as well, which results in a quartic. The superior stability of the $\mathtt{H}_2\mathbf{l}\lambda$ solver (see Figure 5.8a) demonstrates the benefits of the elementary formulation.

Noise Sensitivity

The proposed and state-of-the-art solvers are tested with increasing levels of white noise added to the points extracted (see Section 3.2.3) from the radially-distorted conjugately-translated affine-covariant region correspondences (see Figure 5.9). The amount of white noise is given by the standard deviation of a zero-mean isotropic Gaussian distribution, and the solvers are tested at noise levels of $\sigma \in \{0.1, 0.5, 1, 2\}$. The ground-truth normalized division model parameter is set to $\lambda = -4$, which is typical for GoPro-type imagery in normalized image coordinates.

The solvers are wrapped by a basic RANSAC estimator that minimizes either the RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ (see Figure 5.8b), the RMS transfer error (see Figure 5.9a) $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$, or the relative error of lens distortion (see Figure 5.9b) over 25 minimal samples of affine frames. The RANSAC estimates are summarized in boxplots for 1000 synthetic scenes. The interquartile range is contained within the extents of a box, and the median is the horizontal line dividing the box.

The proposed solvers—$\mathtt{H}_2\mathbf{lu}\lambda$,$\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$,$\mathtt{H}_{22}\mathbf{luv}\lambda$,$\mathtt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$, and $\mathtt{H}_2\mathbf{l}\lambda$—demonstrate excellent robustness to noisy features across all three error measures. In particular, the $\mathtt{H}_2\mathbf{l}\lambda$ solver is the least sensitive to noise of the proposed solvers and gives the best undistortion estimates of any solver in the bench (see Figure 5.9b). Figure 5.8b shows that at the 2 pixel noise level, all the proposed solvers rectify with less than 5 pixel RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ more than half the time. Figure 5.9a shows that radially-distorted conjugate translations are estimated with less than 3 pixel RMS transfer error $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$ error more than half the time. All proposed solvers estimate the correct lens distortion parameter more than half the time (see Figure 5.9b) with the $\mathtt{H}_2\mathbf{l}\lambda$ performing the best of any solver in the bench on this study.

For both the warp error and transfer error studies, the $\mathtt{H}_2\mathbf{l}$ solver of Schaffalitzky et al. [82] shows significant bias since it does not model lens distortion, making it essentially unusable as a minimal solver at GoPro-like levels of radial lens distortion. As expected, the overparmeterized radial-distortion homography solvers of $\mathtt{H}_{22}\lambda$ [26] Fitzgibbon and $\mathtt{H}_{22}\lambda_1\lambda_2$ [45] of Kukelova et al. have significantly higher transfer errors with respect to the proposed solvers, which suggests that the extraneous degrees of freedom are used to explain feature noise by incorrect geometry. In fact, at the two pixel noise level of the transfer error study in Figure 5.9a, the performance of these solvers is worse than the $\mathtt{H}_2\mathbf{l}$ solver, which does not model radial lens distortion.

The state-of-the art solver $\mathtt{H}_{222}^{\mathsf{DES}}\mathbf{l}\lambda$ of Chapter 6 shows slightly better noise robustness than the proposed solvers on the warp and transfer error sensitivity studies. However, the proposed solvers are competitive and require fewer correspondences. In particular, the $\mathtt{H}_2\mathbf{l}\lambda$ reaches near parity with the $\mathtt{H}_{222}^{\mathsf{DES}}\mathbf{l}\lambda$ solver and requires only one region correspondence versus three required by the $\mathtt{H}_{222}^{\mathsf{DES}}\mathbf{l}\lambda$ solver. As is shown in Section 5.7.2, the proposed solvers are magnitudes faster in wall clock time. Given their competitive performance in the sensitivity studies and the fact that they require fewer correspondences and have faster times to solution, the proposed solver should be preferred to the $\mathtt{H}_{222}^{\mathsf{DES}}\mathbf{l}\lambda$ solver for images with radially-distorted conjugate translations.

Note that the $\mathtt{H}_{22}\lambda$ solver of [26] and the $\mathtt{H}_{22}\lambda_1\lambda_2$ solver of [45] are omitted from the warp

Solver	Wall Clock	Relative Speed	Template Size
$\mathtt{H}_2\mathbf{l}\lambda$	$\mathbf{0.5}$ **µs**	$\mathbf{1.0}\times$	N/A
$\mathtt{H}_2\mathbf{lu}\lambda$	3.7 µs	$7.4\times$	14×18
$\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$	6.1 µs	$12.2\times$	24×26
$\mathtt{H}_{22}\mathbf{luv}\lambda$	34.6 µs	$69.2\times$	54×60
$\mathtt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$	66.1 µs	$132.2\times$	76×80
$\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ (Chapter 6)	1076.8 µs	$2153.6\times$	133×187

Table 5.3: *Runtime Analysis.* Wall-clock times are reported for optimized C++ implementations of the proposed solvers versus $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ of Chapter 6, which was the only competitive solver from the noise sensitivity experiments. The EVL solver is $2153.6\times$ faster than $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$, and the other proposed variants are orders of magnitude faster.

error since the vanishing line is not directly estimated.

Each of the $\mathtt{H}_2\mathbf{l}\lambda$ and $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solvers requires the ex-post estimation of vanishing point of the translation direction, which is accomplished by the method proposed in Section 5.4.2. Surprisingly, the sequential estimation used by the proposed $\mathtt{H}_2\mathbf{l}\lambda$ and the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver of Chapter 6 achieve the best performances on the transfer error $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$. This is explainable by the improved performance of the $\mathtt{H}_2\mathbf{l}\lambda$ EVL solver with respect to the EVP solvers on all measures, and the fact that the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver uses three correspondences, the most of any in the bench of solvers (see Table 5.2).

The benefit of best minimal solution selection as proposed in (5.4.2) can be seen by comparing the $\mathtt{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$ and $\mathtt{H}_2\mathbf{l}\lambda$ solvers in all sensitivity studies. To quickly recap, The $\mathtt{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$ solver randomly selects a minimal solution from 10 possible solutions given by the EVL geometry shown in Figure 5.6, while the $\mathtt{H}_2\mathbf{l}\lambda$ chooses the solution that minimizes a geometric error on the unused constraints (see Section 5.4.2 for details). The sensitivity improvements using minimal solution selection are considerable: at the 2 pixel noise levels, the RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ (Figure 5.8b) and RMS transfer error (Figure 5.9a) decreased by 26% and 28%, respectively, and the interquartile range of division model parameter estimates decreased by 61%. In fact, the incorporation of best minimal solution selection puts the performance of the $\mathtt{H}_2\mathbf{l}\lambda$ solver on par with the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver, which requires two more region correspondences.

5.7.2 Computational Complexity

Table 5.3 lists the wall-clock time to solution for the optimized C++ implementations of the proposed solvers and the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver of Chapter 6, which was the only competitive solver from the sensitivity experiments reported in Figures 5.9a, 5.8b, and 5.9b. Also reported for easy comparison are the relative speeds with respect to the $\mathtt{H}_2\mathbf{l}\lambda$ solver and the elimination template sizes, where applicable. The proposed EVL $\mathtt{H}_2\mathbf{l}\lambda$ solver is an astounding $2153.6\times$ faster than the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver and significantly faster than all EVP solvers ($\mathtt{H}_2\mathbf{lu}\lambda$,$\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$,$\mathtt{H}_{22}\mathbf{luv}\lambda$, and $\mathtt{H}_{22}\mathbf{luv}s_{\mathbf{v}}\lambda$), which require the Gröbner basis method to solve polynomial systems of equations. All of the proposed solvers are much faster than the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver, making them more suitable

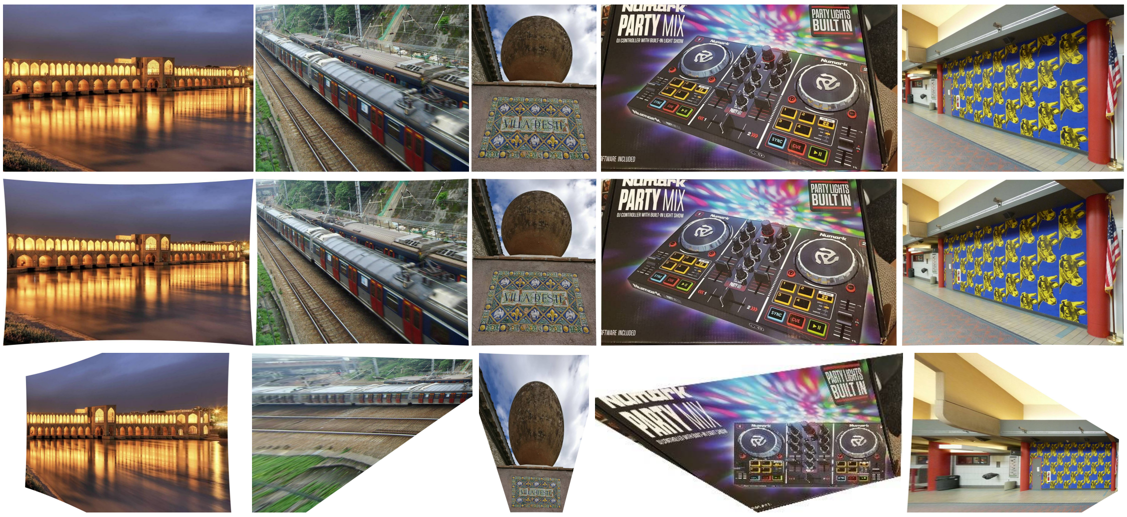


Figure 5.11: *Narrow Field of View and Diverse Scene Content.* The proposed solvers works well if the input image has little or no radial lens distortion. This imagery is typical of consumer cameras and mobile phone cameras. The images are diverse and contain unconventional scene content. Input images are on the top row; undistorted images are on the middle row, and the rectified images are on the bottom. Results were generated with the $\texttt{H}_2\mathbf{lu}\lambda$ solver.

for fast sampling in RANSAC for scenes containing translational symmetries.

5.7.3 Real Images

In the experiments on real images shown in Figures 5.1 and 5.3, we tested the proposed solvers on GoPro4 Hero 4 images with increasing field-of-view settings—medium and wide, where the wider field-of-view setting generates more extreme radial distortion since the full extent of the lens is used. To span the gamut of lens distortions in the field-of-view study of Figure 5.3, we included a Samyang 7.5mm fisheye lens. The consistency of the undistortion estimate at the same GoPro Hero4 field-of-view setting can be seen by comparing the undistortions between the medium GoPro Hero 4 images in Figure 5.3a and the undistortions between the wide GoPro images in Figures 5.1 and 5.3b. Despite significantly different image content and sensor orientation, the undistortions are of comparable magnitude at the same setting. Rectification are accurate for all GoPro Hero 4 images, and the image of the distorted vanishing line is correctly positioned (rendered in green) in the original images. Despite using the 1-parameter division model for lens undistortion (see Section 2.11), an excellent rectification is achieved for the fisheye distorted image taken with the Samyang 7.5mm lens in Figure 5.3c, and the horizon line is perfectly estimated.

Figure 5.7 shows results obtained with 1-correspondence sampling using the proposed $\texttt{H}_2\mathbf{l}\lambda$ EVL solver on very challenging fisheye images. Images from five distinct fisheye lenses are evaluated with Figures 5.7b, 5.7c, and 5.7e having highly oblique viewpoints of the dominant scene plane. Accurate rectifications and undistortions are achieved for all images, and the distorted image of the vanishing line (rendered in green) is correctly positioned. The limitations of the 1-parameter division model can be seen with extreme radial distortions, as, *e.g.*, Figures 5.7c and 5.7d exhibit some mustache distortion, which cannot be modeled with 1 parameter. However, the local optimizer of [74] could be modified to regress a higher-order distortion model using the results of Figure 5.7 as an initial guess. We leave this for future work.

Figures 5.3c, 5.7c, and 5.7e contain imaged scene planes with vanishing lines that pass near the image origin (equivalently, center of distortion), which is a degeneracy of the solver (see Section 5.5). Still excellent results are achieved, which empirically demonstrates that even for vanishing lines passing very close to the image center, the lens distortion is sufficiently observable. In practice the degeneracy does not seem to be a problem.

The experiment shown in Figure 5.10 compares the performance of two of the proposed solvers $\mathtt{H}_2\mathbf{lu}\lambda$ and $\mathtt{H}_{22}\mathbf{luv}\lambda$ to the conjugate translation solver $\mathtt{H}_2\mathbf{l}$ of Schaffalitzky et al. [82] in the coplanar repeat detection and rectification framework of Pritts et al. [74] (see Sec. 5.6) with a GoPro Hero 4 image at the wide field-of-view setting. The two proposed solvers accurately estimate the division-model parameter (see the undistorted reference chessboard in Figure 5.10) and the rectification, while the estimation framework using the $\mathtt{H}_2\mathbf{l}$ solver is unable to recover the lens distortion parameter. The rectification quality is also reflected by the number of inlying features found, which is nearly double for the proposed solvers with respect to the solver of [82]. The experiment demonstrates the non-convexity of the problem, and emphasizes the need for a good initial guess by the minimal solver for the local optimizer of [74].

The narrow field of view and diverse content experiment of Figure 5.11 shows the performance of the proposed method on imagery typical from cell phone cameras and near rectilinear lenses. The left 3 columns of the study are challenging since the conjugate translations and reflections are extracted a small strip of the image. Still the rectifications are accurate.

5.8 Discussion

This chapter proposes a suite of simple high-speed solvers for jointly undistorting and affine-rectifying images containing radially-distorted conjugate translations. The proposed solvers contain variants that relax the assumptions that the preimages of radially-distorted conjugately-translated point correspondences are translated by the same magnitude in the scene plane, and that all point correspondences translate in the same direction. Furthermore, a variant is proposed that admits reflections of similarity-covariant region correspondences, which is helpful for searching for correspondences for semi-metric rectification.

The EVL $\mathtt{H}_2\mathbf{l}\lambda$ solver admits the same point configuration as the one-direction EVP solver $\mathtt{H}_2\mathbf{lu}\lambda$, but is much simpler (*i.e.*, does not require the Gröbner bases method), more stable, and is 7.4× faster in terms of wall-clock time to solution. The improvement is given by the choice to eliminate the vanishing line instead of the vanishing point. The significant difference emphasizes the importance of care in solver design; in particular, the need to simplify the constraint equations. While Gröbner bases related methods are powerful and somewhat general, their blind application for solver generation can result in slow and unstable solvers. *E.g.*, in Chapter 6 we were unable to reduce the degree of their constraint equations used for the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver, which resulted in slow solver (see Table 5.3). Furthermore, stability sampling was required to generate useful solvers [54].

Synthetic experiments show that the EVP and EVL solvers are significantly more robust to noise in terms of the accuracy of rectification and radially-distorted conjugate translation estimation than the radial-distortion homography solvers of Fitzgibbon and Kukelova et al. [26, 45]. The experiment verifies the importance of solving the minimal problem since the extraneous degrees of freedom of the radial-distortion homography solvers are free to explain the noise with incorrect geometry. Furthermore, the proposed solvers are competitive with the robustness of the state-of-the-art $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver of Chapter 6 despite the fact that the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver requires two more region correspondences as input (compared to $\mathtt{H}_2\mathbf{l}\lambda$, $\mathtt{H}_2\mathbf{lu}\lambda$, and $\mathtt{H}_2\mathbf{lu}s_{\mathbf{u}}\lambda$). The advantage of the proposed solvers is more pronounced if the combinatorics of the robust RANSAC estimator are considered, where one correspondence sampling makes it possible to solve scenes with a very-low proportion of good correspondences.

Experiments on difficult images with large radial distortions confirm that the solvers give high-accuracy rectifications if used inside a robust estimator. By jointly estimating rectification and radial distortion, the proposed minimal solvers eliminate the need for sampling lens distortion parameters in RANSAC.

6 Minimal Solvers for Rectifying from Radially-Distorted Scales and Change of Scales

This chapter introduces the first minimal solvers that jointly estimate lens distortion and affine rectification from the image of rigidly-transformed coplanar features. The solvers work on scenes without straight lines and, in general, relax strong assumptions about scene content made by the state of the art. The proposed solvers use the affine invariant that coplanar repeats have the same scale in rectified space. The solvers are separated into two groups that differ by how the equal scale invariant of rectified space is used to place constraints on the lens undistortion and rectification parameters. We demonstrate a principled approach for generating stable minimal solvers by the Gröbner basis method, which is accomplished by sampling feasible monomial bases to maximize numerical stability. Synthetic and real-image experiments confirm that the proposed solvers demonstrate superior robustness to noise compared to the state of the art. Accurate rectifications on imagery taken with narrow to fisheye field-of-view lenses demonstrate the wide applicability of the proposed method. The method is fully automatic.

6.1 Introduction

The state of the art has several approaches for rectifying (or partially calibrating) a distorted image, but these methods make restrictive assumptions about scene content by assuming, *e.g.*, the presence of sets of parallel scene lines [3, 94] or translational symmetries (Chapter 5). The proposed solvers relax the need for specific assumptions about scene content to unknown repeated structures (see Table 6.1).

The proposed minimal solvers exploit the scale constraint: two instances of rigidly-transformed coplanar repeats occupy identical areas in the scene plane and in the affine rectified image of the scene plane (*e.g.*, see the rectifications in Figures 6.2, 6.3, and 6.4). There are two groups of solvers introduced in this chapter: the *directly-encoded-scale* and *change-of-scale* solvers, which are differentiated by the way in which the scale constraint is used. The *directly-encoded-scale* solvers, which we acronymize as the DES solvers for short, encode the unknown area of a rectified region as a dependent function of the measured region, vanishing line, and undistortion parameter (see Section 6.3). The *change-of-scale* solvers – CS solvers for short – linearize the undistorting and rectifying transformation and use its Jacobian determinant to induce constraints on the unknown undistortion and rectification parameters (see Section 6.4). The Jacobian determinant measures the local change-of-scale of the rectifying transformation (and, more generally, of any differentiable transformation).

There are three different minimal configurations of corresponding features that provide a suf-

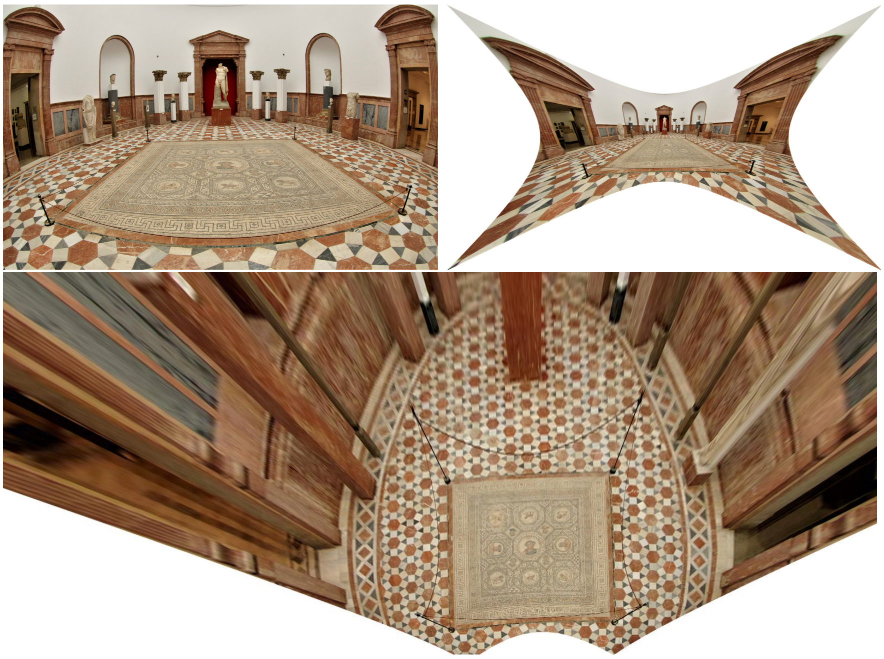


Figure 6.1: *Inputs and Outputs.* Input (top left) is a distorted view of a scene plane, and the outputs (top right, bottom) are the undistorted and rectified scene plane. The method is fully automatic.

ficient number of constraints to solve for the unknown undistortion and rectification parameters (see Section 6.3.3). The minimal configurations are shown in Figure 6.4 and are the same for the DES and CS groups of solvers. We generate solvers for all input configurations for both groups of solvers to provide for flexible sampling during robust estimation. The solvers are fast and robust to noisy feature detections, so they work well in robust estimation frameworks like RANSAC [24].

6.1.1 Previous Work

Several state-of-the-art methods can rectify from imaged coplanar repeated texture, but these methods assume the pinhole camera model [1, 2, 14, 20, 61, 71, 98]. A subset of these methods introduce solvers that use algebraic constraints induced by the equal-scale invariant of affine-rectified space [14, 20, 71] in a similar formulation to the proposed solvers (see Figure 6.2). These methods are members of the change-of-scale (CS) solver group (see Section 6.4) since they use the Jacobian determinant of the affine-rectifying transformation to induce constraints on the imaged scene plane's vanishing line. To complete the family of affine-rectifying minimal solvers for pinhole cameras [14, 20, 71], we also construct and evaluate a novel DES solver that assumes the pinhole camera model in Section 6.3.5.

Chapter 5 proposes minimal solvers that jointly estimate affine rectification and lens undistor-

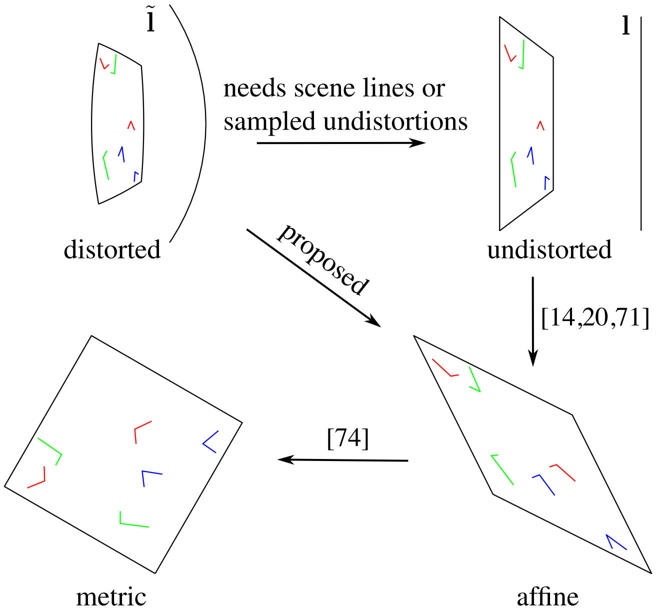


Figure 6.2: *A Shortcut to Affine Rectification.* The hierarchy of rectifications from distorted to metric space is traversed clockwise from the top left. The proposed method is a direct path to affine-rectified space using only rigidly-transformed coplanar repeats, in contrast to the state of the art, which requires scene lines or sampled undistortions. The scene plane's vanishing line is shown in the original and undistorted image ($\tilde{\mathbf{l}}$ and $\mathbf{l}$, respectively). The affine-covariant regions are in the 222-configuration (see Section 6.3.2), where corresponded coplanar regions are the same color. All affine-rectified images are metrically upgraded with the method of [74] for presentation (see Section 6.5.3).

tion, but this method is restricted to scene content with translational symmetries (see Table 6.1). We show that the conjugate translation solvers of Chapter 5 are more noise sensitive than the proposed scale-based solvers (see Figures 6.7 and 6.8).

6.2 Preliminaries

The polynomial system of equations encoding the rectifying constraints is solved using an algebraic method based on Gröbner bases. Automated solver generators using the Gröbner basis method [44, 48] have been used to generate solvers for several camera geometry estimation problems [44, 45, 48, 49], see also Chapter 5. However, the straightforward application of automated solver generators to the proposed constraints resulted in unstable solvers (see Section 6.6). Recently, Larsson et al. [54] sampled feasible monomial bases, which can be used in the action-matrix method. In [54] basis, sampling was used to minimize the size of the solver. We modified

	Wildenauer et al. [94]	Antunes et al. [3]	Chapter 5	Proposed
Feature Type	fitted circles	fitted circles	affine-covariant	affine-covariant
Assumption	3 & 3 parallel lines	3 & 4 parallel lines	2 trans. repeats	4 repeats
Rectification	multi-model	multi-model	direct	direct

Table 6.1: *Scene Assumptions.* Solvers [94, 3] require distinct sets of parallel scene lines as input and multi-model estimation for rectification. Solvers of Chapter 5 are restricted to scenes with translational symmetries. The proposed solvers directly rectify from as few as 4 rigidly transformed repeats (also see Figure 6.4).

the objective of [54] to maximize for solver stability. Stability sampling generated significantly more numerically stable solvers (see Fig. 6.6).

6.3 The Directly-Encoded Scale (DES) Solvers

The proposed DES solvers use the invariant that rectified coplanar repeats have equal scales. In Sections 6.3.1 and 6.3.2 the equal-scale invariant is used to formulate a system of polynomial constraint equations on rectified coplanar repeats with the vanishing line and radial undistortion parameter as unknowns. The radial lens undistortion function is parameterized with the one-parameter division model as defined in Section 2.11. Affine-covariant region detections are used to model repeats since they encode the necessary geometry for scale estimation (see Figure 6.4 and Section 6.5.1). The geometry of an affine-covariant region is uniquely given by an affine frame (see Section 6.3.1). The solvers require 3 points from each detected region to measure the region's scale in the image space. The scale of the rectified coplanar repeat is defined as the area of the triangle defined by the 3 rectified points that represent a corresponding affine-covariant region.

Three minimal cases exist for the joint estimation of the vanishing line and division-model parameter (see Figure 6.4 and Section 6.3.2). These cases differ by the number of affine-covariant regions needed for each detected repetition. The method for generating the minimal solvers for the three variants is described in Section 6.3.4. Finally, in Section 6.3.5, we show that if the undistortion parameter is given, then the constraint equations simplify, which results in a small solver for estimating rectification under the pinhole camera assumption.

6.3.1 Equal Scales Constraint from Rectified Affine-Covariant Regions

Let $\begin{bmatrix}\tilde{\mathbf{x}}_{i,1} & \tilde{\mathbf{x}}_{i,2} & \tilde{\mathbf{x}}_{i,3}\end{bmatrix}$ be the point parameterization of an affine-covariant region $\tilde{\mathcal{R}}_i$ detected in a radially-distorted image (see Section 3.2.3 for a discussion on the point parameterization of a covariant region). Then, by (2.36), the point parameterization of an affine-rectified image of $\tilde{\mathcal{R}}_i$—namely $\underline{\mathcal{R}}_i$—is

$$\begin{bmatrix}\mathtt{H}f(\tilde{\mathbf{x}}_{i,1},\lambda) & \mathtt{H}f(\tilde{\mathbf{x}}_{i,2},\lambda) & \mathtt{H}f(\tilde{\mathbf{x}}_{i,3},\lambda)\end{bmatrix} = \begin{bmatrix}\alpha_{i,1}\underline{\mathbf{x}}_{i,1} & \alpha_{i,2}\underline{\mathbf{x}}_{i,2} & \alpha_{i,3}\underline{\mathbf{x}}_{i,3}\end{bmatrix}, \qquad (6.1)$$

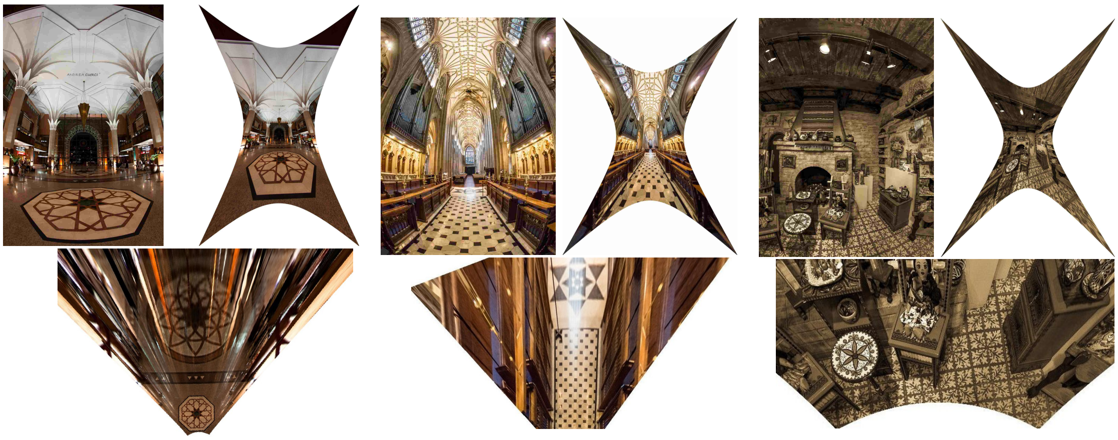


Figure 6.3: *Wide-Angle Results.* Input (top left) is an image of a scene plane. Outputs include the undistorted image (top right) and rectified scene planes (bottom row). The method is automatic.

where $\alpha_{i,j} = \mathbf{l}^{\top} f(\tilde{\mathbf{x}}_{i,j}, \lambda)$. Thus the scale $\underline{s}_i$ of $\underline{\mathcal{R}}_i$ is given as an area of a triangle defined by points in (6.1) as

$$\begin{aligned}\underline{s}_i &= \frac{\det\left(\begin{bmatrix}\alpha_{i,1}\underline{\mathbf{x}}_{i,1} & \alpha_{i,2}\underline{\mathbf{x}}_{i,2} & \alpha_{i,3}\underline{\mathbf{x}}_{i,3}\end{bmatrix}\right)}{\alpha_{i,1}\alpha_{i,2}\alpha_{i,3}} = \frac{1}{\alpha_{i,1}\alpha_{i,2}\alpha_{i,3}} \cdot \begin{vmatrix} \tilde{x}_{i,1} & \tilde{x}_{i,2} & \tilde{x}_{i,3} \\ \tilde{y}_{i,1} & \tilde{y}_{i,2} & \tilde{y}_{i,3} \\ \alpha_{i,1} & \alpha_{i,2} & \alpha_{i,3} \end{vmatrix} \\ &= \frac{\begin{vmatrix} \tilde{x}_{i,2} & \tilde{x}_{i,3} \\ \tilde{y}_{i,2} & \tilde{y}_{i,3} \end{vmatrix}}{\alpha_{i,2}\alpha_{i,3}} - \frac{\begin{vmatrix} \tilde{x}_{i,1} & \tilde{x}_{i,3} \\ \tilde{y}_{i,1} & \tilde{y}_{i,3} \end{vmatrix}}{\alpha_{i,1}\alpha_{i,3}} + \frac{\begin{vmatrix} \tilde{x}_{i,1} & \tilde{x}_{i,2} \\ \tilde{y}_{i,1} & \tilde{y}_{i,2} \end{vmatrix}}{\alpha_{i,1}\alpha_{i,2}}.\end{aligned} \tag{6.2}$$

The numerators of the second and third expressions in (6.2) depend only on the undistortion parameter λ and l_3 due to cancellations in the determinant. The sign of $\underline{s}_i$ depends on the handedness of the detected affine-covariant region. See Section 6.3.7 for a method to use reflected affine-covariant regions with the proposed solvers.

6.3.2 Eliminating the Rectified Scales

The affine-rectified scale in $\underline{s}_i$ (6.2) is a function of the unknown undistortion parameter λ and vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^{\top}$. This encoding of the rectified scale is the motivation for calling this solver group the Directly-Encoded Scale (DES) solvers. A unique solution to (6.2) can be defined by restricting the vanishing line to the affine subspace $l_3 = 1$ or by fixing a rectified scale, *e.g.*, $\underline{s}_1 = 1$. The inhomogeneous representation for the vanishing line is used since it results in degree 4 constraints in the unknowns λ, l_1, l_2 and $\underline{s}_i$ as opposed to fixing a rectified scale, which results in complicated equations of degree 7.

Let $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ be repeated affine-covariant region detections. Then the scales $\underline{s}_i$ and $\underline{s}_j$ of

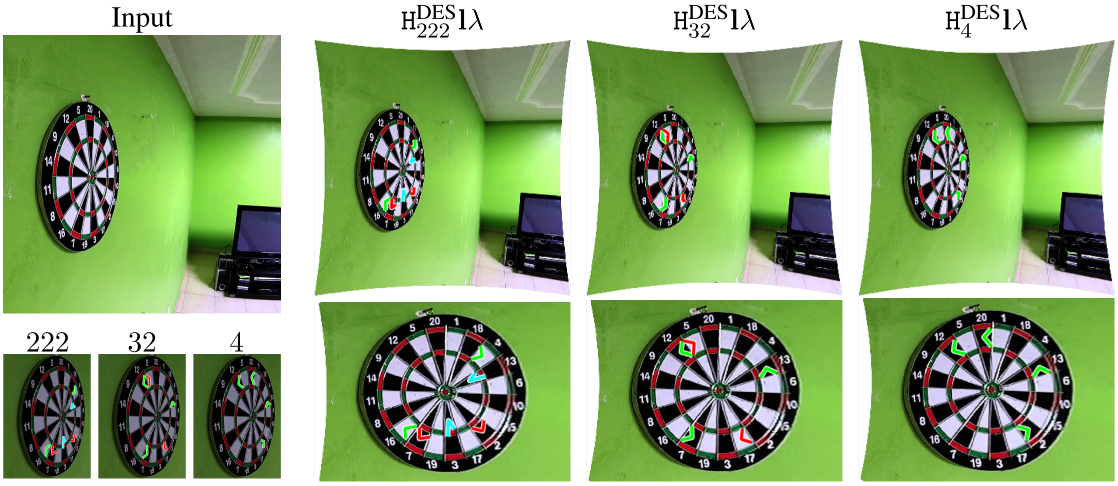

Input
$\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$
$\mathtt{H}_{32}^{\mathrm{DES}}\mathbf{l}\lambda$
$\mathtt{H}_{4}^{\mathrm{DES}}\mathbf{l}\lambda$
222
32
4

Figure 6.4: *Solver Variants.* (top-left image) The input to the method is a single image. (bottom-left triptych, contrast enhanced) The three configurations—$222, 32, 4$—of affine frames that are inputs to the proposed solvers variants. Corresponded frames have the same color. (top row, right) Undistorted outputs of the proposed solver variants. (bottom row, right) Cutouts of the dartboard rectified by the proposed solver variants. The affine frame configurations—$222, 32, 4$—are transformed to the undistorted and rectified images. The rectifications were estimated by the proposed directly-encoded scale (DES) solvers (see Section 6.3), but the input configurations are the same for the proposed change-of-scale (CS) solvers (see Section 6.4).

affine-rectified regions $\underline{\mathcal{R}}_i$ and $\underline{\mathcal{R}}_j$ are equal, namely $\underline{s}_i = \underline{s}_j$. Thus the unknown rectified scales of a corresponded set of n affine-covariant repeated regions $\underline{s}_1, \underline{s}_2, \dots, \underline{s}_n$ can be eliminated in pairs, which gives $n-1$ algebraically independent constraints and $\binom{n}{2}$ polynomial equations that are obtained by cross multiplying the denominators of the rational equations $\underline{s}_i = \underline{s}_j$. After eliminating the rectified scales, 3 unknowns remain, $\mathbf{l} = \left(l_1, l_2, 1\right)^\top$ and λ, so 3 constraints are needed.

6.3.3 Solver Variants

There are 3 minimal configurations for which we derive 3 solver variants: (i) 3 affine-covariant region correspondences, which we denote as the 222-configuration; (ii) 1 corresponded set of 3 affine-covariant regions and 1 affine-covariant region correspondence, denoted the 32-configuration; (iii) and 1 corresponded set of 4 affine-covariant regions, denoted the 4-configuration.

The notational convention introduced for the input configurations — $(222, 32, 4)$ — is extended to the change-of-scale solvers introduced in Section 6.4 and the bench of state-of-the-art solvers evaluated in the experiments (see Section 6.6) to make comparisons between the inputs of all the solvers easier. See Figure 6.4 for examples of all input configurations and results from each corresponding solver variant, and see Table 6.2 for a summary of all the tested solvers.

	Reference	Rectifies	Undistorts	Motion	# Regions	# Sols.	Size	Linearized
$\mathtt{H}_{2}\mathtt{l}$	Chapter 5	✓	✓	trans.	2	2	24x26	
$\mathtt{H}_{22}\mathtt{l}$	Chapter 5	✓	✓	trans.	4	4	76x80	
$\mathtt{H}_{22}\lambda$	[26]		✓	rigid[1]	4	18	18x18	
$\mathtt{H}_{22}^{\mathrm{DES}}\mathtt{l}$		✓		rigid	4	9	12x21	
$\mathtt{H}_{222}^{\mathrm{DES}}\mathtt{l}\lambda$		✓	✓	rigid	6	54	133x187	
$\mathtt{H}_{32}^{\mathrm{DES}}\mathtt{l}\lambda$		✓	✓	rigid	5	45	154x199	
$\mathtt{H}_{4}^{\mathrm{DES}}\mathtt{l}\lambda$		✓	✓	rigid	4	36	115x151	
$\mathtt{H}_{22}^{\mathrm{CS}}\mathtt{l}$	[14]	✓		rigid	4	1	4x4	✓
$\mathtt{H}_{222}^{\mathrm{CS}}\mathtt{l}\lambda$		✓	✓	rigid	6	54	133x187	✓
$\mathtt{H}_{32}^{\mathrm{CS}}\mathtt{l}\lambda$		✓	✓	rigid	5	45	154x199	✓
$\mathtt{H}_{4}^{\mathrm{CS}}\mathtt{l}\lambda$		✓	✓	rigid	4	36	115x151	✓

[1] The preimages of both region correspondences must be related by the same rigid transform in the scene plane.

Table 6.2: *State of the Art vs. Proposed Solvers (shaded in grey).* The proposed solvers return more solutions, but typically only 1 solution is feasible (see Figure 6.9). Note that the directly-encoded-scale (DES) solvers (shaded in light grey, see Section 6.3) have the same template size as the change-of-scale (CS) solvers (shaded in dark grey, see Section 6.4), despite being generated from different constraints. The $\mathtt{H}_{22}^{\mathrm{CS}}\mathtt{l}$ solver of [14] is part of the change-of-scale group of solvers but assumes a pinhole camera model.

The system of equations is of degree 4 regardless of the input configuration and has the form

$$\alpha_{j,1}\alpha_{j,2}\alpha_{j,3}\sum_{k=1}^{3}(-1)^k M_{3,k}^{(i)}\alpha_{i,k} = \alpha_{i,1}\alpha_{i,2}\alpha_{i,3}\sum_{k=1}^{3}(-1)^k M_{3,k}^{(j)}\alpha_{j,k}, \tag{6.3}$$

where $M_{3,k}^{(i)}$ is the $(3,k)$-minor of the rectified point-parameterization matrix $\left[\alpha_{i,1}\underline{\mathbf{x}}_{i,1} \quad \alpha_{i,2}\underline{\mathbf{x}}_{i,2} \quad \alpha_{i,3}\underline{\mathbf{x}}_{i,3}\right]$ defined by (6.1).

Note that the minors $M_{3,\cdot}^{(i)}$ are constant coefficients (see (6.2)). The 222-configuration results in a system of 3 polynomial equations of degree 4 in three unknowns l_1, l_2 and λ; the 32-configuration results in 4 equations of degree 4, and the 4-configuration gives 6 equations of degree 4. Only 3 constraints are needed, but we found that for the 32- and 4- configurations that all $\binom{n}{2}$ equations must be used to avoid spurious solutions that are introduced when the rectified scales are eliminated and the original rational equations $\underline{s}_i = \underline{s}_j$ are multiplied with their denominators. For example, if only the polynomial equations coming from the constraints $\underline{s}_1 = \underline{s}_2$, $\underline{s}_1 = \underline{s}_3$, $\underline{s}_1 = \underline{s}_4$ are used for the 4-configuration

$$\alpha_{i,1}\alpha_{i,2}\alpha_{i,3}\sum_{k=1}^{3}(-1)^k M_{3,k}^{(j)}\alpha_{1,k} = \alpha_{1,1}\alpha_{1,2}\alpha_{1,3}\sum_{k=1}^{3}(-1)^k M_{3,k}^{(i)}\alpha_{i,k} \quad i = 2,3,4, \tag{6.4}$$

then λ can be chosen such that $\sum_{k=1}^{3}(-1)^k M_{3,k}^{(i)}\alpha_{1,k} = 0$, and the remaining unknowns l_1 or

l_2 chosen such that $\alpha_{1,1}\alpha_{1,2}\alpha_{1,3} = 0$, which gives a 1-dimensional family of solutions. Thus, adding two additional equations removes all spurious solutions. Furthermore, including all equations simplified the elimination template construction.

In principle, a solver for the 222-configuration can be applied to the 32- and 4-configurations by duplicating the corresponding points in the input. Depending on how the points are duplicated, different results are obtained. In practice we observed that if, as above, we select the input such that $\underline{s}_1 = \underline{s}_2$, $\underline{s}_1 = \underline{s}_3$, $\underline{s}_1 = \underline{s}_4$, the solver breaks down. This is expected since the ideal is no longer zero-dimensional. However, other input configurations, e.g. $\underline{s}_1 = \underline{s}_2$, $\underline{s}_2 = \underline{s}_3$, $\underline{s}_3 = \underline{s}_4$, allow us to recover the same solutions as the 4-configuration solver in addition to a set of spurious solutions corresponding to some $\sum_{k=1}^{3}(-1)^k M_{3,k}^{(i)}\alpha_{i,k}$ vanishing.

6.3.4 Creating the Solvers

We used the automatic generator from Larsson et al. [48] to make the polynomial solvers for the three input configurations: $222, 32$, and 4. The directly-encoded-scale solver corresponding to each input configuration is denoted $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$, $\mathtt{H}_{32}^{\mathrm{DES}}\mathbf{l}\lambda$, and $\mathtt{H}_{4}^{\mathrm{DES}}\mathbf{l}\lambda$, respectively. The resulting elimination templates were of sizes 101×155 (54 solutions), 107×152 (45 solutions), and 115×151 (36 solutions). The equations have coefficients of very different magnitude. *E.g.*, the center-subtracted image coordinates have magnitude $\tilde{x}_i, \tilde{y}_i \approx 10^3$, and thus the distance to the image center $\tilde{x}_i^2 + \tilde{y}_i^2$ is $\approx 10^6$. To improve numerical conditioning, we re-scaled both the image coordinates and the squared distances by their average magnitudes. Note that this corresponds to a simple re-scaling of the variables in (λ, l_1, l_2), which is inverted once the solutions are obtained.

Experiments on synthetic data (see Section 6.6.1) revealed that using the standard GRevLex bases in the generator of [48] gave solvers with poor numerical stability. To generate stable solvers, we used the basis sampling technique proposed by Larsson et al. [54]. In [54] the authors propose a method for randomly sampling feasible monomial bases, which can be used to construct polynomial solvers. We generated (with [48]) 1,000 solvers with different monomial bases for each of the three variants using the heuristic from [54]. Following the method from Kuang et al. [41], the sampled solvers were evaluated on a test set of 1,000 synthetic instances, and the solvers with the smallest median equation residual were kept. The resulting solvers have slightly larger elimination templates (133×187, 154×199, and 115×151); however, they are significantly more stable. See Section 6.6.1 for a comparison between the solvers using the sampled bases and the standard GRevLex bases (default in [48]).

6.3.5 The Fixed Lens Distortion Variant

Finally, we consider the case of known division-model parameter λ. Fixing λ in (6.3) yields degree 3 constraints in only 2 unknowns l_1 and l_2. Thus only 2 correspondences of 2 repeated affine-covariant regions are needed. The generator of [48] found a stable solver (denoted $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}$) with an elimination template of size 12×21, which has 9 solutions. Basis sampling was not required in this case. There is a second minimal problem for 3 repeated affine-covariant regions; however, unlike the case of unknown distortion, this minimal problem is equivalent to the $\mathtt{H}_{22}$

variant. It also has 9 solutions and can be solved with the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver by duplicating a region in the input. The proposed $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}$ solver contrasts to the solvers from [71, 20, 14] in that it is generated from constraints directly induced by the rectifying homography rather than its linearization.

6.3.6 Degeneracies

We observed three important degeneracies for the DES solvers. First, if the vanishing line passes through the image origin, *i.e.* $\mathbf{l} = \left(l_1, l_2, 0\right)^\top$, then the radial term in the homogeneous coordinate of (2.36) is canceled. In this case, it is not possible to recover the radial distortion using the equations in (6.3). However, the degeneracy does not arise from the problem formulation. An affine transform can be applied to the undistorted image such that the vanishing line $\mathbf{l}$ in the affine-transformed space has $l_3 \neq 0$. As future work, we will investigate how to remove this degeneracy from the solvers.

Secondly, the problem degenerates if the scene plane is already fronto-parallel to the camera and the corresponding points from the affine-covariant regions fall on circles centered at the image center. Since the corresponding points have the same radii, they will undergo the same scaling due to radial distortion (see (2.35)). In this case, the radial distortion parameter again becomes unobservable since it is impossible to disambiguate the scale of the features from the scaling of the lens distortion.

Third, suppose that (i) $\mathtt{H}$ is a rectifying homography other than the identity matrix, (ii) that the image has no radial distortion, (iii) and that all corresponding points from repeated affine-covariant regions fall on a single circle centered at the image center. As in the second case, applying the division model (see Section 2.11) uniformly scales the points about the image center. Given $\lambda \neq 0$, for a transformation by $f(\cdot, \lambda)$ defined in (2.35) of the points lying on the circle there is a scaling matrix $\mathtt{S}(\lambda) = \mathrm{diag}(1/\lambda, 1/\lambda, 1)$ that maps the points back to their original positions. Thus there is a 1D family of rectifying homographies given by $\mathtt{HS}(\lambda)$ for the corresponding set of undistorted images given by $f(\cdot, \lambda)$.

6.3.7 Reflections

In the derivation of (6.3), the rectified scales $\underline{s}_i$ were eliminated with the assumption that they had equal signs (see Sec. 6.3.4). Reflections will have oppositely signed rectified scales; however, reversing the orientation of left-handed affine frames in a simple pre-processing step that admits the use of reflections. Suppose that $\det\left(\left[\tilde{\mathbf{x}}_{i,1}\, \tilde{\mathbf{x}}_{i,2}\, \tilde{\mathbf{x}}_{i,3}\right]\right) < 0$, where $(\tilde{\mathbf{x}}_{i,1}, \tilde{\mathbf{x}}_{i,2}, \tilde{\mathbf{x}}_{i,3})$ is a distorted left-handed point parameterization of an affine-covariant region. Then reordering the point parameterization as $(\tilde{\mathbf{x}}_{i,3}, \tilde{\mathbf{x}}_{i,2}, \tilde{\mathbf{x}}_{i,1})$ results in a right-handed point-parameterization such that $\det\left(\left[\tilde{\mathbf{x}}_{i,3}\, \tilde{\mathbf{x}}_{i,2}\, \tilde{\mathbf{x}}_{i,1}\right]\right) > 0$, and the scales of corresponded rectified reflections will be equal.

6.4 The Change-of-Scale (CS) Solvers

The proposed change-of-scale (CS) solvers use the Jacobian determinant of the rectifying transformation to induce local constraints on the imaged vanishing line and the unknown parameter

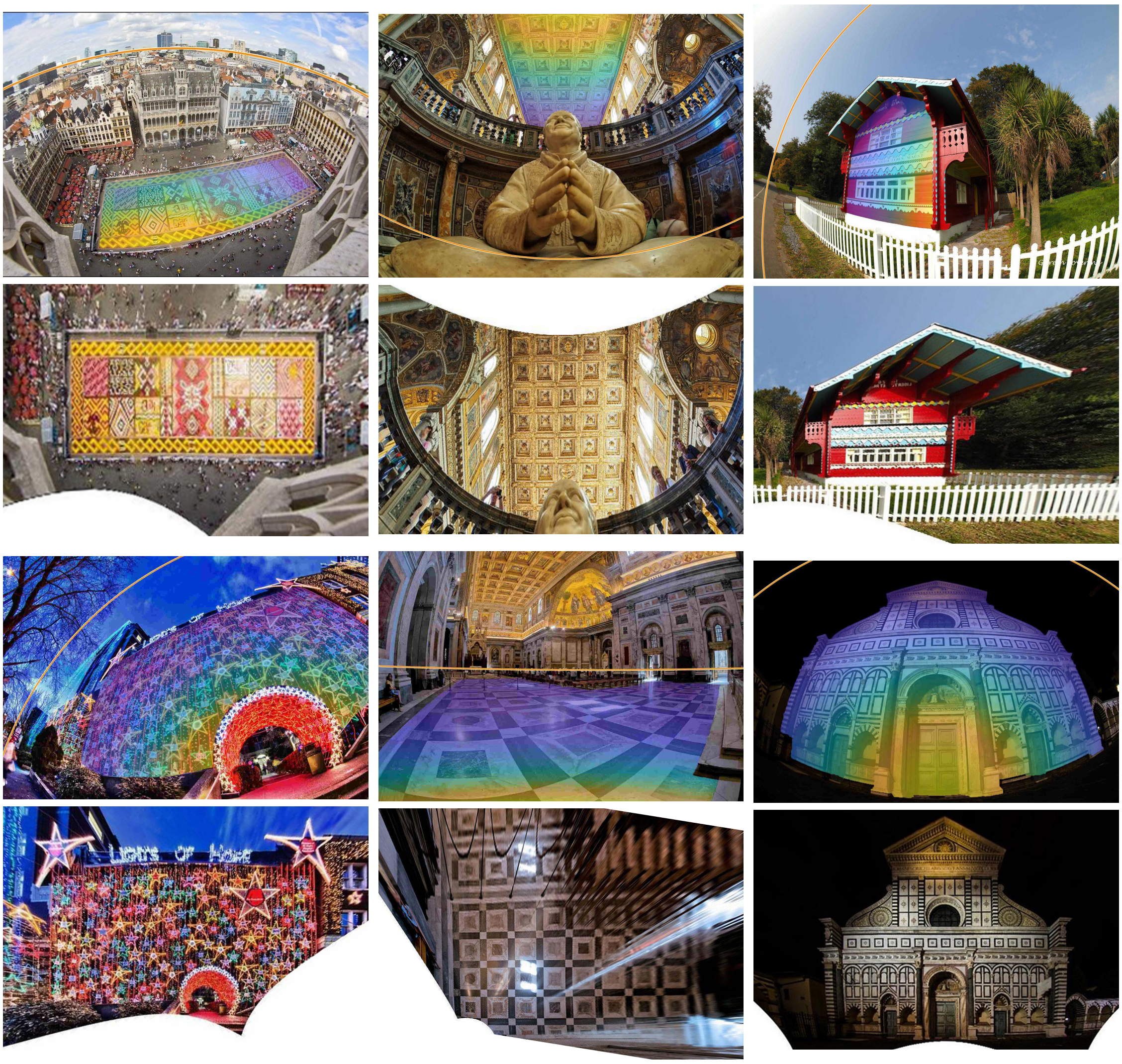


Figure 6.5: *Change-of-Scale Solver Results.* The input images are on the first and third rows and show the distorted image of the vanishing line in orange and the dense change of scale (see Section 6.4.5) in the parula color map that is alpha blended on the scene plane. Purple corresponds to the smallest relative scale change due to the imaging of the scene plane and yellow to the largest with respect to a chosen reference point on the plane. The second and fourth rows contain the rectified results from the $\mathtt{H}_{222}^{\mathrm{CS}}\mathbf{l}\lambda$ change-of-scale solver (see Section 6.4).

for the division model of radial lens distortion (see Section 2.11). In particular, the derivation uses the fact that the unknown division model parameter is encoded exclusively in the third coordinate (see (2.35)), which results in a formulation that is tractable for automatic solver generators.

In fact, there are several related works that linearize the homography and impose constraints on the Jacobian determinant [6, 14, 71, 38, 39]; however, the proposed CS solvers are the first solvers to incorporate lens distortion with this approach. The Jacobian determinant gives the change of scale of a function at a point, which motivates the name Change of Scale (CS) for the solvers proposed in this section. It is a surprising discovery that the combined effects of severe lens distortion and perspective imaging from oblique views can be linearized over regions with scales that are typical for covariant region detections (see Figure 6.5), which measure the relative scale change between coplanar repeats due to imaging. In fact, the change-of-scale solvers are used to rectify near fisheye distortions effectively (see Figure 6.5).

The CS solvers have the advantage over the DES solvers in that they admit strictly scale-covariant regions detections, whereas the DES solvers require affine-covariant region detections. As with the DES solvers in Section 6.3.4, the solvers restore the affine invariant that coplanar repeated regions have the same scale.

6.4.1 The Change-of-Scale Formulation

The Euclidean coordinates $(\underline{x}_i, \underline{y}_i)^\top$ of the rectified point $\underline{\mathbf{x}}_i = \alpha_i \left(\underline{x}_i, \underline{y}_i, 1\right)^\top = \mathtt{H} f(\tilde{\mathbf{x}}_i, \lambda)$ (refer to (2.36)), of any imaged point $\tilde{\mathbf{x}}_i = (\tilde{x}_i, \tilde{y}_i, 1)^\top$ on the scene plane is given by the vector-valued nonlinear function

$$\underline{\mathbf{x}}(\tilde{x}, \tilde{y}) = \left(\underline{x}(\tilde{x}, \tilde{y}), \underline{y}(\tilde{x}, \tilde{y})\right)^\top = \left(\tfrac{\tilde{x}}{\mathbf{l}^\top f(\tilde{\mathbf{x}}, \lambda)}, \tfrac{\tilde{y}}{\mathbf{l}^\top f(\tilde{\mathbf{x}}, \lambda)}\right)^\top .$$

The function $\underline{\mathbf{x}}$, which returns the inhomogeneous coordinates of the undistorted and rectified point $\left(\underline{x}, \underline{y}\right)$, can be linearized at $(\tilde{x}, \tilde{y})$ with the first-order Taylor expansion,

$$\underline{\mathbf{x}}(\tilde{x} + \delta_{\tilde{x}}, \tilde{y} + \delta_{\tilde{y}}) = \underline{\mathbf{x}}(\tilde{x}, \tilde{y}) + \mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l}, \lambda)|_{(\tilde{x}, \tilde{y})} \cdot \left(\delta_{\tilde{x}}, \delta_{\tilde{y}}\right)^\top .$$

The Jacobian determinant $\det\left(\mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l}, \lambda)|_{(\tilde{x}_i, \tilde{y}_i)}\right)$ gives the approximate change of scale of the rectifying and undistorting function $\underline{\mathbf{x}}$ near the point $(\tilde{x}, \tilde{y})^\top$. Let $\tilde{s}_i$ be the scale of an image region $\tilde{\mathcal{R}}_i$ with its centroid at $\left(\tilde{x}_i, \tilde{y}_i\right)^\top$, where the preimage $\underline{\mathcal{R}}_i$ of $\tilde{\mathcal{R}}_i$ is on some scene plane Π. Let $\underline{s}_i$ be the rectified scale of $\underline{\mathcal{R}}_i$. Then the unknown rectified scale $\underline{s}_i$ can be expressed in terms of the distorted scale $\tilde{s}_i$ and the Jacobian determinant as

$$\underline{s}_i = \tilde{s}_i \cdot \det\left(\mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l}, \lambda)|_{(\tilde{x}_i, \tilde{y}_i)}\right) = \frac{-\tilde{s}_i(\lambda(\tilde{x}_i^2 + \tilde{y}_i^2) - 1)}{(\lambda(\tilde{x}_i^2 + \tilde{y}_i^2) + l_1\tilde{x}_i + l_2\tilde{y}_i + 1)^3}. \tag{6.5}$$

6.4.2 Eliminating the Rectified Scale

The equation for the rectified scale given in (6.5) defines the unknown geometric quantities: (i) division-model parameter λ, (ii) scene-plane vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$, (iii) and the rectified scale $\underline{s}_i$ of the rectified image region $\underline{\mathcal{R}}_i$. The distorted scale $\tilde{s}_i$ of imaged region $\tilde{\mathcal{R}}_i$ is measured by some scale-covariant region detector, *e.g.*, the SIFT or Hessian Affine detector [60, 65]. Let $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ be detected repeated coplanar regions. Then the scales of their rectified preimages $\underline{\mathcal{R}}_i$ and $\underline{\mathcal{R}}_j$ are equal, namely $\underline{s}_i = \underline{s}_j$. A unique solution is defined by restricting the vanishing line to the affine subspace $l_3 = 1$, which results in degree 4 constraints. The alternative of fixing the rectified scale $\underline{s}_i$ is rejected since it results in higher degree constraints. Thus, the unknown rectified scales of a group of n co-planar repeats $\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_n$ can be eliminated in pairs (see (6.6)), which gives $n-1$ algebraically independent constraints and $\binom{n}{2}$ polynomial equations that are obtained by cross multiplying the denominators of the rational equations $\underline{s}_i = \underline{s}_j$.

6.4.3 Creating the solver

After eliminating the rectified scales 3 unknowns remain, namely $\mathbf{l} = \left(l_1, l_2, 1\right)^\top$ and λ, so 3 equations are needed. The minimal configurations are the same as the DES solvers and an analogous naming scheme is adopted for the CS solvers. The CS solvers can be obtained from 3 correspondences of 2 coplanar repeats, denoted $\mathtt{H}_{222}^{\mathrm{CS}}\mathbf{l}\lambda$, 1 corresponded set of 3 and 1 correspondence of 2 coplanar repeats, denoted $\mathtt{H}_{32}^{\mathrm{CS}}\mathbf{l}\lambda$, or 1 corresponded set of 4 coplanar repeats, denoted $\mathtt{H}_{4}^{\mathrm{CS}}\mathbf{l}\lambda$ (see the comparison in Table 6.2). The system of equations contains rational expressions of the form

$$\tilde{s}_i \cdot \det\left(\mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l}, \lambda)|_{(\tilde{x}_i, \tilde{y}_i)}\right) = \tilde{s}_j \cdot \det\left(\mathtt{J}_{\underline{\mathbf{x}}}(\mathbf{l}, \lambda)|_{(\tilde{x}_j, \tilde{y}_j)}\right). \tag{6.6}$$

After multiplying equations (6.6) by common denominators we obtain a system of three quartic polynomial equations in three unknowns, namely l_1, l_2 and λ. Again we used the automatic generator from Larsson et al. [48] to create the polynomial solvers for all of the minimal configurations. The structure of the change-of-scale solvers turned out to be similar to the DES solvers (*i.e.*, same monomials and number of solutions, but the coefficients in equations are computed differently).

6.4.4 Degeneracies

The change-of-scale solvers suffer from the same degeneracies that are listed in Section 6.3.6 for the DES solvers. There are likely different degeneracies between the two families of solvers, but an exhaustive analysis is difficult.

6.4.5 Dense Change of Scale Due to Imaging

Up to a global scale ambiguity, the rectified scale $\underline{s}$ of an imaged scene plane region can be approximated with (6.5). The projective and radial lens distortion components of the imaging transformation are linearized in (6.5), so the approximation of the rectified scale $\underline{s}$ is more accurate for smaller regions.

The combined change-of-scale effects of lens distortion and perspective warping due to the imaging of a scene plane can be seen in Figure 6.5. The reference point is the image of the centroid of the convex hull of rectified coplanar covariant regions. The dense relative change of scale is rendered by the alpha-blended parula colormap in the original images of Figure 6.5. Regions with larger scale change due to imaging are orange; regions close to the scale change of the imaged reference point are blue, and regions with vanishing relative scale change are purple.

The purple regions will be expanded in the rectified image and the yellow regions shrunk such that the affine rectification restores the affine invariant that coplanar regions whose preimages are of equal scale are the same scale in the rectified image.

For pinhole cameras, regions undergoing an equal change of scale from imaging are projected to isolines [20]. However, as seen in Figure 6.5, for radially-distorted cameras parameterized by the division model (see Section 2.11), regions undergoing equal change of scale from imaging are constrained to circles. This is consistent with the fact that scene lines are imaged as circles under the division model of radial lens distortion [10, 26, 86, 92]. The distorted image of the vanishing line as a circle under the division model is shown in a synthetic scene of Figure 6.2 and in real images in Figures 6.5 and 6.14 (the orange circular segments).

The dense relative change of scale is useful for automatic rectification. *E.g.*, in images where the image of the vanishing line intersects the image extents, regions approaching the vanishing line rectify to arbitrarily large scales. Thus a bound on the rectified scale is needed to prevent the rectified image from blowing up. Using (6.5), an image can be masked such that any masked point has a relative change of scale bounded by some user threshold, which can be used to generate reasonably sized rectifications. All images in this document were automatically generated with this method.

6.5 Robust Estimation

The solvers are used in a LO-RANSAC-based robust-estimation framework [15]. Affine rectifications and undistortions are jointly hypothesized by one of the proposed solvers. A metric upgrade is attempted, and models with maximal consensus sets are locally optimized by an extension of the method introduced in [74]. The metric-rectifications are presented in the results.

6.5.1 Local Features and Descriptors

We use the Maximally-Stable Extremal Region and Hessian-Affine detectors as detailed in Sections 3.2.4 and 3.2.6 [62, 65]. The affine-covariant regions are given by an affine transform (see Section 6.3.1), equivalently 3 distinct points, which defines an affine frame in the image space [69]. The image patch local to the affine frame is embedded into a descriptor vector by the RootSIFT transform [4, 60] (see Section 3.3.1).

6.5.2 Appearance Clustering and Sampling

Affine frames are tentatively labeled as repeated texture by their appearance. The RootSIFT descriptors are agglomeratively clustered, and the pair-wise tentative correspondences are established among connected components as detailed in Section 3.3.2.

Sample configurations for the proposed minimal solvers are illustrated in Figure 6.4 and detailed in Section 6.3.3. To recap, the solver variants for the proposed undistorting and rectifying minimal solvers—either from the DES or CS family—are 3 correspondences of 2 covariant regions (the 222-solvers), a corresponded set of 3 covariant regions and a correspondence of 2 covariant regions (the 32-solvers), and a corresponded set of 4 covariant regions (the 4-solvers).

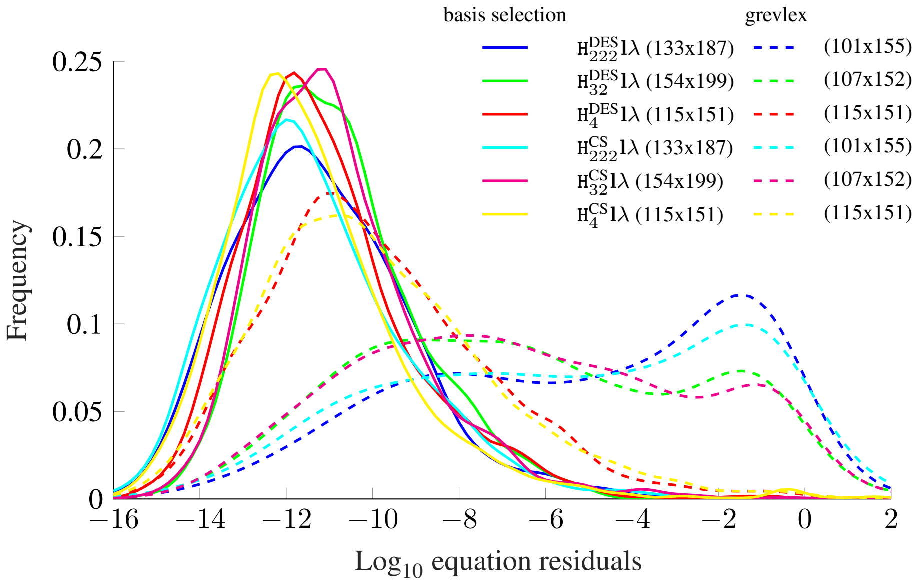


Figure 6.6: *Stability Study for Noiseless Features.* The equation residuals (deviation from 0) for the particular polynomial system of equations solved by each of the DES and CS solvers is used to measure solver stability (see Sections 6.3.2 and 6.4.2, respectively) . The minimal solution closest to the ground truth is evaluated and reported for 1000 noiseless synthetic scenes. The basis selection method of [54] is essential for stable solver generation.

For each RANSAC trial, appearance clusters are selected with the probability given by its relative size to the other appearance clusters, and the required number of correspondences or corresponded sets are drawn from the selected clusters.

6.5.3 Metric Upgrade and Local Optimization

The affine-covariant regions that are members of the minimal sample are affine rectified by each feasible model returned by the solver; typically there is only 1 (see Figure 6.9). A metric upgrade is estimated from the affine-rectified minimal sample set using the linear solver introduced in [74]. Then all affine-covariant regions are metrically-upgraded using the estimate. The consensus set is measured in the metric-rectified space by verifying the congruence of the basis vectors of the corresponded affine frames. Congruence is an invariant of metric-rectified space and is a stronger constraint than the equal-scale invariant of affine-rectified space that was used to derive the proposed solvers. The metric upgrade essentially comes for free by inputting the affine-covariant regions sampled for the proposed solvers to the linear metric-upgrade solver proposed in [74]. By using the metric-upgrade, the verification step of RANSAC can enforce the congruence of corresponding affine-covariant region extents (equivalently, the lengths of the linear basis vectors) to estimate an accurate consensus set. Models with the maximal consensus

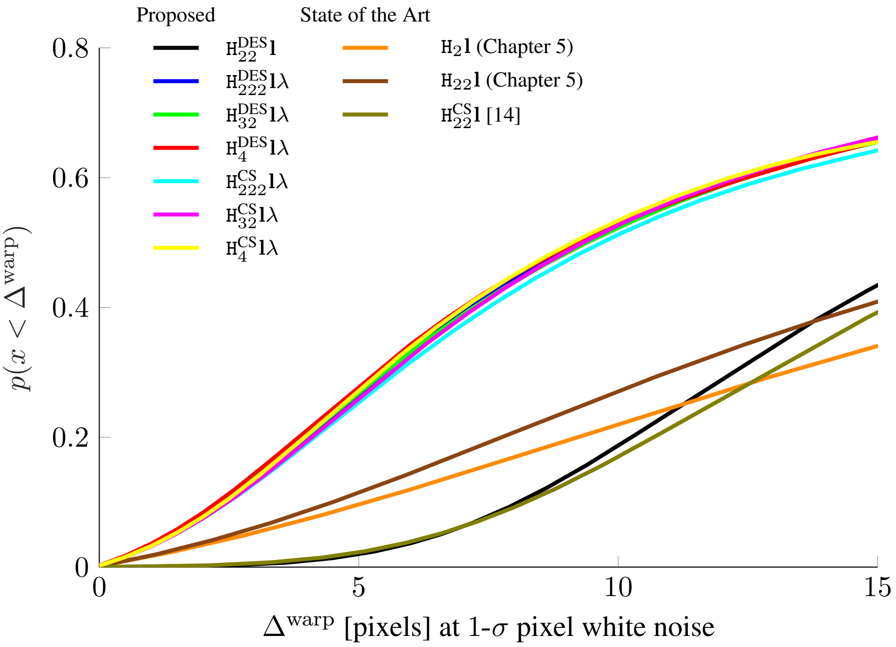


Figure 6.7: *Warp Errors for Fixed 1-σ Pixel Noise.* Reports the cumulative distributions of raw warp errors Δ^{warp} (see Section 2.12) for the bench of solvers on 1000 synthetic scenes with 1-σ pixel of imaging white noise added. The proposed solvers (with undistortion estimation) give significantly better proposals than the state of the art.

set are locally optimized in a method similar to [74].

6.6 Experiments

The stabilities and noise sensitivities of the proposed solvers are evaluated on synthetic data. We compare the proposed solvers to a bench of 4 state-of-the-art solvers (see Table 6.2). We apply the denotations for the solvers introduced in Section 6.3.3 to all the solvers in the benchmark; *e.g.*, a solver requiring 2 correspondences of 2 affine-covariant regions will be prefixed by $\mathtt{H}_{22}$, while the proposed solver requiring 1 corresponded set of 4 affine-covariant regions is prefixed by $\mathtt{H}_{4}$.

Included are two state-of-the-art single-view solvers for radially-distorted conjugate translations, denoted $\mathtt{H}_{2}\mathbf{l}$ and $\mathtt{H}_{22}\mathbf{l}$(see Chapter 5); a full-homography and radial distortion solver, denoted $\mathtt{H}_{22}\lambda$ [26]; and the change-of-scale solver for affine rectification of [14], denoted $\mathtt{H}_{22}^{\mathrm{CS}}\mathbf{l}$.

The sensitivity benchmarks measure the performance of rectification accuracy by the warp error (see Section 2.12) and the relative error of the division parameter estimate. Stability is measured by the equation residuals of the solution that is closest to ground truth. The $\mathtt{H}_{22}\lambda$ solver is omitted from the warp error since the vanishing line is not estimated, and the $\mathtt{H}_{22}^{\mathrm{CS}}\mathbf{l}$ and $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}$ solvers are omitted from benchmarks involving lens distortion since the solvers assume a pinhole camera.

6.6.1 Synthetic Data

The performance of the proposed solvers on 1000 synthetic images of 3D scenes with known ground-truth parameters is evaluated. A camera with a random but realistic focal length is randomly placed with respect to a scene plane such that it is mostly in the camera's field-of-view. The image resolution is set to 1000x1000 pixels. The noise sensitivity of the solvers are evaluated both on conjugately-translated and rigidly-transformed coplanar repeats (see Figure 6.8). Scenes with conjugately-translated coplanar repeats are evaluated so that the proposed solvers can be compared to state-of-the-art solvers proposed in Chapter 5. For either motion type, affine frames are generated on the scene plane such that their scale with respect to the scene plane is realistic. The modeling choice reflects the use of affine-covariant region detectors on real images (see Section 6.3.1).

The image is distorted according to the division model. For the sensitivity experiments, isotropic white noise is added to the distorted affine frames at increasing levels. Performance is characterized by the relative error of the estimated distortion parameter and by the warp error, which measures the accuracy of the affine-rectification.

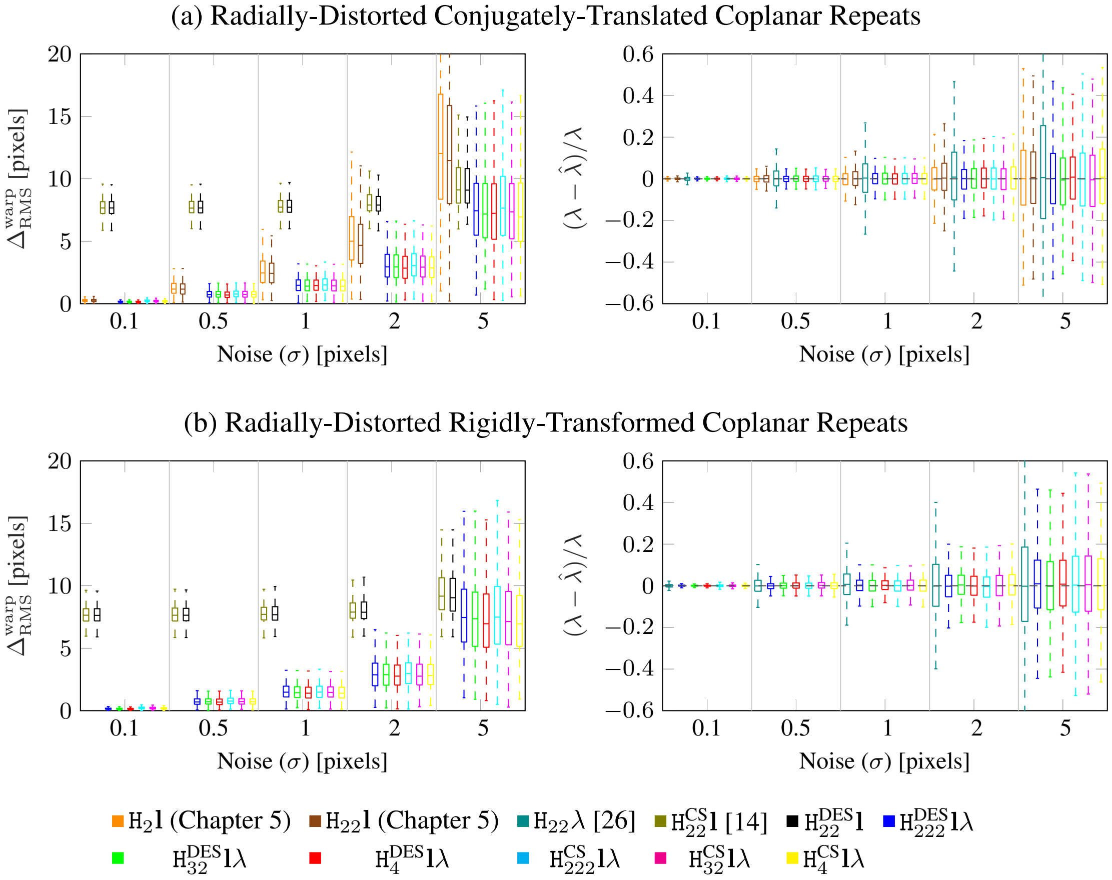
Numerical Stability

The stability study of Figure 6.6 compares compares the solver variants generated using the standard GRevLex bases versus solvers generated using the basis selection method of [54] (also see Section 6.3.4). The generator of Larsson et al. [48] was used to generate both sets of solvers. Stability is measured as the equation residual (equivalently, deviation from 0) of the polynomial system of equations associated with each solver (see Sections 6.3.2 and 6.4.2) for the solution that is closest to ground truth for noiseless affine-frame correspondences across realistic synthetic scenes, which are generated as described in the introduction of Section 6.6.1.

The normalized ground-truth parameter of the division model λ is set to -4, a value typical for wide field-of-view cameras like the GoPro, where the image is normalized by $1/(\text{width} + \text{height})$. Figure 6.6 reports the histogram of $\log_{10}$ equation residuals and shows that the basis selection method of [54] significantly improves the stability of the generated solvers. The basis-sampled solvers are used for the remainder of the experiments.

Noise Sensitivity

The proposed and state-of-the-art solvers are tested with increasing levels of white noise added to the point parameterizations (see Section 6.3.1) of the affine-covariant region correspondences that are either translated or rigidly-transformed on the scene plane (see Figure 6.8). The amount of white noise is given by the standard deviation of a zero-mean isotropic Gaussian distribution, and the solvers are tested at noise levels of $\sigma \in \{0.1, 0.5, 1, 2, 5\}$. The ground-truth normalized division model parameter is set to $\lambda = -4$, which is typical for GoPro-type imagery in normalized image coordinates.

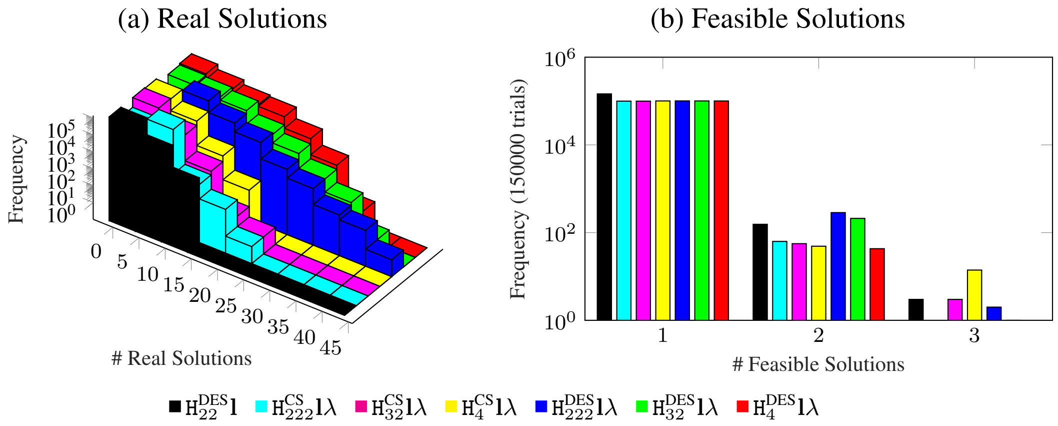

(a) Radially-Distorted Conjugately-Translated Coplanar Repeats

(b) Radially-Distorted Rigidly-Transformed Coplanar Repeats

Figure 6.8: *Sensitivity Benchmark.* Comparison of two error measures after 25 iterations of a simple RANSAC for different solvers with increasing levels of white noise added to the affine frame correspondences, where the normalized division model parameter is set to -4 (see Section 2.11), which is similar to the distortion of a GoPro Hero 4. (top row) Shows results for translated coplanar repeats, and (bottom row) shows results for rigidly-transformed coplanar repeats. (left column) Reports the root mean square warp error $\Delta^{\text{warp}}_{\text{RMS}}$, and (right column) reports the relative error of the estimated division model parameter. The proposed solvers are significantly more robust for both types of repeats on both error measures.

The cumulative distributions of warp errors in Figure 6.7 show that for 1-pixel white noise on conjugately-translated affine frames, the proposed solvers—$\mathtt{H}^{\text{DES}}_{222}\mathbf{l}\lambda$, $\mathtt{H}^{\text{DES}}_{32}\mathbf{l}\lambda$, $\mathtt{H}^{\text{DES}}_{4}\mathbf{l}\lambda$,$\mathtt{H}^{\text{CS}}_{222}\mathbf{l}\lambda$, $\mathtt{H}^{\text{CS}}_{32}\mathbf{l}\lambda$ and $\mathtt{H}^{\text{CS}}_{4}\mathbf{l}\lambda$—give significantly more accurate estimates than the state-of-the-art conjugate translation solvers proposed in Chapter 5. Interestingly, all of the proposed undistorting variants from both the DES and CS families of rectifying solvers have nearly identical performance.

If 5 pixel RMS warp error is fixed as a threshold for a good model proposal, then 30% of the models given by the proposed solvers are good versus roughly 10% for the solvers of Chapter 5.



(a) Real Solutions

(b) Feasible Solutions

Figure 6.9: *Real and Feasible Solutions.* 1. The histograms of the number of real solutions returned by the proposed solvers. 2. Typically, only 1 solution is feasible. Feasibility is determined by checking that the division model parameter falls in a reasonable interval. The frequencies were calculated on results from 150,000 trials on different scenes with varying levels of imaged white-noise.

The proposed $\mathtt{H}_{22}^{\mathrm{DES}}\mathtt{1}$ solver and the $\mathtt{H}_{22}^{\mathrm{CS}}\mathtt{1}$ of [14] each give biased proposals since they cannot estimate lens distortion.

The solvers are wrapped by a basic RANSAC estimator that minimizes the RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ over 25 minimal samples of affine frames for each of the conjugately-translated and rigidly-transformed coplanar repeat sensitivity studies in Figure 6.8. The RANSAC estimates are summarized in boxplots for 1000 synthetic scenes. The interquartile range is contained within the extents of a box, and the median is the horizontal line dividing the box. As shown in Figure 6.8, the proposed solvers — $\mathtt{H}_{222}^{\mathrm{DES}}\mathtt{1}\lambda$, $\mathtt{H}_{32}^{\mathrm{DES}}\mathtt{1}\lambda$, $\mathtt{H}_{4}^{\mathrm{DES}}\mathtt{1}\lambda$,$\mathtt{H}_{222}^{\mathrm{CS}}\mathtt{1}\lambda$, $\mathtt{H}_{32}^{\mathrm{CS}}\mathtt{1}\lambda$ and $\mathtt{H}_{4}^{\mathrm{CS}}\mathtt{1}\lambda$— again give the most accurate lens distortion and rectification estimates. In fact, the proposed solvers are superior to the state of the art at all noise levels. The proposed distortion-estimating solvers give solutions with less than 5-pixel RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ 75% of the time and estimate the correct division model parameter more than half the time at the 2-pixel noise level. The proposed fixed-lens distortion solver $\mathtt{H}_{22}^{\mathrm{DES}}\mathtt{1}$ and the $\mathtt{H}_{22}^{\mathrm{CS}}\mathtt{1}$ of [14] give biased solutions since they assume the pinhole camera model. The vanishing line is not directly estimated by the solver $\mathtt{H}_{22}\lambda$ of [26], so it is not reported.

### Feasible Solutions and Runtime

This study shows the number of real and feasible solutions given by the proposed solvers for 150000 trials across 1000 scenes at varying noise levels with a fixed normalized division model parameter of $\lambda = -4$. Figure 6.9 (left) shows the number of real solutions, and Figure 6.9 (right) shows the subset of feasible solutions as defined by the estimated normalized division-model parameter solution falling in the interval $[-8, 0.5]$. All solutions are considered feasible

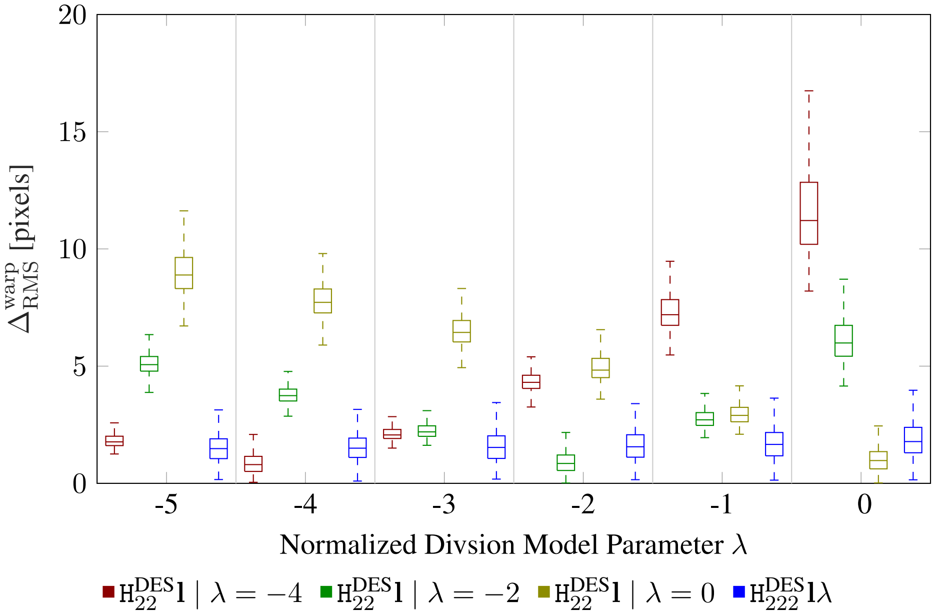


Figure 6.10: *Distortion Study.* Reports the root-mean-square warp error $\Delta_{\text{RMS}}^{\text{warp}}$ (see Section 2.12) for 1000 synthetic scenes imaged by cameras with varying normalized division model parameter with 1-σ pixel white noise. Solvers $\mathtt{H}_{22}^{\text{DES}}\mathtt{l} \mid \lambda = -4$, $\mathtt{H}_{22}^{\text{DES}}\mathtt{l} \mid \lambda = -2$, and $\mathtt{H}_{22}^{\text{DES}}\mathtt{l} \mid \lambda = 0$ rectify the pinhole image that is undistorted with the given fixed division model parameter. The $\mathtt{H}_{222}^{\text{DES}}\mathtt{l}\lambda$ solver is competitive even for the case where the fixed division model parameter matches ground truth and gives stable performance across all distortion levels.

for the $\mathtt{H}_{22}^{\text{DES}}\mathtt{l}$ solver. Figure 6.9 (right) shows that in 97% of the scenes only 1 solution is feasible, which means that nearly all incorrect solutions can be quickly discarded.

The runtimes of the DES family of solvers are reported. The MATLAB implementation of the solvers on a standard desktop are 2 ms for $\mathtt{H}_{222}^{\text{DES}}\mathtt{l}\lambda$, 2.2 ms for $\mathtt{H}_{32}^{\text{DES}}\mathtt{l}\lambda$, 1.7 ms for $\mathtt{H}_{4}^{\text{DES}}\mathtt{l}\lambda$, and 0.2 ms for $\mathtt{H}_{22}^{\text{DES}}\mathtt{l}$. Due to the similar structure in the equations, the CS solvers have comparable performance.

6.6.2 Distortion Study

The distortion study evaluates the accuracy of rectifications as measured by the warp error (see Section 2.12) over a normalized ground truth division model parameter from $\lambda \in \{-5, -4, -3, -2, -1, 0\}$, which are values that are characteristic of near-fisheye to pinhole lenses (see Section 2.11). The images have fixed 1px-σ white noise added. The methodology of scene generation is the same as detailed in Section 6.6.1.

Since the sensitivity experiments of Section 6.6.1 show that the performance of the proposed solvers is essentially the same with respect to noise, we choose $\mathtt{H}_{222}^{\text{DES}}\mathtt{l}\lambda$ as their representative. It is evaluated against 3 solvers—$\mathtt{H}_{22}^{\text{DES}}\mathtt{l} \mid \lambda = -4$, $\mathtt{H}_{22}^{\text{DES}}\mathtt{l} \mid \lambda = -2$, and $\mathtt{H}_{22}^{\text{DES}}\mathtt{l} \mid \lambda = 0$—

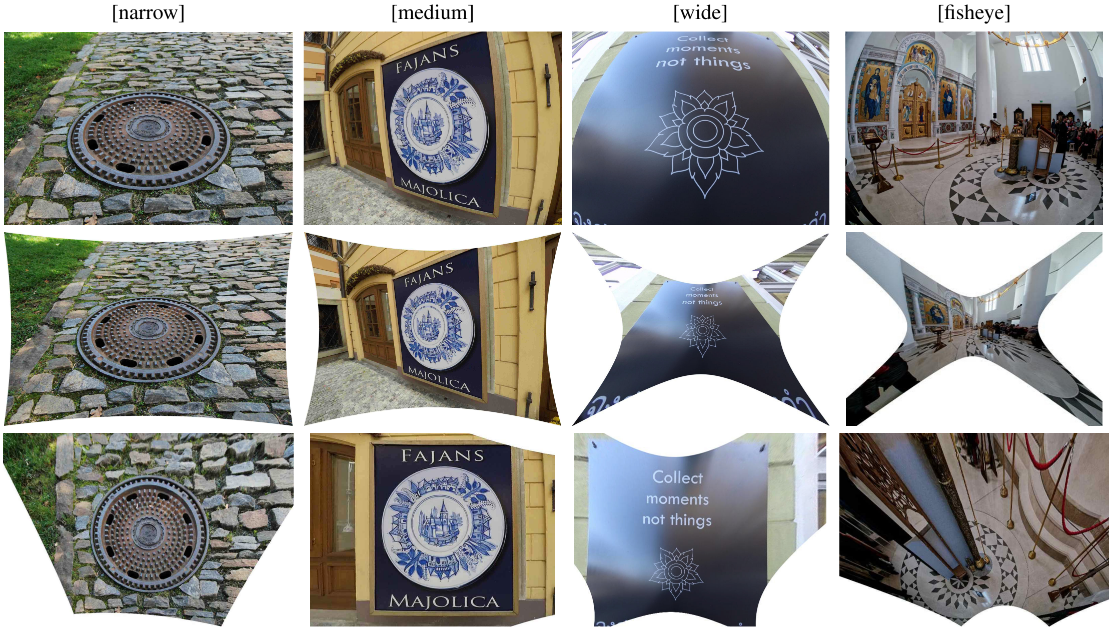


Figure 6.11: *Field-of-View Study.* The proposed solver $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{1}\lambda$ gives accurate rectifications across all fields-of-view: (left-to-right) Nikon D60, GoPro Hero 4 at the medium- and wide-FOV settings, and a Panasonic DMC-GM5 with a Samyang 7.5mm fish-eye lens. The outputs are the undistorted (middle row) and rectified images (bottom row).

each of which undistort at a different fixed normalized division model parameter, namely $\lambda \in \{-4, -2, 0\}$, respectively. The fixed distortion solvers estimate the affine rectification with the proposed $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{1}$ (see Section 6.3.5) using the undistorted minimal sample, which is computed with the given fixed division model parameter of the solver.

Figure 6.10 shows that even for the case where the fixed division model parameter of the solver is equivalent to the ground truth, the best solutions of the proposed $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{1}\lambda$ are equivalent to rectifying with known ground truth. Furthermore, the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{1}\lambda$ is stable, giving the same performance at a fixed noise level across all ground truth division model parameters. As expected, the warp error quickly increases for the $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{1} \mid \lambda = -4$, $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{1} \mid \lambda = -2$, and $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{1} \mid \lambda = 0$ solvers as the ground truth division model parameter differs from the fixed division model parameter.

6.6.3 Real Images

The field-of-view experiment of Figure 6.11 evaluates the proposed $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{1}\lambda$ solver on real images taken with narrow, medium, wide-angle, and fish-eye lenses. Images with diverse scene content were chosen. Figure 6.11 shows that the $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{1}\lambda$ gives accurate rectifications for all lens types. Additional results for wide-angle and fisheye lenses are included in Figure 6.14 in the end of this chapter.

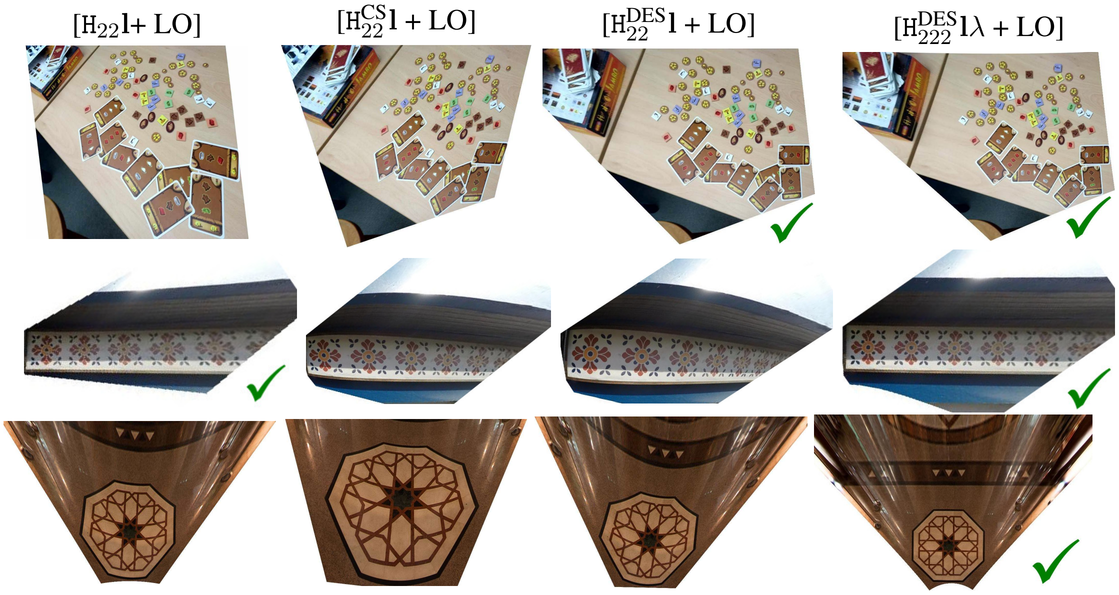

$[\mathtt{H}_{22}\mathbf{l}+\text{LO}]$
$[\mathtt{H}_{22}^{\mathrm{CS}}\mathbf{l}+\text{LO}]$
$[\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}+\text{LO}]$
$[\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda+\text{LO}]$
✓
✓
✓
✓
✓

Figure 6.12: *Solver Comparison.* The state-of-the art solvers $\mathtt{H}_{22}\mathbf{l}$(see Chapter 5) and $\mathtt{H}_{22}^{\mathrm{CS}}\mathbf{l}$ [14] are compared with the proposed solvers $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ and $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}$ on images containing either translated or rigidly-transformed coplanar repeated patterns with increasing amounts of lens distortion. (top) small distortion, rigidly-transformed; (middle) medium distortion, translated; (bottom) large distortion, rigidly-transformed. Accurate rectifications for all images are only given by the proposed $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$.

Figure 6.12 compares the proposed $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ and $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}$ solvers to the state-of-the-art solvers on images with increasing levels of radial lens distortion (top to bottom) that contain either translated or rigidly-transformed coplanar repeated patterns. Only the proposed $\mathtt{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ accurately rectifies on both pattern types and at all levels of distortion. The results are after a local optimization and demonstrate that the method of Pritts et al. [74] is unable to accurately rectify without a good initial guess at the lens distortion. The proposed fixed-distortion solver $\mathtt{H}_{22}^{\mathrm{DES}}\mathbf{l}$ gave a better rectification than the change-of-scale solver $\mathtt{H}_{22}^{\mathrm{CS}}\mathbf{l}$ of Chum et al. [14].

Figure 6.13 shows the rectifications of a deceiving picture of a landmark taken by wide-angle and fisheye lenses. From the wide-angle image, it is not obvious which lines are really straight in the scene making undistortion with the plumb-line constraint difficult.

6.7 Discussion

This chapter proposes two groups of solvers (DES and CS) that extend affine-rectification to radially-distorted images that contain essentially arbitrarily repeating coplanar patterns. Both solver groups use the invariant that imaged coplanar repeats have the same scale if rectified. Despite using the equal scale invariant of rectified coplanar repeats in different ways to impose constraints on the undistortion and rectification parameters, the generated solvers have identical

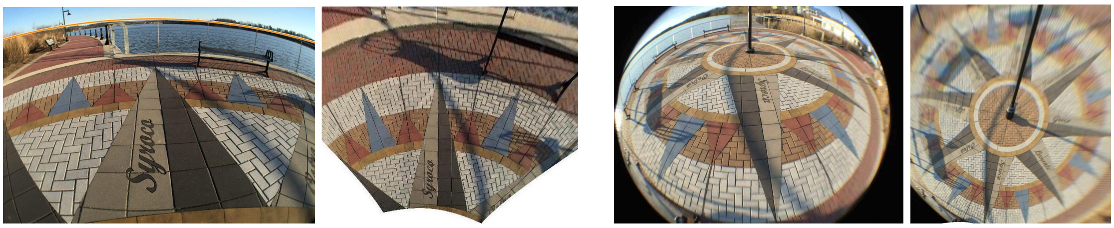


Figure 6.13: *Straight Lines Don't Have to be Straight.* (left pair) It is difficult to disentangle the effects of radial lens distortion from the projections of curvilinear forms in the image. *E.g.*, the waterfront, fence and compass tile mosaic are circles, which violate the plumb-line assumption and cannot be used for undistortion or rectification [22]. However, the imaged rigidly-transformed coplanar repeats can be used to rectify this image with the solvers proposed in this chapter. (right pair) Note that the distortion center is clearly decentered in the third image, but a good rectification is still achieved for the fisheye image.

structure and similar stability and robustness to imaging noise. This was a surprising finding since the CS solvers linearize the undistorting and rectifying transformation to generate the constraint equations. Given the results for the CS solvers on synthetic benchmarks and challenging images, it can be concluded that the first-order approximation of the rectifying transformation is sufficient to handle the effect of severe lens distortion of an obliquely imaged scene plane. Equivalently, the linearization is reasonable over a measurement region that is typical for an affine-covariant region detection.

Synthetic experiments show that both groups of proposed solvers are more robust to noise with respect to the state of the art, give stable estimates across a wide range of distortions, and are applicable to a broader set of image content. The paper also demonstrates that robust solvers can be generated with the basis selection method of [54] by maximizing for numerical stability. We expect basis selection to become a standard procedure for improving solver stability. Experiments on difficult images with large radial distortions confirm that the solvers give high-accuracy rectifications if used inside a robust estimator. By jointly estimating rectification and radial distortion, the proposed minimal solvers eliminate the need for sampling lens distortion parameters in RANSAC.

In future work, we will attempt to remove the degeneracies from the solvers unrelated to the problem formulation. Another future direction, similar to the recent work of [12], is to generate a set of hybrid solvers by combining constraint equations from the DES and CS and the conjugate translation solvers proposed in Chapter 5. The constraint equations for the DES and CS solvers may be sensitive to different properties of the inputted covariant regions, such as their size, shape and relative orientation. During sampling, the most robust solver given the properties of the minimal sample (as listed above) can be chosen to hypothesize the model.

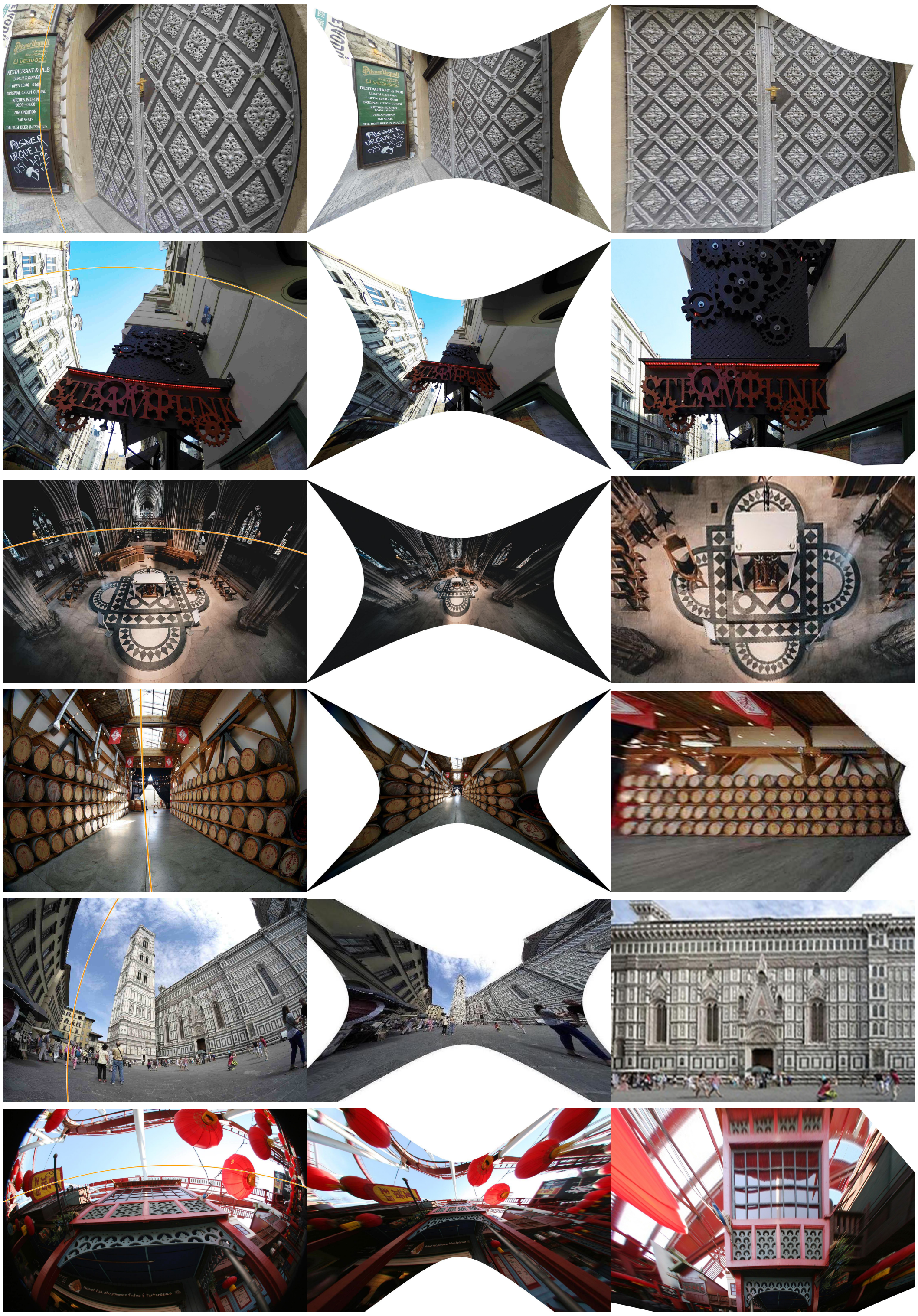


Figure 6.14: *Wide-angle and fisheye results.* Input images (left) with the estimated distorted vanishing line (orange), undistorted (middle) and rectified (right). Results are produced with the proposed $\mathtt{H}_{222}^{\mathrm{DES}}1\lambda$ solver.

7 Coplanar Repeats by Energy Minimization

This chapter proposes an automated method to detect, group and rectify arbitrarily arranged coplanar repeated elements via energy minimization. The proposed energy functional combines several features that model how planes with coplanar repeats are projected into images and captures global interactions between different coplanar repeat groups and scene planes. An inference framework based on a recent variant of α-expansion is described and fast convergence is demonstrated. We compare the proposed method to two widely-used geometric multi-model fitting methods using a new dataset of annotated images containing multiple scene planes with coplanar repeats in varied arrangements. The evaluation shows a significant improvement in the accuracy of rectifications computed from coplanar repeats detected with the proposed method versus those detected with the baseline methods.

7.1 Introduction

Most state-of-the-art repeat detection and modeling methods take a greedy approach that follows appearance-based clustering of extracted local affine frames (LAFs) with geometric verification. Greedy methods have a common drawback: Sooner or later the wrong choice will be made in a sequence of threshold tests resulting in an irrevocable error, which makes a pipeline approach too fragile for use on large image databases.

We propose a global energy model for grouping coplanar repeats and scene plane detection. The energy functional combines features encouraging: (i) the geometric and appearance consistency of coplanar repeated elements, (ii) the spatial and color cohesion of detected scene planes, (iii) and a parsimonious model description of coplanar repeat groups and scene planes. The energy is minimized by block-coordinate descent, which alternates between grouping extracted LAFs into coplanar repeats by labeling (see Figures 7.1 and 7.3) and regresses the continuous parameters that model the geometries and appearances of coplanar repeat groups and their underlying scene planes. Inference is fast even for larger problems (see Section 7.6).

Comparison to state-of-the-art coplanar repeat detection methods is complicated by the fact that many prior methods were either evaluated on small datasets, include only qualitative results, or were restricted to images with repeats having a particular symmetry. We evaluate the proposed method on a new annotated dataset of 113 images. The images have from one to five scene planes containing translation, reflection, or rotation symmetries that repeat periodically or arbitrarily. Performance is measured by comparing the quality of rectifications computed from detected coplanar repeat groups versus rectifications computed from the annotated coplanar repeat groups of the dataset.

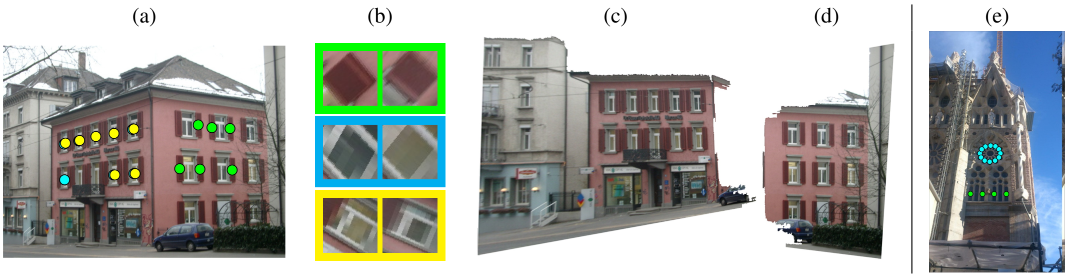


Figure 7.1: *Grouping and Rectification of Coplanar Repeats.* (a) A subset of the detected coplanar repeats is denoted by colored dots. (b) Rectification of the most distant LAF pairs grouped as coplanar repeats—repeat group membership is encoded by the colored border. (c,d) Rectified and segmented scene planes. (e) Translation and rotation symmetric local affine frames labeled as distinct coplanar repeats.

7.2 Related Work

Two frequently cited approaches use energy minimization frameworks. Park et al. [73] minimize an energy that measures the compatibility of a deformable lattice to imaged uniform grids of repetitions. Wu et al. [97] refine vanishing point estimates of an imaged building facade by minimizing the difference between detected symmetries across repetition boundaries of the facade.

None of the reviewed approaches globally model repeats; rather, there is an assumption that a dominant plane is present, or repeat grouping proceeds greedily by detecting scene planes sequentially. A significant subset of the reviewed literature requires the presence of special scene structure like parallel scene lines or lattices, which limits their applicability.

7.3 Scene Model

The scene model has three types of outputs. The first output is a grouping of detected local affine frames (see Figure 7.3) into coplanar repeats (see Figures 7.1). Random variables Y^K jointly assign local affine frames to LAF groups with mutually compatible geometry and appearance and to planar surfaces. Each random variable of Y^K is from the set $\mathcal{Y}_K = \{\, 1 \ldots N_G, \varnothing \,\} \times \{\, 1 \ldots N_V, b \,\}$. Here N_G is the number of clusters of local affine frames that were grouped based on their similarity in appearance, and N_V is the estimated number of planar surfaces in the scene (see Table 7.1). A particular labeling of Y^K is denoted $\mathbf{y}^K$. The assignment of the i-th local affine frame to a compatible LAF cluster is indexed as $\mathbf{y}^K_{ig}$, and its assignment to a scene plane is indexed as $\mathbf{y}^K_{iv}$. The empty set $\varnothing$ is assigned if the local affine frame i does not repeat, $\mathbf{y}^K_{ig} = \varnothing$, and the token b is assigned to a local affine frame if it does not lie on a planar surface. Background local affine frames cannot be assigned to a repeat group, so they are assigned the ordered pair $(\, \varnothing, b \,)$. The non-planar surfaces are collectively called the background. The sets of local affine frames assigned to the same LAF cluster and scene plane are the coplanar repeated patterns.

Term	Description	Term	Description
$\mathbf{x}_i^K$	local affine frame (LAF), see Figure 7.3	$\varnothing$	LAF is a singleton
$\mathbf{x}_{iw}^K$	point of a LAF	N_G	number of LAF clusters
$\mathbf{x}_j^R$	image region, see Figure 7.3	N_V	number of scene planes
$\mathbf{y}_{ig}^K$	LAF ↔ cluster	$\beta^K(\mathbf{y}^K)$	geom./app. parameters for repeats
$\mathbf{y}_{iv}^K$	LAF ↔ scene plane	$\beta^R(\mathbf{y}^K, \mathbf{y}^R)$	geom./app. parameters for planes
$\mathbf{y}_i^K$	LAF label, $(\mathbf{y}_{ig}^K, \mathbf{y}_{iv}^K)$, see Figure 7.1	$\beta(\mathbf{y})$	joint parameter vector
$\mathbf{y}_j^R$	region ↔ scene plane, see Figure 7.1	$\psi(\cdot)$	joint feature vector
$\mathbf{y}$	joint labeling	$\mathbf{l}_n$	scene plane vanishing line
b	LAF/region is on background	$H_{\mathbf{l}_n}(\cdot)$	rectifying transform from $\mathbf{l}_n$

Table 7.1: *Common Scene Model Denotations.*

The second output is a labeling of image regions as planar surfaces and background. The image regions are small and connected areas of similar color that are detected as SEEDS superpixels [90] (see Figure 7.3). Random variables Y^R assign image regions to planar surfaces and the background, where each random variable of Y^R is from the set $\mathcal{Y}_R = \{1 \ldots N_V, b\}$. As before, N_v and b are the estimated number of planar surfaces and the background token, respectively. A particular labeling of Y^R is denoted $\mathbf{y}^R$, and the labeling partitions the image regions into larger components that correspond to contiguous planar surfaces of the scene or background. The assignment of the j-th region to a scene plane or to background is indexed as $\mathbf{y}_j^R$.

The third output is a set of continuous random variables modeling the geometries and appearances of the sets of coplanar repeats and the scene planes. The geometries and appearances of coplanar repeats are functions of the local affine frame assignments and are given by the dependent random variables $B^K(Y^K)$. The corresponding parameter estimates are denoted as $\beta^K(\mathbf{y}^K)$. The geometries and appearances of the scene planes are functions of Y^K and Y^R, and are given by dependent random variables $B^R(Y^K, Y^R)$. The parameters $\beta^R(\mathbf{y}^K, \mathbf{y}^R)$ represent the colors of the scene surfaces and the orientations of scene planes.

The joint labeling and parameter vector for the entire model are respectively denoted $\mathbf{y} = \mathbf{y}^{K\frown}\mathbf{y}^R$ and $\beta(\mathbf{y}) = \beta^K(\mathbf{y}^K)^\frown\beta^R(\mathbf{y}^K, \mathbf{y}^R)$.

xsy

7.3.1 Energy Function

The joint feature vector $\psi(\cdot)$ encodes potentials that measure: (i) coplanar repeats consist of local affine frames that have similar appearance and the same area in the preimage, (ii) the scene planes and background should consist of image regions with the same color distributions, (iii) surfaces should be contiguous and that nearby repeated content should be on the same surface, (iv) and scenes should have a parsimonious description.

A minimal energy labeling $\mathbf{y}$ and parameter set $\beta(\mathbf{y})$ are sought by solving the energy mini-

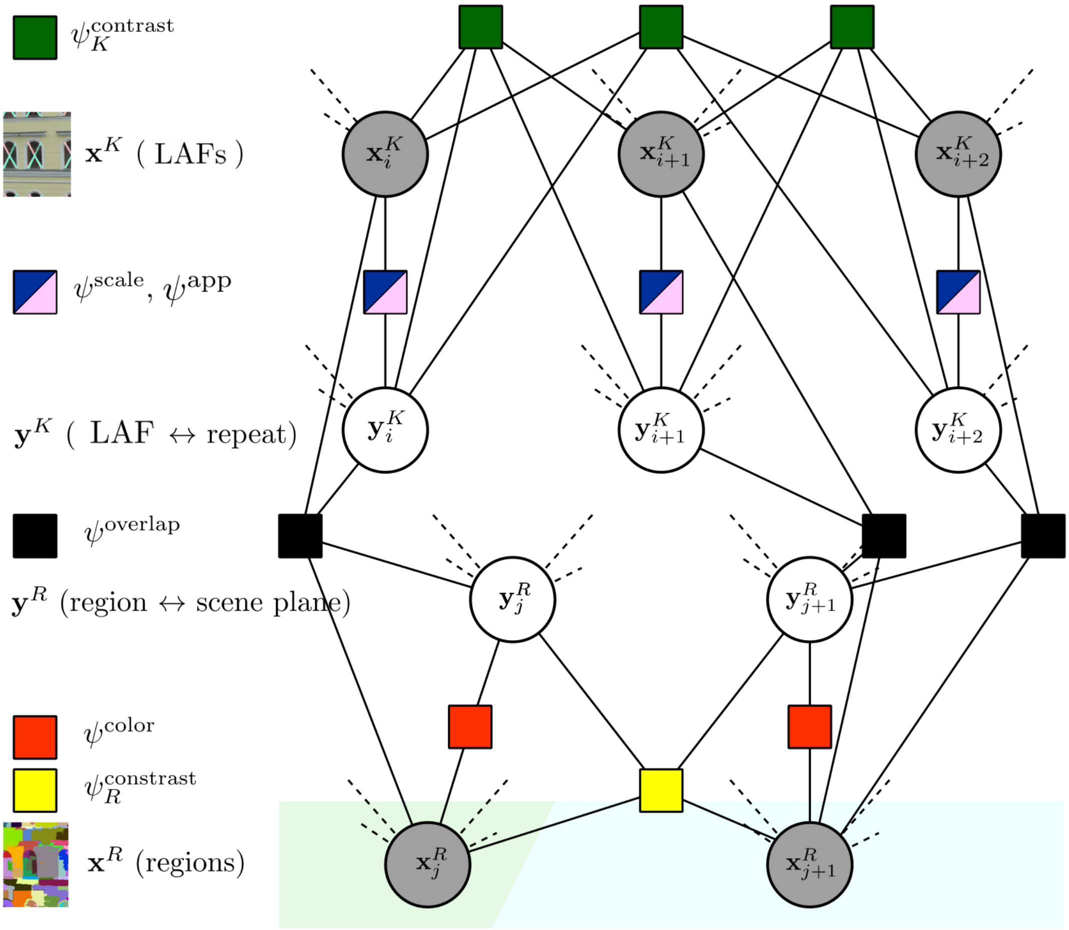


Figure 7.2: *Factor Graph for Unaries and Pairwise terms.* The energy function incorporates (i) unaries for geometric and appearance consistency of coplanar repeats, (ii) pairwise terms for spatial and color cohesion of detected scene planes

mization task

$$\operatorname*{argmin}_{\mathbf{y},\beta} \underbrace{\mathbf{w}^\top \psi(\mathbf{x}, \mathbf{y}, \beta(\mathbf{y}))}_{E \text{ (energy)}}, \tag{7.1}$$

where $\mathbf{x}$ are the detected salient image patches and over-segmented regions of the image, and $\mathbf{w}$ is a weight vector. The components of $\mathbf{w}$ take on different meanings depending on their paired features and are discussed in Sections 7.3.3 and 7.3.5. Figure 7.2 is a factor graph defining the interactions of the unaries and pairwise energies in the energy function. It will be a useful reference as the energy terms are defined.

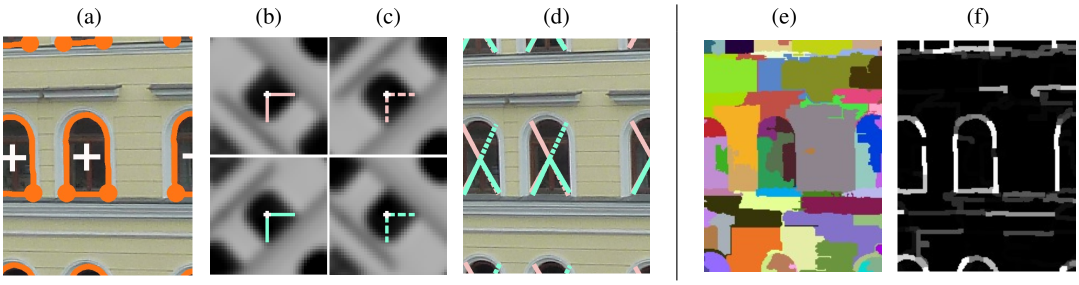


Figure 7.3: *Region Detection and Description.* (a) Center of gravity (white cross) and curvature extrema (orange circles) of a detected MSER (orange contour [62], see also Section 3.2.4). Patches are normalized to a square and oriented to define an affine frame as in [63], (b) Bases are reflected for detecting axial symmetries. The RootSIFT transform embeds the local texture [4, 60]. (c) Affine frames are mapped back into image. (d) Over-segmentation by SEEDS superpixels. (e) The contrast feature ψ_T^{contrast}, where intensity is proportional to edge response along superpixel boundaries.

7.3.2 Measurements

Affine-covariant regions [62, 65, 69] are extracted from the image as good candidates for representing repeated content. (see Figures 7.3). The shapes of the detected patches are summarized by local affine frames, or, equivalently, 3-tuples of points, and are given by measurements $\mathbf{x}^K$. One type of local affine frame construction is illustrated in Figure 7.3 (left). The image is over-segmented by SEEDS superpixels [90] to provide measurements on regions where local affine frame detection is unlikely as illustrated in Figure 7.3 (right). The segmented regions are denoted by $\mathbf{x}^R$. The local affine frames and regions are concatenated to give the joint measurement $\mathbf{x} = \mathbf{x}^{K\frown}\mathbf{x}^R$, which is an argument to the energy defined in (7.1).

7.3.3 Unary Features for Repeats and Surfaces

The perspective skew of each scene plane π_n is given by its vanishing line, which is an analog to the horizon line for a scene plane at any orientation. Vanishing lines are encoded in the parameters of the scene planes $\beta^R(\mathbf{y}^K, \mathbf{y}^R)$. Explicitly they are the set $\{\, \mathbf{l}_n \mid \mathbf{l}_n \in \mathcal{P}^2 \,\}_{n=1}^{N_V}$, where N_V is the number of scene planes and $\mathcal{P}^2$ is the real projective plane.

Scale of coplanar repeats. A coplanar repeat group C is the set of local affine frames from the same pattern that co-occur on a scene plane, namely $C = \{\, \mathbf{x}_i^K \mid \mathbf{y}_{ig}^K = m \wedge \mathbf{y}_{iv}^K = n \,\}$, where $n \neq b$. The local affine frames of C are called *coplanar repeats*. The coplanar repeats of C are of equal scale (equiareal) if their perspective skew is removed, which is accomplished by transforming the vanishing line of the underlying scene plane $\mathbf{l}_n$ so that it is coincident with the principal axis of the camera (see Chum et al. [14]). The scale feature ψ^{scale} measures the mutual compatibility of coplanar repeats with the scale constraint. Let $H_{\mathbf{l}_n}(\cdot)$ be a transformation that removes perspective skew from plane π_n by orienting $\mathbf{l}_n$ to the principal axis and $s(\cdot)$ be the function that computes the scale of a local affine frame. Then the scale feature for the scene's coplanar repeats is

$$\psi^{\text{scale}} = -\sum_{m=1}^{N_G}\sum_{n=1}^{N_V}\sum_{i}[\mathbf{y}^K_{ig}=m]\cdot[\mathbf{y}^K_{iv}=n]\cdot\big(\log s(H_{\mathbf{l}_n}(\mathbf{x}^K_i)) - \log\ \bar{s}(n,\mathbf{y}^K_{ig})\big)^2, \quad (7.2)$$

where $\bar{s}(n,\mathbf{y}^K_{ig})$ is the geometric mean of the local affine frames in pattern $\mathbf{y}^K_{ig}$ rectified by transformation $H_{\mathbf{l}_n}(\cdot)$, which is part of the estimated parameters of the repeated scene content encoded in $\beta^F(\mathbf{y}^K)$.

Appearance of patterns. The appearance of the image patches containing LAFs $\mathbf{x}^K$ are described by RootSIFT [4, 60]. The corresponding RootSIFT of a local affine frame is given by the function $\mathbf{r}(\cdot)$ (see Section 3.3.1). The appearance affinity of the local affine frame $\mathbf{x}^K_i$ to a pattern is given by the normalized Euclidean distance between the RootSIFT descriptor of the local affine frame and mean RootSIFT descriptor of the pattern. The appearance feature for patterns is

$$\psi^{\text{app}} = \sum_{m=1}^{N_G}\sum_{i}[\mathbf{y}^K_{ig}=m]\cdot\frac{\|\mathbf{r}(\mathbf{x}^K_i)-\bar{\mathbf{r}}(\mathbf{y}^K_{ig})\|_2^2}{\sigma_1^2}, \quad (7.3)$$

where $\bar{\mathbf{r}}(\mathbf{y}^K_{ig})$ is the mean of the RootSIFTs of the local affine frames in pattern $\mathbf{y}^K_{ig}$, which is part of the estimated parameters of the repeated scene content encoded in $\beta^F(\mathbf{y}^K)$. The variance σ_1^2 is set empirically.

Color of scene surfaces. The color distribution of each scene surface is modeled with a RGB Gaussian mixture model (GMM) with K components, $\gamma = \{\mu_{nk},\Sigma_{nk},\tau_{nk},\}$, where $nk \in \{1\dots N_V,b\}\times\{1\dots K\}$ and $\mu_{nk},\Sigma_{nk},\tau_{nk}$ are the mean RGB color, full color covariance and mixing weight for component k of surface v. The set of GMM parameters γ is part of the estimated parameters of the appearance and geometry for scene planes encoded in $B^R(Y^K,Y^R)$. The color feature for the scene surfaces is

$$\psi^{\text{color}} = \sum_{n\in\{1\dots N_V,b\}}\sum_{j}\sum_{j'}\frac{[\mathbf{y}^R_{jv}=n]}{|\mathbf{x}^R_j|}\cdot\underbrace{\min_{k\in\{1,\dots,K\}}\big\{-\log\big(p_n(\mathbf{x}^R_{jj'}|k)\cdot\tau_{nk}\big)\big\}}_{\text{approximately}\,\propto\,-\log p_n(\mathbf{x}^R_{jj'})}, \quad (7.4)$$

where $\mathbf{x}^R_{jj'}$ is the j'-th member pixel of region $\mathbf{x}^R_j$ with $|\mathbf{x}^R_j|$ number of pixels and the conditional likelihood of a pixel $\mathbf{x}^R_{jj'}$ given a mixture component k is normally distributed, $\mathbf{x}^R_{jj'}|k \sim \mathcal{N}(\mu_{nk},\Sigma_{nk})$. The feature ψ^{color} uses the same approximation for the log-likelihood as Grabcut [81] to make the maximum-likelihood estimation of GMM parameters faster. Connected components of regions with the same surface assignment segment the image into contiguous planar and background regions.

Planar and background singletons. Singletons are local affine frames that do not repeat. A weighted cost for each singleton is assessed, which is the maximum unary energy that can be considered typical for a coplanar repeat. For a complete geometric parsing of the scene, it is necessary to assign each singleton to its underlying scene plane or to the background surface. Singletons induce no single-view geometric constraints nor appearance constraints because they are not part of a repeat group, so their assignments to scene planes are based on their interactions with neighboring local affine frames and regions, which are defined in Section 7.3.4 as assignment regularization functions. An additional weighted cost for each planar singleton is assessed, which is the minimum amount of required evidence obtained through interactions with neighboring local affine frames and regions to consider a singleton planar.

7.3.4 Pairwise

The pairwise features are a set of bivariate Potts functions that serve as regularizers for local affine frame and region assignment to scene model components.

Local affine frame contrast. The local affine frame contrast feature penalizes models that over-segment similar looking repeats. The local affine frame contrast of the scene is

$$\psi_F^{\text{contrast}} = \sum_{i \neq i'} [\mathbf{y}_{iv}^K \neq \mathbf{y}_{i'v}^K] \cdot \exp\left[-\frac{\|\mathbf{r}(\mathbf{x}_i^K) - \mathbf{r}(\mathbf{x}_{i'}^K)\|_2^2}{\sigma_2^2}\right], \tag{7.5}$$

where the variance σ_2^2 is set empirically.

Region contrast. Regions have bounded area, so there may be large areas of low texture on a scene plane or in the background that are over-segmented. Regions that span low-texture areas can be identified by a low cumulative edge response along their boundary. The cumulative edge response between two regions, denoted $\phi(\mathbf{x}_j^R, \mathbf{x}_{j'}^R)$, is robustly calculated so that short but extreme responses along the boundary do not dominate (see Figures 7.3). The region contrast of the image is given by the feature

$$\psi_R^{\text{contrast}} = \sum_{j \neq j'} [\mathbf{y}_{jv}^R \neq \mathbf{y}_{j'v}^R] \cdot \exp\left[-\frac{\phi(\mathbf{x}_j^R, \mathbf{x}_{j'}^R)^2}{\lambda}\right]. \tag{7.6}$$

A larger constant λ increases the amount of smoothing and is set as $\lambda = 2 \cdot \bar{\phi}^2$, which puts the crossover point of smoothing at the mean contrast of regions.

Local affine frame overlap. A local affine frame that overlaps a region is coplanar or co-occurs on the background surface with the overlapped region, which is encoded as a pairwise constraint. A penalty for each violation of the coplanarity constraint is assessed.

7.3.5 Label subset costs

Parsimonious scene models are encouraged by assessing a cost for each scene model part. Equivalence classes of the label set are defined by labels that share a scene model part, *e.g.*, the set of labels that have the same vanishing line. A label subset cost is assessed if at least one label from an equivalence class is used, which is equivalent to accumulating a weighted count of the number of unique scene model components in the scene.

7.4 Energy Minimization

The energy minimization task of (7.1) is solved by alternating between finding the best labeling $\mathbf{y}$ and regressing the scene model components β in a block-coordinate descent loop until the energy converges. Alternating between finding the minimal energy labeling and regressing continuous model parameters has notably been used in segmentation and multi-model geometry estimation by Rother et al. and Isack et al. [81, 35].

7.4.1 Labeling and Regression

The scene model parameters are fixed to the current estimate for the labeling problem, $\hat{\mathbf{y}} = \operatorname{argmin}_{\mathbf{y}} E(\mathbf{x}, \mathbf{y}, \beta(\mathbf{y}) = \hat{\beta})$. Finding the minimal-energy labeling is NP-hard [9]. An extension to alpha-expansion by Delong et al. [9, 21, 37] that accommodates label subset costs (defined in Section 7.3.5) is used to find an approximate solution.

The labeling is fixed to the current estimate for the regression subtask $\hat{\beta} = \operatorname{argmin}_{\beta} E(\mathbf{x}, \mathbf{y} = \hat{\mathbf{y}}, \beta(\hat{\mathbf{y}}))$. Each continuous parametric model must be regressed with respect to its dependent unary potentials so that the energy does not increase during a descent iteration. In particular, the vanishing lines, surface color distributions and the representative appearance for patterns and rectified scale for coplanar repeats are updated as detailed in the following paragraphs. The updated parameters are aggregated in $\hat{\beta}$.

Vanishing lines. All local affine frames assigned to the same planar surface are used to refine the surface's vanishing line orientation. The objective is the same as the unary defined in eq. 7.2 and encodes the affine scale invariant defined in Chum et al. [14]. The vanishing line is constrained to the unit sphere and so that all local affine frames are on the same side of the oriented vanishing line,

$$\mathbf{l}_n^* = \operatorname*{argmin}_{\mathbf{l}} \sum_{i\,:\,\hat{\mathbf{y}}_{iv}^K = n} \left(\log s(H_{\mathbf{l}}(\mathbf{x}_i^K)) - \frac{1}{\sum_{i'} [\hat{\mathbf{y}}_{ig}^K = \hat{\mathbf{y}}_{i'g}^K]} \log \sum_{i'} [\hat{\mathbf{y}}_{ig}^K = \hat{\mathbf{y}}_{i'g}^K] \cdot s(H_{\mathbf{l}}(\mathbf{x}_{i'}^K)) \right)^2 \tag{7.7}$$

$$\begin{aligned} \text{s.t.} \quad \mathbf{l}^\top \mathbf{x}_{iw}^K > 0, &\quad w \in \{1 \dots 3\} \\ \mathbf{l}^\top \mathbf{l} = 1, & \end{aligned} \tag{7.8}$$

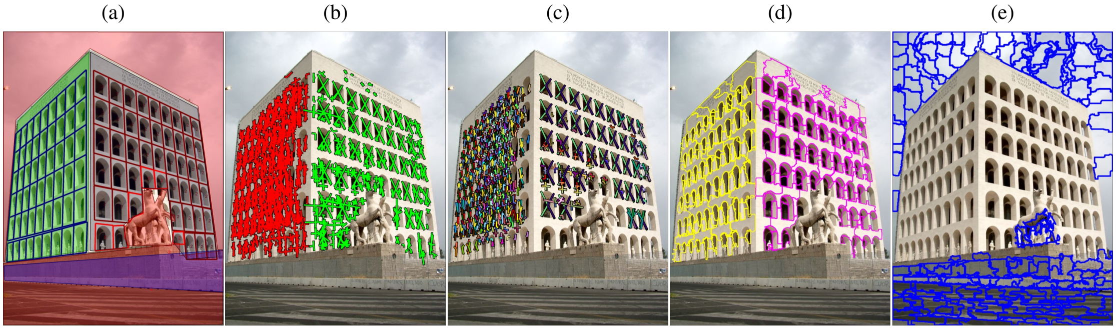


Figure 7.4: *The Hierarchical Annotations.* The annotations are included with the 113 image dataset. (a) Translation symmetries are annotated by grids, regions that cannot share coplanar repeats are colored differently. (b) Detected local affine frames to vanishing line assignment. (c) Groups of coplanar repeated local affine frames found by annotation-assisted inference. (d) Image regions (SEEDS superpixels [90]) to vanishing line assignment. (e) Background image regions, which coplanar repeats cannot overlap.

for all scene planes n that have patterns assigned, where $s(\cdot)$ is the scale of a local affine frame and $H_\mathbf{l}(\cdot)$ is the rectifying transform as defined in Section 7.3.3, and $\mathbf{x}^K_{iw}$ denotes the individual homogeneous coordinates that define local affine frame $\mathbf{x}^K_i$. The constrained nonlinear program is solved with the MATLAB intrinsic FMINCON.

Coplanar repeats and patterns. For features ψ^{scale} eq. (7.2) and ψ^{app} eq. (7.3) that are sums of squared differences, the parameters are estimated as a mean of the respective values.

Surface color distribution. The parameters of the color distribution of a surface are estimated from the member pixels of regions assigned to the surface. The approximate log-likelihood defined for the unary ψ^{color} in eq. 7.4 is maximized to estimate the Gaussian mixture for each surface that has region assignments,

$$\{\,\Sigma^*_{nk}, \mu^*_{nk}, \pi^*_{nk}\,\}_{k=1}^K = \operatorname*{argmax}_{\{\Sigma_{nk},\mu_{nk},\tau_{nk}\}_{k=1}^K} \prod_{j:\hat{\mathbf{y}}^R_j=n} \prod_{j'} \max_{k'} p_v(\mathbf{x}^R_{jj'} \mid k'; \Sigma_{nk}, \mu_{nk}, \tau_{nk}) \cdot \tau_{nk'}. \tag{7.9}$$

The objective defined in eq. 7.9 is maximized by block-coordinate ascent in a manner similar to Lloyd's algorithm: The mixture component assignments are fixed to estimate the means and covariances and then vice-versa in alternating steps. A fixed number of iterations is performed.

7.4.2 Proposals

The initial minimal labeling energy requires a guess β^0 at the continuous parameters $\beta(\mathbf{y})$. This is provided by a proposal stage in which the local affine frames $\mathbf{x}^K$ are clustered by their

RootSIFT descriptors and sampled to generate vanishing line hypotheses as in Chum et al. [14]. The clustered regions are verified against the hypothesized vanishing lines to create a putative collection of coplanar repeats that are scale-consistent after affine rectification by a compatible sampled vanishing line. The proposed coplanar repeat groups do not partition the local affine frames, which is a constraint enforced by the minimal energy labeling $\hat{\mathbf{y}}$. The initial color model for each detected surface (equivalently proposed vanishing lines and background) is estimated from the image patches of local affine frames from the proposed coplanar repeat groups.

7.5 Dataset

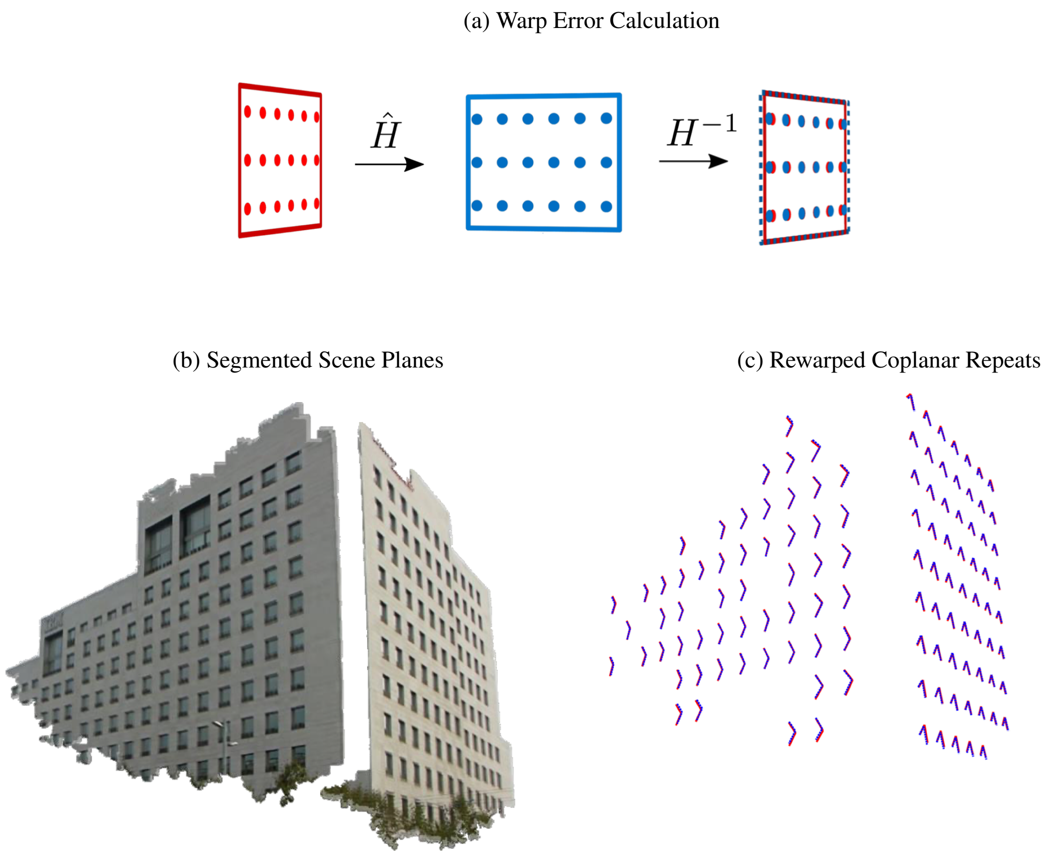
We introduce a dataset of 113 images containing from one to five scene planes with translated, reflected and rotated coplanar repeats occurring periodically or arbitrarily (see Figure 7.4). The dataset includes images from the ZuBuD database of Shao et al. and the CVPR 2013 symmetry database assembled by Liu et al. [83, 58, 72]. The manual assignment of local affine frames to coplanar repeat groups is infeasible since a typical image will have thousands of extracted local affine frames. Direct annotation is also undesirable since setting changes of the local affine frame detectors would invalidate the assignments. Instead, the annotations are designed to constrain the search for coplanar repeated local affine frames, making annotations agnostic to the local affine frame type. The annotations hierarchically group parallel scene planes, individual scene planes, and areas within a scene plane that cannot mutually have the same coplanar repeats, *i.e.* denoting distinct patterns. Clutter and non-planar surfaces are also segmented. LAF-level assignment to coplanar repeat groups is achieved using a RANSAC-based estimation framework which leverages the annotations to constrain the search for correspondences to choose the correct transformation type.

7.6 Experiments

We evaluate the proposed method against two state-of-the-art geometric multi-model fitting methods: J-Linkage and MultiRANSAC [99, 89]. Both estimators are hypothesize-and-verify variants. A model hypothesis consists of a vanishing line and tentatively grouped local affine frames of similar appearance. Coplanar repeat group assignments are verified by a threshold test on the similarity measure for repeated local affine frame detection proposed by Shi et al. [84]. However, the rectified scale constraint defined in Eq. 7.2 is used in lieu of the scale kernel used by [84]. We provide the number of scene planes present in each image to MultiRANSAC.

The accuracy of rectifications constructed from vanishing lines computed from detected coplanar repeat groups are used to compare the methods. Two necessary conditions for accurate rectifications are that: (i) no outliers are included in the detected coplanar repeat groups, (ii) and detected coplanar repeat groups densely cover the extents of the scene plane where there are coplanar repeat groups annotated in the dataset. Thus the rectification accuracy of coplanar repeats serves as a proxy measure for the precision and recall of coplanar repeat detection.

The accuracy of the rectification is evaluated with the warp error (see Section 2.12), where the annotated coplanar repeats are used to compute the ground truth rectifying transformation.



(a) Warp Error Calculation

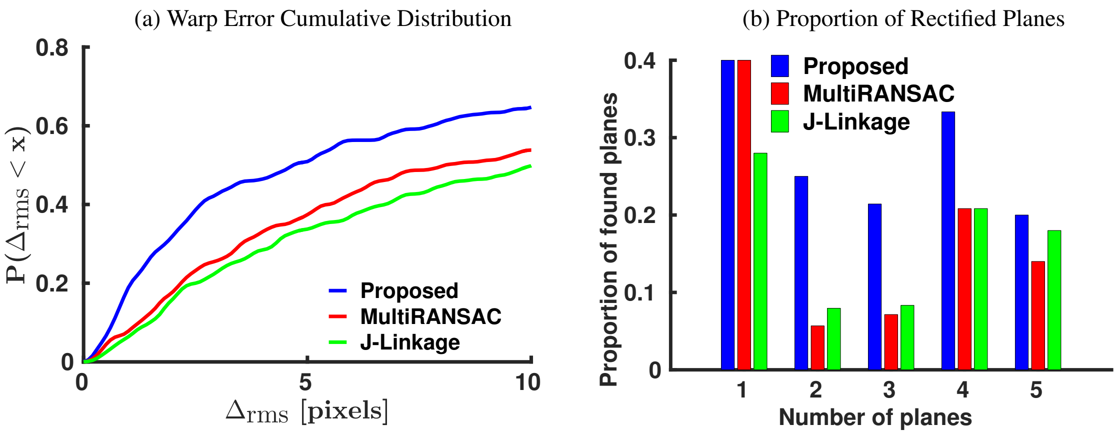
(b) Segmented Scene Planes

(c) Rewarped Coplanar Repeats

Figure 7.5: *Warp Error Evaluation.* (a) Rectification accuracy is measured with the warp error. Annotated coplanar repeats are rectified with the ground-truth transformation $\hat{H}(\cdot)$ and rewarped with the inverse estimated rectification $H^{-1}(\cdot)$ (see also Section 2.12). (b) The segmented scene planes contain annotated coplanar repeats that are used for calculating the warp error. (c) The annotated coplanar repeats (colored red) are rewarped (colored blue) and used to compute the warp error.

A rectilinear camera is assumed, so the division model parameter in (2.37) is set to zero, namely $\lambda = 0$. The set of annotated coplanar repeats that is the largest proportion of the detected coplanar repeats is used to match the rectification computed from detected coplanar repeats to a rectification computed from annotated coplanar repeats (see Figures 7.5b and 7.5c).

The cumulative distribution of the warp error on the dataset (truncated at 10 pixels) is shown in Figure 7.6a. At 1 pixel of the warp error, the proposed method solves 163% more scene planes than the next best; at 2 pixels, 94% more; and at 5 pixels, which can be considered a threshold for meaningful rectification, 51% more scene planes. Figure 7.6b plots the proportion of scene planes rectified with less than 2 pixels of the warp error with respect to the number of scene


(a) Warp Error Cumulative Distribution

(b) Proportion of Rectified Planes

Figure 7.6: *Evaluation: Accuracy of Rectifications.* (a) Cumulative distribution function of the warp error Δ_{rms} computed on the 113 image dataset. (b) Proportion of planes rectified with less than 2 pixels of the warp error in images with 1 to 5 scene planes

planes in the image. Clearly the proposed method excels when there are multiple scene planes present.

Figure 7.7 plots the cumulative runtime of the labeling step for images as function of the number of local affine frames and image regions, denoted *sites*, and the number of active model proposals, denoted *labels*. Inference ranges from under a second to two minutes for the largest problems in the dataset.

7.7 Discussion

The proposed energy minimization formulation demonstrates a distinct increase in the quality of rectifications estimated from detected coplanar repeat groups on the evaluated dataset with respect to two state-of-the-art geometric multi-model fitting methods. The advantage can be attributed to the global scene context that is incorporated into the energy functional of the proposed method. The evaluation was performed on a new annotated dataset of images with coplanar repeats in diverse arrangements.

Despite a significant improvement over the baseline, the proposed method failed to solve roughly half of the dataset with less than 5 pixels of the warp error. Future work will incorporate constraints specific to reflected and rotated local affine frames and parallel scene lines, which would add significant geometric discrimination to the model. Learning the feature weight vector $\mathbf{w}$, which was hand tuned, could also give a significant performance boost. However, the complete annotation of coplanar repeated local affine frames in an image is probably infeasible. This means structured output learning must be performed with partial annotations, which complicates the learning task considerably.

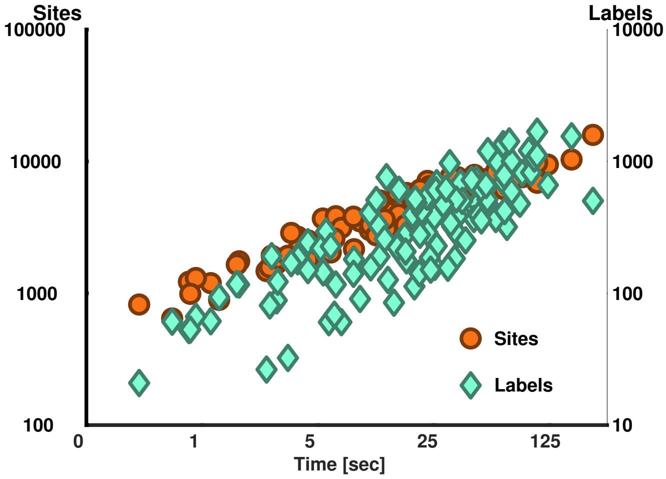


Figure 7.7: *Evaluation: Cumulative Runtime.* Cumulative wall time in seconds for the labeling task of energy minimization.

7.8 Annotation-Assisted Repeat Grouping

The annotations provided by the 113 image dataset referenced in the paper are discussed in detail The annotations hierarchically segment the image into parts that: (i) are scene planes (ii) are the union of scene planes that share the same vanishing line (iii) contain repeated content (iv) are the union of repeated content annotations that are distinctly different from other repeated content in the remainder of the image. In particular the repeated content annotations are specific to the type of symmetry exhibited by the repeat: namely annotations for translational and rotational symmetries are provided. In addition lattices are provided for translationally symmetric periodic repeats.

Individual salient features (*e.g.* Hessian Affine Detections or MSERs) are not grouped or annotated, so the annotations are feature agnostic, which is preferable since settings adjustments would invalidate such annotations. Rather, the annotations are used to assist a RANSAC-based inference algorithm to establish coplanar repeat groups. The annotations constrain the search for correspondences, which gives a much higher inlier percentage among tentative groupings that are inputted to RANSAC. Since the transform type is known from the annotations, the transform with the fewest required constraints can be used, which improves the probability of proposing a transform estimated from all-inliers. The vanishing line is estimated, and, depending on the annotation tag, either a translation or rotation and translation, which maps repeats onto each pointwise. The annotations are tagged so that the correct transformation can be estimated during annotation-assisted inference.

Even with this relaxed standard of annotation, it is impossible to group repeats at their highest frequency of recurrence. Depending on the features extracted, *e.g.*, corners of facade ornamentation may be detected, where only the windows were marked as repeated. Thus any performance evaluation must not penalize methods that correctly identify repeats that recur at higher frequencies than the annotations. Reflections and rotational symmetries, in particular, exacerbate this problem. Perhaps the most common example in the dataset are window panes, which have axial symmetry, and if square, rotational symmetry. It is not practical to annotate all such occurrences (not just restricted to windows) in a large dataset. The annotations also group oversegmentations of the image (*i.e.* superpixels in this context) into contiguous components of planes, sets of parallel planes and background surface. These annotations are not currently used in the evaluation, but would be useful for learning the regularization weights in the energy function.

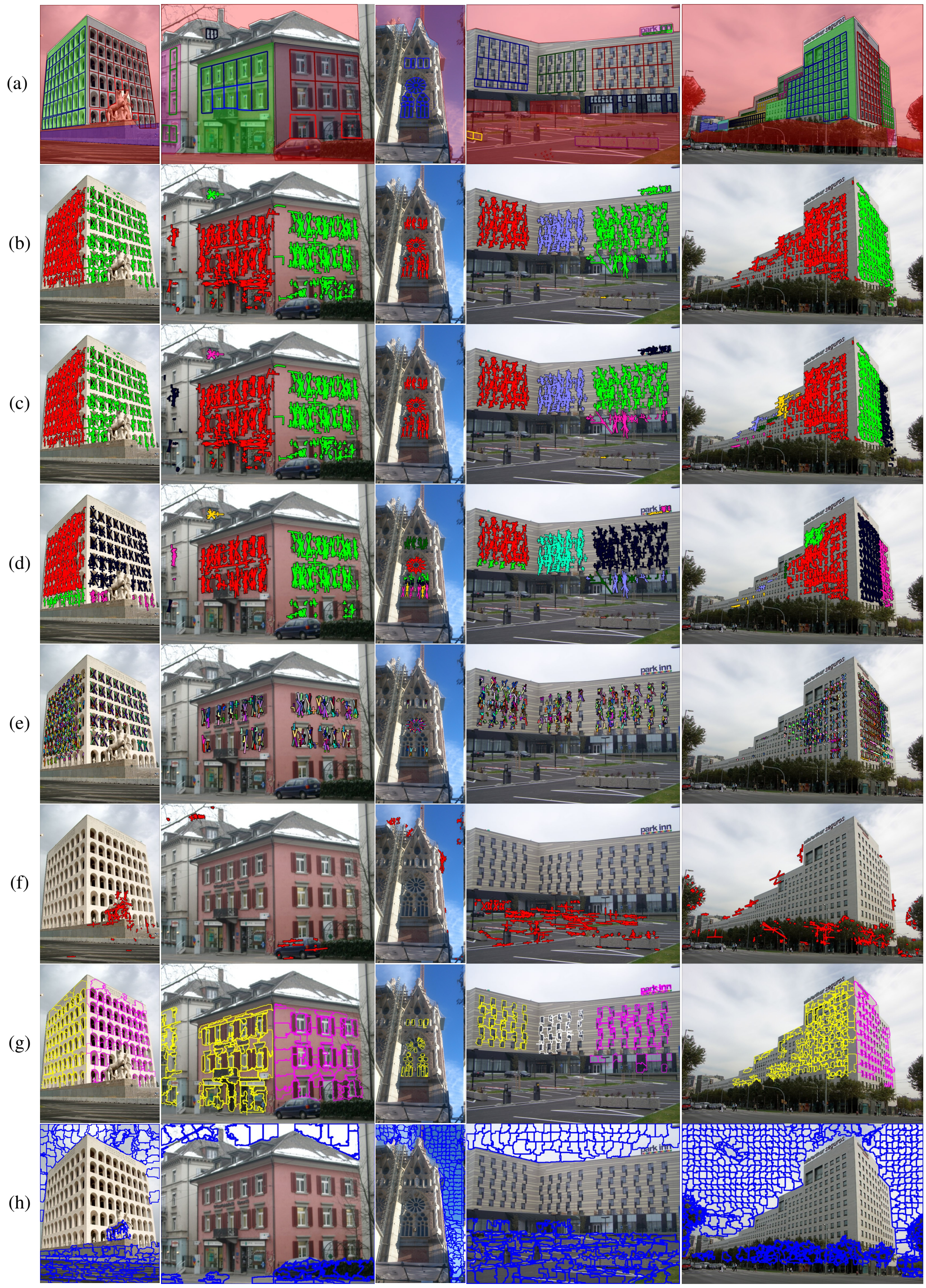


Figure 7.8: *Annotations.* (a) Constraints coplanar repeat grouping. (b) Vanishing line assignment. (c) Plane assignment. (d) Mutually distinct repeated content. (e) Coplanar repeats found by annotation-assisted inference. (f) Features on the background surface. (g) Vanishing line assignment for regions. (h) Regions on the background surface.

8 Conclusions

This thesis proposes a suite of minimal solvers for rectifying from the radially-distorted images of scene planes with repeated texture. The solvers differ by the assumptions made on the configuration of the repeated texture. In particular, minimal problems are solved from novel constraints on affine rectification induced by the distorted images of translational symmetries, reflections and rigidly-transformed coplanar repeated textures. The solvers are generated using techniques from both elementary algebraic geometry and the state-of-the-art methods in automated solver generation. In general, the solvers have excellent stability and robustness to measurement noise. The solvers extend rectification to distorted images with repeated coplanar scene content, which may not contain a sufficient number of detectable straight lines for the state of the art to rectify.

In addition, an energy functional is proposed that incorporates terms for global scene context and model parsimony with a geometric unary that measures the consistency of scene plane models proposed by the rectifying solvers. The proposed global model has two important properties: (a) the energy minimization jointly evaluates rectification models proposed by rectifying minimal solvers, and (b) higher-order terms encouraging smooth scene-plane segmentation and model parsimony regularize the relatively weak geometric unary. These two properties were shown to provide significant gains in rectification and segmentation accuracy over a greedy method and global method that used the geometric unary alone on scenes with multiple planes. Furthermore, the proposed minimal solvers can be plugged into the energy minimization framework to extend their applicability to scenes without a dominant plane.

Bibliography

[1] S. Ahmad and L.-F. Cheong. Robust detection and affine rectification of planar homogeneous texture for scene understanding. *International Journal of Computer Vision*:1–33, 2018.

[2] D. Aiger, D. Cohen-Or, and N. J. Mitra. Repetition maximization based texture rectification. *Computer Graphics Forum*, 31(2pt2):439–448, 2012.

[3] M. Antunes, J. P. Barreto, D. Aouada, and B. Ottersten. Unsupervised vanishing point detection and camera calibration from a single manhattan image with radial distortion. In *CVPR*, July 2017.

[4] R. Arandjelović and A. Zisserman. Three things everyone should know to improve object retrieval. In *CVPR*, 2012.

[5] B. Sturmfels. Solving systems of polynomial equations. In *American Mathematical Society, CBMS Regional Conferences Series, No 97*, 2002.

[6] D. Barath and L. Hajder. A theory of point-wise homography estimation. *Pattern Recognition Letters*, 94:7–14, 2017. ISSN: 0167-8655.

[7] A. Baumberg. Reliable feature matching across widely-separated views. In *CVPR*, 2000.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004. ISBN: 0521833787.

[9] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.

[10] F. Bukhari and M. N. Dailey. Automatic radial distortion estimation from a single image. *Journal of Mathematical Imaging and Vision*, 45(1):31–45, January 2013. ISSN: 1573-7683.

[11] M. Byröd, K. Josephson, and K. Åström. Improving numerical accuracy of gröbner basis polynomial equation solvers. In *International Conference on Computer Vision (ICCV)*. IEEE, 2007.

[12] F. Camposeco, A. Cohen, M. Pollefeys, and T. Sattler. Hybrid camera pose estimation. In *CVPR*, June 2018.

[13] J. Cech, J. Matas, and M. Perdoch. Efficient sequential correspondence selection by cosegmentation. In *CVPR*, 2008.

[14] O. Chum and J. Matas. Planar affine rectification from change of scale. In *ACCV*, 2010.

[15] O. Chum, J. Matas, and Š. Obdržálek. Enhancing RANSAC by generalized model optimization. In *ACCV*, 2004.

[16] C. Colombo. The Family of Rectifying Homographies Generated by the Circular Points:2.

[17] D. A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2015, page 646. ISBN: 978-3-319-16720-6.

[18] D. Cox, J. Little, and O. D. Using algebraic geometry. In second edition. Springer, 2004.

[19] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. English. Springer, 2nd edition, 2005. URL: http://www.cs.amherst.edu/~dac/uag.html.

[20] A. Criminisi and A. Zisserman. Shape from texture: homogeneity revisited. In *BMVC*, 2000.

[21] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *IJCV*, 96:1–27, 2012.

[22] F. Devernay and O. Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 13(1):14–24, August 2001. ISSN: 1432-1769.

[23] I. Eichhardt and D. Chetverikov. Affine correspondences between central cameras for rapid relative pose estimation. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[24] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 24(6):381–395, June 1981.

[25] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. ISSN: 0001-0782. URL: http://doi.acm.org/10.1145/358669.358692.

[26] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *CVPR*, 2001.

[27] K. Fukuda, A. Jensen, and R. Thomas. Computing gröbner fans. English. *Math. Comput.*, 76(260):2189–2212, 2007. URL: http://www.ams.org/journals/mcom/2007-76-260/S0025-5718-07-01986-2/home.html.

[28] C. Funk, S. Lee, M. R. Oswald, S. Tsogkas, W. Shen, A. Cohen, S. Dickinson, and L. Y. 2017 ICCV challenge: detecting symmetry in the wild. In *ICCV Workshop*, 2017.

[29] D. Grayson and M. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[30] R. I. Hartley. Chirality. *International Journal of Computer Vision*, 26(1):41–61, January 1998. ISSN: 1573-1405.

[31] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997. ISSN: 0162-8828.

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

[33] R. Hartley and H. Li. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2303–2314, 2012. ISSN: 0162-8828. DOI: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2012.43.

[34] D. G. Hook and P. R. McAree. *Using Sturm Sequences to Bracket Real Roots of Polynomial Equations*. eng. Academic Press, January 1990. ISBN: 978-0-12-286165-9.

[35] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *IJCV*, 97(2):123–147, 2012.

[36] A. N. Jensen. Gfan, a software system for Gröbner fans. URL: http://home.imf.au.dk/ajensen/software/gfan/gfan.html.

[37] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *PAMI*, 26:65–81, 2004.

[38] K. Köser, C. Beder, and R. Koch. Conjugate rotation: parameterization and estimation from an affine feature correspondence. In *CVPR*, 2008.

[39] K. Köser and R. Koch. Differential spatial resection - pose estimation using a single local image feature. In *ECCV*, 2008.

[40] Y. Kuang and K. Åström. Numerically stable optimization of polynomial solvers for minimal problems. In *European Conference on Computer Vision (ECCV)*. Springer, 2012.

[41] Y. Kuang and K. Åström. Numerically stable optimization of polynomial solvers for minimal problems. In *ECCV*, 2012.

[42] Z. Kukelova. *Algebraic Methods in Computer Vision*. PhD thesis, Czech Technical University in Prague, 2013.

[43] Z. Kukelova, M. Bujnak, J.Heller, and T. Pajdla. Singly-bordered block-diagonal form for minimal problem solvers. In *Asian Conference on Computer Vision (ACCV 2014)*. (Singapore), 2014.

[44] Z. Kukelova, M. Bujnak, and T. Pajdla. Automatic generator of minimal problem solvers. In *European Conference on Computer Vision (ECCV 2008), Proceedings, Part III*. (Marseille, France), volume 5304 of *Lecture Notes in Computer Science*, 2008. ISBN: 978-3-540-88689-1.

[45] Z. Kukelova, J. Heller, B. M., and T. Pajdla. Radial distortion homography. In *CVPR*, 2015.

[46] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial Eigenvalue Solutions to Minimal Problems in Computer Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1381–1393, July 2012.

[47] Z. Kukelova, J. Kileel, B. Sturmfels, and T. Pajdla. A clever elimination strategy for efficient minimal solvers. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[48] V. Larsson, K. Åström, and M. Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *CVPR*, 2017.

[49] V. Larsson, K. Åström, and M. Oskarsson. Polynomial solvers for saturated ideals. In *ICCV*, 2017.

[50] V. Larsson and K. Åström. Uncovering symmetries in polynomial systems. In *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[51] V. Larsson, K. Åström, and M. Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.

[52] V. Larsson, K. Åström, and M. Oskarsson. Polynomial solvers for saturated ideals. In *International Conference on Computer Vision (ICCV)*, 2017.

[53] V. Larsson, Z. Kukelova, and Y. Zheng. Making minimal solvers for absolute pose estimation compact and robust. In *International Conference on Computer Vision (ICCV)*, 2017.

[54] V. Larsson, M. Oskarsson, K. Astrom, A. Wallis, Z. Kukelova, and T. Pajdla. Beyond grobner bases: basis selection for minimal solvers. In *CVPR*, 2018.

[55] V. Larsson, M. Oskarsson, K. Åström, A. Wallis, Z. Kukelova, and T. Pajdla. Beyond grobner bases: basis selection for minimal solvers. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 3945–3954, 2018. DOI: 10.1109/CVPR.2018.00415. URL: http://openaccess.thecvf.com/content%5C_cvpr%5C_2018/html/Larsson%5C_Beyond%5C_Grobner%5C_Bases%5C_CVPR%5C_2018%5C_paper.html.

[56] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *CVPR*, 1998.

[57] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994. ISBN: 0792394186.

[58] J. Liu, G. Slota, G. Zheng, Z. Wu, M. Park, S. Lee, I. Rauschert, and Y. Liu. Symmetry detection from real world images competition 2013: summary and results. Workshop on Second Symmetry Detection from Real Images Competition:1–6, June 2013.

[59] Y. Liu, H. Hel-Or, C. S. Kaplan, and L. V. Gool. Computational symmetry in computer vision and computer graphics. *Foundations and Trends® in Computer Graphics and Vision*, 5(1–2):1–195, 2010. ISSN: 1572-2740. DOI: 10.1561/0600000008. URL: http://dx.doi.org/10.1561/0600000008.

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

[61] M. Lukáč, D. Sýkora, K. Sunkavalli, E. Shechtman, O. Jamriška, N. Carr, and T. Pajdla. Nautilus: recovering regional symmetry transformations for image editing. *ACM Trans. Graph.*, 36(4):108:1–108:11, July 2017. ISSN: 0730-0301.

[62] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, 2002.

[63] J. Matas, S. Obdržálek, and O. Chum. Local affine frames for wide-baseline stereo. In *ICPR*, 2002.

[64] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10):1615–1630, 2005.

[65] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1):63–86, 2004.

[66] D. Mishkin, F. Radenovic, and J. Matas. Repeatability is not enough: learning affine regions via discriminability. In *ECCV*, 2018.

[67] F. Mora and L. Robbiano. The Gröbner fan of an ideal. English. *Journal of Symbolic Computation*, 6(2-3):183–208, 1988.

[68] O. Naroditsky and K. Daniilidis. Optimizing polynomial solvers for minimal geometry problems. In *International Conference on Computer Vision (ICCV)*. IEEE, 2011.

[69] Š. Obdržálek and J. Matas. Object recognition using local affine frames on distinguished regions. In *BMVC*, 2002.

[70] S. Obdrzálek and J. Matas. Local affine frames for image retrieval. In *Proceedings of the International Conference on Image and Video Retrieval*, CIVR '02, pages 318–327, Berlin, Heidelberg. Springer-Verlag, 2002. ISBN: 3540438998.

[71] T. Ohta, K. Maenobu, and T. Sakai. Obtaining surface orientation from texels under perspective projection. In *IJCAI*, 1981.

[72] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu. Translation-symmetry-based perceptual grouping with applications to urban scenes. In *ACCV*, 2010.

[73] M. Park, K. Brocklehurst, R. Collins, and Y. Liu. Deformed lattice detection in real-world images using mean-shift belief propagation. *PAMI*, 2009.

[74] J. Pritts, O. Chum, and J. Matas. Detection, rectification and segmentation of coplanar repeated patterns. In *CVPR*, 2014.

[75] J. Pritts, Z. Kukelova, V. Larsson, and O. Chum. Radially-distorted conjugate translations. In *CVPR*, 2018.

[76] J. Pritts, Z. Kukelova, V. Larsson, and O. Chum. Rectification from radially-distorted scales. In *ACCV*, 2018.

[77] J. Pritts, D. Rozumnyi, M. P. Kumar, and O. Chum. Coplanar repeats by energy minimization. In *BMVC*, 2016.

[78] J. Pritts, O. Chum, and J. Matas. Approximate models for fast and accurate epipolar geometry estimation. In *IVCNZ*, 2013.

[79] J. Pritts, Z. Kukelova, V. Larsson, Y. Lochman, and O. Chum. Minimal solvers for rectifying from radially-distorted conjugate translations. In 2019. arXiv: 1911.01507 [cs.CV].

[80] J. Pritts, Z. Kukelova, V. Larsson, Y. Lochman, and O. Chum. Minimal solvers for rectifying from radially-distorted scales and change of scales, 2019. arXiv: 1907.11539 [cs.CV].

[81] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, August 2004.

[82] F. Schaffalitzky and A. Zisserman. Geometric grouping of repeated elements within images. In *BMVC*, 1998.

[83] H. Shao, T. Svoboda, and L. Van Gool. Zubud-zurich buildings database for image based recognition. *Computer Vision Lab, Swiss Federal Institute of Technology, witzerland, Tech. Rep*, 260, 2003.

[84] M. Shi, Y. Avrithis, and H. Jégou. Early burst detection for memory-efficient image retrieval. In *CVPR*, 2015.

[85] H. Stewenius. *Gröbner Basis Methods for Minimal Problems in Computer Vision*. eng, volume 2005:1. Centre for Mathematical Sciences, Lund University, 2005. ISBN: 978-91-628-6410-1.

[86] R. Strand and E. Hayman. Correcting radial distortion by circle fitting. In *BMVC*, 2005.

[87] B. Sturmfels. *Gröbner Bases and Convex Polytopes*. University Lecture Series. American Mathematical Society, Providence, RI, USA, 1996.

[88] G. Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138, November 1991. ISSN: 1939-3539. DOI: 10.1109/34.103273.

[89] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *ECCV*, 2008.

[90] M. Van den Bergh, X. Boix, G. Roig, B. de Capitani, and L. Van Gool. Seeds: superpixels extracted via energy-driven sampling. In *ECCV*, 2012.

[91] A. Vedaldi and B. Fulkerson. VLFeat: an open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[92] A. Wang, T. Qiu, and L. Shao. A simple method of radial distortion correction with centre of distortion estimation. *Journal of Mathematical Imaging and Vision*, 35(3):165–172, 2009.

[93] H. Wildenauer and A. Hanbury. Robust camera self-calibration from monocular images of manhattan worlds. In *CVPR*, pages 2831–2838. IEEE, 2012.

[94] H. Wildenauer and B. Micusík. Closed form solution for radial distortion estimation from a single vanishing point. In *BMVC*, 2013.

[95] R. G. Willson and S. Shafer. What is the center of the image? *Journal of the Optical Society of America A*, 11(11):2946–2955, November 1994.

[96] C. Wu, J. M. Frahm, and M. Pollefeys. Repetition-based dense single-view reconstruction. In *CVPR*, 2011.

[97] C. Wu, J. Frahm, and M. Pollefeys. Detecting large repetitive structures with salient boundaries. In *ECCV*, 2010.

[98] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma. TILT: transform invariant low-rank textures. *International Journal of Computer Vision*, 99(1):1–24, 2012.

[99] M. Zuliani, C. Kenney, and M. B. The multiransac algorithm and its application to detect planar homographies. In *ICIP*, 2005.