Minimal Solvers for Rectifying from Radially-Distorted Conjugate Translations

James Pritts, Zuzana Kukelova, Viktor Larsson,Yaroslava Lochman, and Ondřej Chum

Abstract—This paper introduces minimal solvers that jointly solve for radial lens undistortion and affine-rectification using local features extracted from the image of coplanar translated and reflected scene texture, which is common in man-made environments. The proposed solvers accommodate different types of local features and sampling strategies, and three of the proposed variants require just one feature correspondence. State-of-the-art techniques from algebraic geometry are used to simplify the formulation of the solvers. The generated solvers are stable, small and fast. Synthetic and real-image experiments show that the proposed solvers have superior robustness to noise compared to the state of the art. The solvers are integrated with an automated system for rectifying imaged scene planes from coplanar repeated texture. Accurate rectifications on challenging imagery taken with narrow to wide field-of-view lenses demonstrate the applicability of the proposed solvers.

Index Terms—rectification, radial distortion, minimal solvers, symmetry, repeated patterns, local features

◆

1 INTRODUCTION

S CENE-plane rectification is used in many classic computer-vision tasks, including symmetry detection and discovery of near-regular textures [1], [2]; inpainting [3]; single-view 3D reconstruction [4]; using repetitions to improve multi-label segmentation [5], [6], and single-view auto-calibration using the Manhattan scene assumption [7], [8], [9]. 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 [10], [11]. There is only an affine transformation between the affine-rectified imaged scene plane and its real-world counterpart. The removal of the effects of perspective imaging is helpful to understanding the geometry of the scene plane.

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 pinhole camera models [8], [12]. Lens distortion can be estimated by performing a camera calibration apriori, but a fully automated method is desirable. Furthermore, in the case of Internet imagery, the camera and its metadata are often unavailable for use with off-line calibration techniques.

This paper proposes minimal solvers that jointly estimate lens undistortion and affine rectification using point correspondences extracted from the image of coplanar repeated texture that is related by translations or reflections on the scene plane. See

- J. Pritts is with Czech Institute of Informatics, Robotics and Cybernetics (CIIRC) at Czech Technical University in Prague.
  E-mail: prittjam@cvut.cz
- Z. Kukelova and O. Chum are with the Visual Recognition Group (VRG) in the Faculty of Electrical Engineering at Czech Technical University in Prague.
- V. Larsson is with the Computer Vision and Geometry Group (CVG) in the Department of Computer Science at ETH Zurich.
- Y. Lochman is with The Machine Learning Lab at Ukrainian Catholic University in Lviv, Ukraine.
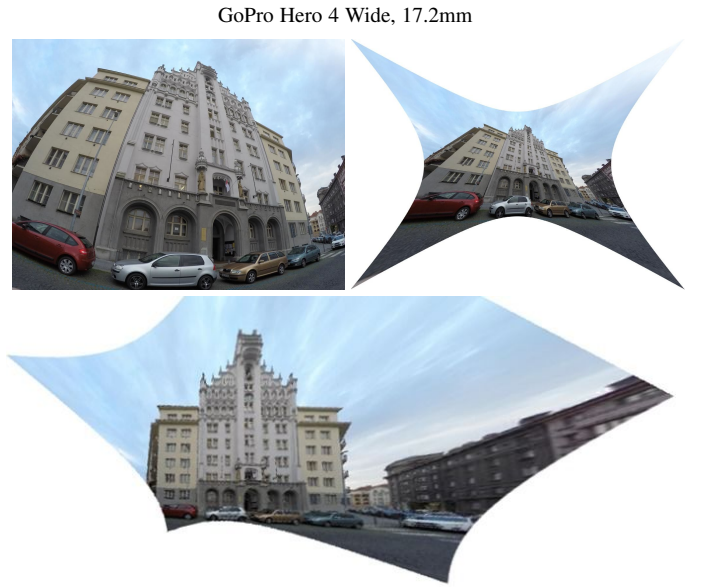
GoPro Hero 4 Wide, 17.2mm

Fig. 1. 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.

Fig. 1 for an example rectification. The proposed solver variants differ by the expected configuration and the number of required point correspondences. The solvers can rectify from distorted local features that are translated on the scene plane in one or two directions, where some of the point correspondences can translate with arbitrary distances. There is also variant that admits reflections. Fig. 3 shows an example of each of the four configurations of point correspondences on the scene plane that are handled by the proposed solvers. Fig. 2 shows these configurations in each stage of the rectification hierarchy and that the solver variants can use these configurations directly affine rectify the distorted image of the scene plane. These configurations often occur in man-made settings, where there are many symmetries. *E.g.*, windows have
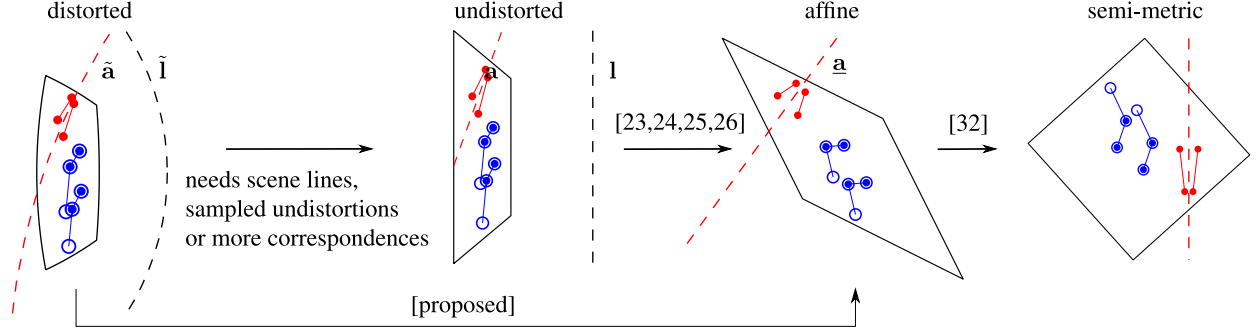
Fig. 2. Direct Affine Rectification. The rectification hierarchy is traversed from left to right. The proposed solvers directly rectify from distorted local features. The state-of-the art requires sampled undistortions, scene lines [8], [9], [13], or more region correspondences [14]. Color denotes how coplanar repeats relate: blue for translation and red for reflection. Marker type denotes the possible correspondence configurations: circles for three translated point correspondences and filled circles for two pairs of two translated point correspondences on a scene plane. Reflections (in red) can be detected as two point correspondences translating over different distances in the same direction. The distorted and undistorted images of the scene plane's vanishing line are denoted $\tilde{l}$ and l, and the distorted and undistorted reflection axis is similarly denoted $\tilde{a}$ and a, where $\underline{a}$ is its rectification. Point correspondences (circles) are extracted from region correspondences (solid polylines), which reduces the required input to one or two region correspondences.

scene structure that is consistent with all four correspondence configurations.

The straightforward application of solver generators to this problem results in unstable solvers [15], [16], [17]. Stable solver generation requires the elimination of some unknowns from the formulations of each variant. The generated solvers differ significantly in time-to-solution and noise sensitivity with respect to the choice of eliminated unknowns. These different formulations are derived and analyzed. In particular, an exceptionally fast and robust solver is generated compared to the solvers introduced in [18] for feature configuration (a) of Fig. 3. All solvers are optimized in C++ and are available at https://github.com/prittjam/repeats.
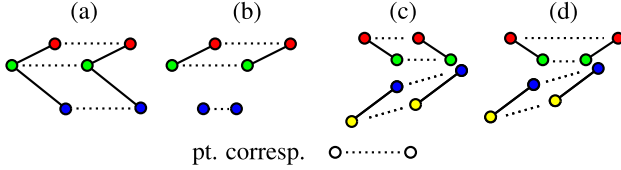

Fig. 3. The proposed solver variants can use: (a) three points translated over the same distance in one-direction, (b) three points translated in one direction with one point (blue) at an arbitrary distance, (c) two point pairs translated over two distances in two directions, and (d) two point pairs translated in two directions with one pair translating the same distance and one pair at different distances.

Automatic scene-plane rectification is an ill-posed single-view geometry estimation task that is further complicated if there is lens distortion. Robust sampling schemes like RANSAC are required to obtain high-accuracy rectifications from noisy or bad local feature correspondences [5], [19], [20], [21]. Reducing the cardinality of the minimal sample set can greatly reduce the number of trials needed by RANSAC to find a good solution [10]. Region correspondences can be used to simultaneously establish multiple point correspondences, which reduces the the minimal sample size of point-based solvers [22], [23], [24], [25]. The proposed solvers are designed such that this substitution is straightforward. The solid polylines connecting points in Fig. 3 denote the cases in which a region correspondence can replace multiple point correspondences. The proposed solvers require either one or two region correspondences, have fast times-to-solution, and are robust to noisy features, which makes them well suited for use in a RANSAC-based estimator.

The RANSAC-based metric-rectification framework of [11] is extended to use the proposed minimal solvers to synthesize the fronto-parallel images presented in this paper. The extension includes a method to use extra available point correspondence constructions from a region correspondence to improve rectification accuracy. The correspondences are partitioned into complementary subsets of minimal samples and unused correspondences, and cross validation is used to select the best rectification. In addition, a method for reliably finding affine-rectified regions that can be used for the estimation of the metric upgrade is proposed.

1.1 Previous Work

The problem of rectification is closely coupled with the detection of coplanar repeats in a classic chicken-and-egg scenario: rectification is easy if the repeats are grouped, and repeats are more easily grouped if the affine invariants of the rectified plane are available [11]. Most local-feature based state-of-the-art methods tentatively group coplanar repeats by their texture, which are verified by testing affine or metric invariants with a hypothesized rectification [3], [20], [21], [27], [28], [29], [30], [31]. These methods assume a pinhole camera model, which makes them of limited use on lens-distorted images. In particular, the method of Schaffalitzky et al. [20] is the most similar to the solvers proposed in this paper since it also uses constraints induced by imaged translational symmetries. A survey of minimal solvers that rectify using imaged coplanar repeats is given in Table 3.

The state-of-the-art has extended minimal rectifying solvers to lens-distorted images [8], [9], [14], [26]. The common choice of parameterization for lens undistortion in ill-posed settings is the division model since it has only one parameter and can effectively model a wide range of radial lens undistortions [32]. The proposed solvers also use the division model. The capabilities of the proposed versus state-of-the-art solvers for affine rectification are illustrated in Fig. 2.

Pritts et al. [14], [26] introduced solvers that rectify from the distorted image of rigidly-transformed coplanar repeats. However, these solvers are about 2000 times slower than the fastest of the proposed solvers, require three region correspondences, and cannot use individual point correspondences (they are strictly region based). Wildenauer et al. [8] and Antunes et al. [9] are two contour-based methods that rectify from the distorted images

TABLE 1
Scene Assumptions

	Wildenauer et al. [8]	Antunes et al. [9]	Pritts et al. [14], [26]	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

Solvers rectifying from distorted images by [8], [9] require distinct sets of parallel scene lines as input. Pritts et al. [14], [26] requires three region correspondences. The proposed solvers can rectify from only one region correspondence and admit point correspondences.

of parallel scene lines. These methods require two sets of imaged parallel lines, which must have different mutual orientations. In particular, the solver [8] requires five lines total and [9] requires seven. These are strong scene content assumptions. In addition, unlike points or regions, contours cannot be reliably corresponded based on appearance. These rectifying solvers that use the division model of lens undistortion are compared with the proposed solvers in Table 1.

The method of Li et al. [33] is a two-view method that jointly estimates undistortion and the fundamental matrix. Both Li et al. and the proposed solvers construct a matrix from the unknown lens undistortion parameters and use the fact that it must be singular to impose constraints on the unknowns.

The method of Pritts et al. [11] uses a non-linear refinement step to generalize a rectification estimated under the pinhole camera assumption to include lens undistortion. However, even with relaxed thresholds, a robust estimator like RANSAC [19] discards measurements around the boundary of the image since this region is the most affected by radial distortion and cannot be accurately modeled with a pinhole camera. Neglecting lens distortion during the labeling of good and bad measurements, as done during the verification step of RANSAC, can give fits that are biased [12], which degrades rectification accuracy.

2 PRELIMINARIES

Without loss of generality, coplanar scene points are assumed to be on the scene plane $z = 0$. This permits the camera matrix 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

$$\underbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix}}_{P^{3\times4}} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix}}_{P} \underbrace{\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}}_{\mathbf{X}}, \quad (1)$$

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

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 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 $\mathbf{v}$ and the vanishing translation directions of scene-plane translations $\mathbf{U}$ or $\mathbf{V}$ as imaged by P, respectively. Matrices are in typewriter font; *e.g.*, a homography is H and a conjugate

TABLE 2
Common Denotations

Term	Description
P	3×3 camera matrix viewing $z = 0$ (see (1)).
$\mathbf{X}$	homogeneous scene point in $\mathbb{RP}^2$
$\mathbf{x}, \tilde{\mathbf{x}}$	homogeneous pinhole and distorted image point
$\underline{\mathbf{x}}$	affine-rectified point (see (6))
$\mathbf{x} \leftrightarrow \mathbf{x}'$	$\mathbf{x},\mathbf{x}'$ 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 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
T	homogeneous translation matrix
$\mathbf{l}, \tilde{\mathbf{l}}$	image of vanishing line and distorted vanishing line
$\mathbf{l}_\infty$	the line at infinity
H	affine-rectifying homography
$H_{\mathbf{u}}$	conjugate translation in the imaged trans. direction $\mathbf{u}$
λ	division model parameter for undistortion (see Sec. 3.1)
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

translation (also a homography) with vanishing point $\mathbf{u}$ is denoted $H_{\mathbf{u}}$ (see Sec. 4).

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. A covariant region detection (see Sec. 7.1) is a distorted function of some region from the pinhole image and is denoted $\tilde{\mathcal{R}}$. Likewise, a distorted point extracted from a region detection is denoted $\tilde{\mathbf{x}} = \left(\tilde{x}, \tilde{y}, 1\right)^\top$. The affine-rectified images of homogeneous points and regions are denoted as $\underline{\mathbf{x}} = \left(\underline{x}, \underline{y}, 1\right)^\top$ and $\underline{\mathcal{R}}$, respectively. Table 2 summarizes this notation.

2.1 The Hidden Variable Trick

Each of the proposed solvers uses the hidden variable trick to transform its polynomial constraint equations into a tractable form. The hidden variable trick is a resultant technique in algebraic geometry that is used to eliminate subsets of variables from multivariate polynomial systems of equations [34]. Suppose that a multivariate polynomial system of m equations in n unknowns is given. The hidden variable trick works by assuming that a set ξ_i of $k < n$ unknowns are parameters that are used to construct a coefficient matrix $M(\xi_1, \ldots, \xi_k) \in \mathbb{R}^{m \times l}$, such that the system can be rewritten as

$$M(\xi_1, \ldots, \xi_k)\mathbf{y} = \mathbf{0}, \quad (2)$$

where $\mathbf{y} \in \mathbb{R}^l$ is a vector of l monomials in the remaining $n - k$ unknowns (*i.e.*, monomials of unknowns not appearing in M). A nontrivial solution to (2) exists only if M is rank-deficient. The problem has been simplified since the $n - k$ unknowns in $\mathbf{y}$ are eliminated from solving $\det M = 0$. In the case where $k > 1$, the

(a) GoPro Hero 4 Medium, 21.9mm   (b) GoPro Hero 4 Wide, 17.2mm   (c) Samyang, 7.5mm

Fig. 4. 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.

$l \times l$ minors of M can be used to generate the necessary number of polynomial constraint equations to solve for the unknowns $\{\xi_1, \ldots, \xi_k\}$. Once ξ_i are recovered, the original system of equations (2) can be solved for $\mathbf{y}$ by back substitution.

2.2 Solving Systems of Polynomial Equations

The polynomial systems of equations encoding the rectifying constraints for the Eliminated Vanishing Point (EVP) solvers are solved using an algebraic method based on Gröbner bases. Automated solver generators using the Gröbner basis method [15], [16] have been used to generate solvers for several camera geometry estimation problems (see e.g. [12], [14], [15], [16], [18], [26], [35]). However, the straightforward application of automated solver generators to the proposed constraints resulted in unstable solvers. In [35], Larsson et al. proposed a method for creating polynomial solvers for problems with unwanted solutions using ideal saturation. We use the hidden variable trick together with ideal saturation to eliminate unknowns from the polynomial system of equations arising in the formulations of the EVP solvers. This results in significantly more numerically stable solvers compared to the solvers generated directly from the original constraint equations. For more details about Gröbner bases and how they are used in polynomial solvers we refer the interested reader to [15], [16], [34], [35].

2.3 Solver Naming Convention

We apply the solver naming convention of Pritts et al. [14], [26] to the proposed and state-of-the-art solvers evaluated in this paper.

The minimal configuration of region correspondences is given as the subscript to H (denoting a homography); e.g., a solver requiring 3 affine-covariant region correspondences is denoted H_{222}. The unknowns that are recovered by the solver are suffixed to $\mathrm{H}_{\cdot}$, e.g., the proposed solver requiring one region correspondence and returning the vanishing line $\mathbf{l}$ and division model parameter λ of lens distortion is denoted $\mathrm{H}_2\mathbf{l}\lambda$.

3 PROBLEM FORMULATION

An affine-rectifying homography 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 $\mathbf{l}_\infty = \left(0, 0, 1\right)^\top$ [10]. Thus any homography H satisfying the constraint

$$\eta \mathbf{l} = \mathrm{H}^\top \mathbf{l}_\infty = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \eta \neq 0, \qquad (3)$$

and where $\mathbf{l}$ is an imaged scene plane's vanishing line, is an affine-rectifying homography. Constraint (3) 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 H. Thus, assuming $l_3 \neq 0$ [10], the affine-rectification of image point $\mathbf{x}$ to the affine-rectified point $\underline{\mathbf{x}}$ can be defined as

$$\alpha \underline{\mathbf{x}} = \left(\alpha \underline{x}, \alpha y, \alpha\right)^\top = \mathrm{H}(\mathbf{l})\mathbf{x}$$

$$\text{s.t.} \quad \mathrm{H}(\mathbf{l}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ & \mathbf{l}^\top & \end{bmatrix} \quad \text{and} \quad \alpha \neq 0. \qquad (4)$$

3.1 Radial Lens Undistortion

Affine rectification as given in (4) is valid only if $\mathbf{x}$ is imaged by a pinhole camera. Cameras always have some lens distortion, and the distortion can be significant for wide-angle lenses. For a lens distorted point, denoted $\tilde{\mathbf{x}}$, an undistortion function f is needed to transform $\tilde{\mathbf{x}}$ to the pinhole point $\mathbf{x}$. We use the one-parameter division model to parameterize the radial lens undistortion function,

$$\gamma\mathbf{x} = f(\tilde{\mathbf{x}}, \lambda) = \left(\tilde{x}, \tilde{y}, 1 + \lambda(\tilde{x}^2 + \tilde{y}^2)\right)^\top \quad (5)$$

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 [32] for the joint estimation of two-view geometry and non-linear lens distortion. The division model is especially suited for minimal solvers since it is able to express a wide range of distortions (*e.g.*, see second row of Fig. 4) 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. [36] and Fitzgibbon [32] showed that the precise positioning of the distortion center does not strongly affect image correction. Furthermore, we will see in the experiments in Sec. 8 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 [36]. 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 is discussed in detail in Sec. 6.2.

Affine rectified points $\underline{\mathbf{x}}_i$ can be expressed in terms of distorted points $\tilde{\mathbf{x}}_i$ by substituting (5) into (4), which gives

$$\alpha\underline{\mathbf{x}} = \left(\alpha\underline{x}, \alpha\underline{y}, \alpha\right)^\top = \mathtt{H}(\mathbf{l})f(\tilde{\mathbf{x}}, \lambda) = \\ \left(\tilde{x}, \tilde{y}, l_1\tilde{x} + l_2\tilde{y} + l_3(1 + \lambda(\tilde{x}^2 + \tilde{y}^2))\right)^\top . \quad (6)$$

Interestingly, the rectifying function $\mathtt{H}(\mathbf{l})f(\tilde{\mathbf{x}}, \lambda)$ in (6) also acts radially about the distortion center, but unlike the division model in (5), it is not rotationally symmetric.

The distortion function of the lens as parameterized by the division model is denoted $f^d(\cdot, \lambda)$. Under the division model, the radially-distorted image of the vanishing line is a circle and is denoted $\tilde{\mathbf{l}}$ [13], [32], [37], [38]. Figs. 4 and 7 render the distorted vanishing line in the source images, which affirm the accuracy of the rectifications by the proposed solvers.

3.2 Covariant Region Parameterization

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. The geometry of an affine-covariant region $\mathcal{R}$ is given by a right-handed affine basis in the image coordinate system called a *local affine frame* (LAF). The affine frame is minimally parameterized by three points $\{\mathbf{o}, \mathbf{x}, \mathbf{y}\}$. For similarity-covariant regions, there is the additional constraint that $\mathbf{x} - \mathbf{o} \perp \mathbf{y} - \mathbf{o}$ (see [39]). This

construction is also referred to as an *oriented circle*, where $\mathbf{o}$ is the origin of the circle and $\mathbf{x}$ defines the circle's orientation and radius. Similarity-covariant regions are minimally parameterized by two points. Examples of both frame constructions are shown in Figs. 2, 3, and 6, and an example of affine frames constructed from the combined methods of [40], [41], [42] are shown in Fig. 8.

4 CONJUGATE TRANSLATIONS

Assume that the scene plane Π and a camera's image plane π are related point-wise by the camera $\mathtt{P}$ 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 $\mathtt{T}(\mathbf{U})$, the points $\mathbf{X}$ and $\mathbf{X}'$ as imaged by camera $\mathtt{P}$ can be expressed as

$$\alpha\mathbf{x}' = \mathtt{P}\mathbf{X}' = \mathtt{P}\mathtt{T}(\mathbf{U})\mathbf{X} = \mathtt{P}\mathtt{T}(\mathbf{U})\mathtt{P}^{-1}\mathbf{x} = \mathtt{H}_{\mathbf{u}}\mathbf{x}$$
$$\text{s.t.} \quad \mathtt{T}(\mathbf{U}) = \begin{pmatrix} 1 & 0 & u_x \\ 0 & 1 & u_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (7)$$

where the homography $\mathtt{H}_{\mathbf{u}} = \mathtt{P}\mathtt{T}(\mathbf{U})\mathtt{P}^{-1}$ 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 $\mathtt{H}_{\mathbf{u}}$ [10], [20].

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

$$\alpha\mathbf{x}' = \mathtt{H}_{\mathbf{u}}\mathbf{x} = \left[\mathtt{P}\mathtt{I}_3\mathtt{P}^{-1} + \mathtt{P}\begin{pmatrix} u_x \\ u_y \\ 0 \end{pmatrix}\left[\mathtt{P}^{-\top}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right]^\top\right]\mathbf{x}$$
$$= [\mathtt{I}_3 + s^{\mathbf{u}}\mathbf{u}\mathbf{l}^\top]\mathbf{x} \quad (8)$$

where $\mathtt{I}_3$ is the 3×3 identity matrix, and, also consulting Fig. 5 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 $\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ [20].

4.1 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

$$\alpha\mathbf{m}_i = \alpha\left(\mathbf{x}_i \times \mathbf{x}'_i\right) = (\mathtt{P}\mathbf{X}_i \times \mathtt{P}\mathbf{X}'_i)/|\mathtt{P}| = \\ (\mathtt{P}\mathbf{X}_i \times \mathtt{P}(\mathbf{X}_i + \mathbf{U}))/|\mathtt{P}| = \mathtt{P}^{-\top}(\mathbf{X}_i + \mathbf{U}), \quad (9)$$

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

Using (9) to express the meet of joins $\mathbf{m}_i$ and $\mathbf{m}_j$ in terms of the camera $\mathtt{P}$ and joined scene point correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ and $\mathbf{X}_j \leftrightarrow \mathbf{X}'_j$ gives

$$\alpha_i\mathbf{m}_i \times \alpha_j\mathbf{m}_j = \left(\mathtt{P}^{-\top}(\mathbf{X}_i + \mathbf{U})\right) \times \left(\mathtt{P}^{-\top}(\mathbf{X}_j + \mathbf{U})\right) = \\ \mathtt{P}((\mathbf{X}_i + \mathbf{U}) \times (\mathbf{X}_j + \mathbf{U}))/|\mathtt{P}| = \\ \mathtt{P}\left(\mathbf{U}^\top(\mathbf{X}_i \times \mathbf{X}_j)\right)\mathbf{U}/|\mathtt{P}| = \beta\mathtt{P}\mathbf{U} = \eta\mathbf{u}, \quad (10)$$

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 Fig. 5).
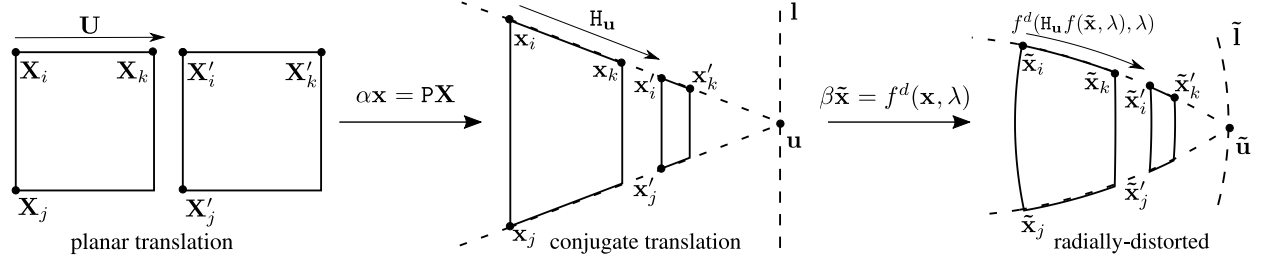
Fig. 5. The Geometry of a Radially-Distorted Conjugate Translations. A translation of coplanar scene points $\{\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k\}$ by $\mathbf{U}$ induces a conjugate translation $\mathtt{H_u}$ in the undistorted image as viewed by camera $\mathtt{P}$. Joined conjugately-translated 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. 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 $f^d(\mathtt{H_u} f(\tilde{\mathbf{x}}, \lambda), \lambda)$, where $f^d(\cdot, \lambda)$ is the division-model distortion function.

In general (10) 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 Fig. 5 were used in lieu of $\mathbf{x}_j \leftrightarrow \mathbf{x}_j'$ in (10), 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 Sec. 6.

Since $\mathbf{U}$ is coincident with $\mathbf{l}_\infty$ by construction (see Fig. 5) and point-line incidence is invariant under projection by $\mathtt{P}$ [10], $\mathbf{u}$ and $\mathbf{l}$ are also coincident,

$$\mathbf{l}^\top\mathbf{u} = 0. \tag{11}$$

The EVL solver introduced in 5.2 uses the relation between conjugately-translated points and vanishing points derived in (9) and (10) and the vanishing point-vanishing line incidence equation of (11) to place constraints on $\mathbf{l}$.

4.2 Radially-Distorted Conjugate Translations

Conjugate translations as defined in (8) can be written in terms of radially-distorted conjugately-translated point correspondences undistorted by (5) as

$$\alpha f(\tilde{\mathbf{x}}', \lambda) = \mathtt{H_u} f(\tilde{\mathbf{x}}, \lambda) = [\mathtt{I}_3 + s^\mathbf{u}\mathbf{u}\mathbf{l}^\top]f(\tilde{\mathbf{x}}, \lambda), \tag{12}$$

$\tilde{\mathbf{x}} \leftrightarrow \tilde{\mathbf{x}}'$ is a radially-distorted point correspondence that is consistent with the conjugate translation $\mathtt{H_u}$. 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 Sec. 5.1 uses the relation defined in (12) and the vanishing point-vanishing line incidence equation of (11) to place constraints on $\mathbf{l}$ and λ.

5 SOLVERS

This section introduces 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. The designs of the solver variants are motivated by the types of covariant feature detectors that can be used to extract point correspondences, which give the constraints needed to jointly solve for the division model parameter of lens undistortion, vanishing line and the vanishing point of the translation direction(s).

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, and (ii) a radially-distorted conjugate translation is a transformation with exactly four degrees of freedom.

The proposed solvers can also be differentiated by the choice to use the hidden variable trick 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 [34]. The solver groups are eponymously named after their eliminated unknowns.

The group of Eliminated Vanishing Point (EVP) solvers hide the lens undistortion parameter and vanishing line parameters and have the vanishing point eliminated. They 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 (see Fig. 3). In addition, there is an EVP variant that admits reflections.

The Eliminated Vanishing Line (EVL) solver hides the lens undistortion parameter and eliminates the vanishing line parameters (the vanishing points are recovered by construction). The EVL solver jointly recovers undistortion and rectification from radially-distorted conjugately-translated point correspondences in one direction. The elimination of the vanishing line results in a solver that is exceptionally fast, stable and robust to feature noise.

Table 3 summarizes the geometric assumptions, inputs and complexity of the proposed solvers with respect to the state of the art. From the undistorting solvers, the proposed solvers require the fewest correspondences, and are much simpler than the undistorting and rectifying solver of [14].

The following sections show significant differences between the two groups with respect to solver complexity, time to solution, stability and noise sensitivity. The following sections describe how the solvers are generated, and, in particular, 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.

5.1 The Eliminated Vanishing Point (EVP) Solvers

The model for radially-distorted conjugate translations in (12) defines the unknown geometric quantities: (i) division–model parameter λ, (ii) imaged scene-plane vanishing line $\mathbf{l} = \begin{pmatrix} l_1, l_2, l_3 \end{pmatrix}^\top$, (iii) vanishing point of the translation direction

TABLE 3
Proposed Solvers (shaded in grey) vs. State of the Art

	Reference	Rectifies	Undistorts	Motion	# Correspondences Regions	# Correspondences Points	# Solutions	Size
H_2l	[20]	✓		translation	1	2	1	closed form
$H_{22}l$	[21]	✓		rigid	2	N/A	1	closed form
$H_2l\lambda$		✓	✓	translation	1	3	4	closed form
$H_2lu\lambda$		✓	✓	translation	1	3	4	14×18
$H_2lus^{\mathbf{u}}\lambda$		✓	✓	translation	1	3	2	24×26
$H_{22}luv\lambda$		✓	✓	translation	2	4	6	54×60
$H_{22}luvs^{\mathbf{v}}\lambda$		✓	✓	translation	2	4	4	76×80
$H_{22}\lambda$	[32]		✓	rigid[1]	2	5	18	18×18
$H_{22}\lambda_1\lambda_2$	[12]		✓	rigid[1]	2	5	5	16×21
$H_{222}^{DES}l\lambda$	[14], [26]	✓	✓	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.

The proposed solvers require a few as 1 region correspondence instead of three and are significantly simpler than the undistorting and rectifying solver $H_{222}^{DES}l\lambda$ of [14], [26]. The homography solvers of [12], [32] 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.

$\mathbf{u} = (u_1, u_2, u_3)^\top$, (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 l is constrained to the affine subspace $l_3 = 1$ of the real-projective plane, which makes it unique. This inhomogeneous choice of 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 l. See Sec. 6 for a more detailed discussion of the degeneracies.

The vanishing direction $\mathbf{u}$ must meet the vanishing line l, which defines a subspace of solutions for $\mathbf{u}$. The magnitude of $\mathbf{u}$ is set to the magnitude of conjugate translation $s_1^{\mathbf{u}}$ of the first correspondence $\tilde{\mathbf{x}}_1 \leftrightarrow \tilde{\mathbf{x}}_1'$, which defines a unique solution

$$l^\top \mathbf{u} = l_1 u_1 + l_2 u_2 + u_3 = 0 \quad \wedge \quad \|\mathbf{u}\| = s_1^{\mathbf{u}}. \quad (13)$$

The relative scale of translation $\bar{s}_i^{\mathbf{u}}$ for each correspondence $\tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{x}}_i'$ with respect to the magnitude of $\|\mathbf{u}\|$ is defined so that $\bar{s}_i^{\mathbf{u}} = s_i^{\mathbf{u}}/\|\mathbf{u}\|$. Note that $\bar{s}_1^{\mathbf{u}} = 1$. The relationship between magnitude of translation in the scene plane and the magnitude of conjugate translation is derived in the Sec. 8.1.2 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. Solver $H_2lu\lambda$ assumes that all point correspondences have the same relative scales of translation, i.e. $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = \bar{s}_3^{\mathbf{u}} = 1$. Solver $H_2lus^{\mathbf{u}}\lambda$ relaxes the equal relative scale of translation assumption of the $H_2lu\lambda$ solver. In particular, solver $H_2lus^{\mathbf{u}}\lambda$ assumes that two of the point correspondences have the same magnitude of conjugate translation (i.e. $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = 1$), and the third point correspondence has an unknown relative scale of the translation $\bar{s}_3^{\mathbf{u}}$. The $H_2lus^{\mathbf{u}}\lambda$ 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 and thus

provide the necessary 4 point correspondences.

Solver $H_{22}luv\lambda$ 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 $H_{22}luvs^{\mathbf{v}}\lambda$ requires four point correspondences, equivalently, two similarity-covariant region correspondences, and relaxes the assumption of the $H_{22}luv\lambda$ solver that both point correspondences in the $\mathbf{v}$ direction have the same magnitudes of conjugate translation. In particular, $H_{22}luvs^{\mathbf{v}}\lambda$ 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 Fig. 3).

In all of the proposed solvers the scalar values α_i are eliminated from (12). This is done by multiplying (12) 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,

$$\begin{bmatrix} 0 & -\tilde{w}_i' & \tilde{y}_i' \\ \tilde{w}_i' & 0 & -\tilde{x}_i' \\ -\tilde{y}_i' & \tilde{x}_i' & 0 \end{bmatrix} \begin{bmatrix} 1 + \bar{s}_i^{\mathbf{u}} u_1 l_1 & \bar{s}_i^{\mathbf{u}} u_1 l_2 & \bar{s}_i^{\mathbf{u}} u_1 \\ \bar{s}_i^{\mathbf{u}} u_2 l_1 & 1 + \bar{s}_i^{\mathbf{u}} u_2 l_2 & \bar{s}_i^{\mathbf{u}} u_2 \\ \bar{s}_i^{\mathbf{u}} u_3 l_1 & \bar{s}_i^{\mathbf{u}} u_3 l_2 & 1 + \bar{s}_i^{\mathbf{u}} u_3 \end{bmatrix} \begin{pmatrix} \tilde{x}_i \\ \tilde{y}_i \\ \tilde{w}_i \end{pmatrix} = \mathbf{0}, \quad (14)$$

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 (14) 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 [34]. In particular, we used the automatic generators proposed in [15], [16]; 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 [15], [16], 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 [15], [16] 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 $H_2 \mathbf{lu}\lambda$, $H_2 \mathbf{lu}s^{\mathbf{u}}\lambda$, $H_{22}\mathbf{luv}\lambda$ and $H_{22}\mathbf{luv}s^{\mathbf{v}}\lambda$ problems, respectively. Therefore, we use the hidden-variable trick to eliminate the vanishing translation directions together with ideal saturation to eliminate parasitic solutions [34], [35]. 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 3, and wall clock timings for the simplified solvers are reported in Sec. 8.2. Optimized C++ implementations for all the proposed solvers are provided.

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

5.1.1 One-Direction EVP Solvers

For the one-direction $H_2 \mathbf{lu}s^{\mathbf{u}}\lambda$ solver we have $\bar{s}_1^{\mathbf{u}} = \bar{s}_2^{\mathbf{u}} = 1$. Therefore the constraints (14) 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 (13). 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

$$\mathtt{M}(l_1, l_2, \bar{s}_3^{\mathbf{u}}, \lambda) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{pmatrix} = \mathbf{0}, \qquad (15)$$

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 (15) introduces a one-dimensional family of false solutions. These are not present in the original system and correspond 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 of the same form as in (15).

These false solutions can be removed by saturating [35] any of the 3×3 cofactors from the first three columns of $\mathtt{M}$. The matrix $\mathtt{M}$ has the following form,

$$\mathtt{M}(l_1, l_2, \bar{s}_3^{\mathbf{u}}, \lambda) = \begin{bmatrix} m_{11} & m_{12} & 0 & m_{14} \\ m_{21} & m_{22} & 0 & m_{24} \\ m_{31} & 0 & m_{33} & m_{34} \\ m_{41} & 0 & m_{43} & m_{44} \\ m_{51} & m_{52} & 0 & m_{54} \\ m_{61} & 0 & m_{63} & m_{64} \\ l_1 & l_2 & 1 & 0 \end{bmatrix}, \qquad (16)$$

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. [35], 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 $H_2 \mathbf{lu}\lambda$ 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 (15) 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 (14), e.g., for $i = 3$, and the matrix $\mathtt{M}$ in (15) 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 (16) 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.

5.1.2 Two-Direction EVP Solvers

In the case of the two-direction $H_{22}\mathbf{luv}s^{\mathbf{v}}\lambda$ solver, the input equations for two vanishing translation directions $\mathbf{u} = (u_1, u_2, u_3)^{\top}$ and $\mathbf{v} = (v_1, v_2, v_3)^{\top}$ 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 (13), $i.e.$, 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

$$\mathtt{M}_1(l_1, l_2, \lambda) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{pmatrix} = \mathbf{0}, \quad \mathtt{M}_2(l_1, l_2, \bar{s}_4^{\mathbf{v}}, \lambda) \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ 1 \end{pmatrix} = \mathbf{0}, \quad (17)$$

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 (16) 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 $H_{22}\mathbf{luv}\lambda$ 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 (14), e.g., for $i = 4$. Therefore, the matrix $\mathtt{M}_2$ from (17) 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 [43] revealed that for this particular formulation, it is sufficient to only saturate the top-left 2×2 cofactor of $\mathtt{M}_1$
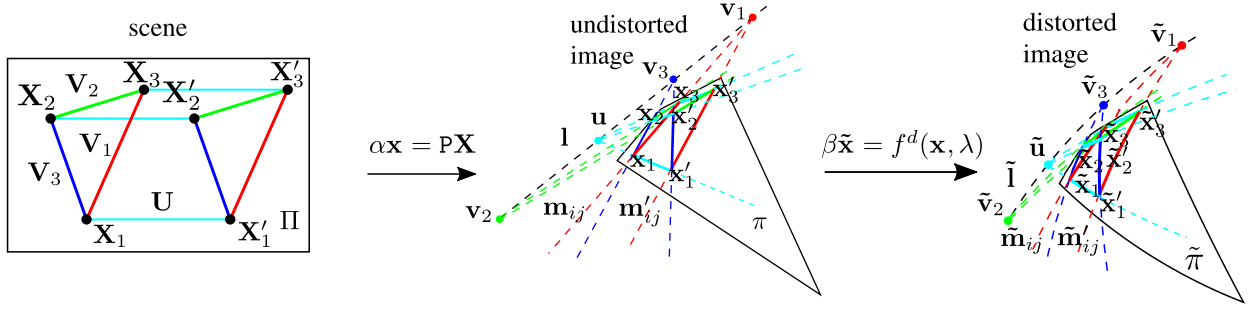
Fig. 6. The Geometry of the EVL Constraints. The scene plane Π contains the preimage of radially-distorted conjugately-translated affine-covariant regions, equivalently, 3 translated points in the direction $\mathbf{U}$. This configuration had 3 additional translation directions $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ 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 in the EVL formulation.

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.

5.2 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 Fig. 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 Fig. 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 of pairs of conjugate-translations that share the same translation direction in the scene plane, *e.g.*,

$$\gamma \mathbf{v}_1 = (\mathbf{x}_1 \times \mathbf{x}_3) \times (\mathbf{x}_1' \times \mathbf{x}_3'). \tag{18}$$

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 Sec. 4.1, 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 $\mathbf{l}$.

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 (11)

$$\mathbf{v}_i^\top \mathbf{l} = \left((\mathbf{x}_i \times \mathbf{x}_j) \times (\mathbf{x}_i' \times \mathbf{x}_j') \right)^\top \mathbf{l} = 0, \tag{19}$$

where $i < j$ and $i, j \in \{\, 1 \dots 3 \,\}$. The homogeneity of (19) is used to eliminate any non-zero scalars. Substituting radially-distorted points for undistorted points in (19) using (5) gives

$$\left(f(\tilde{\mathbf{x}}_i, \lambda) \times f(\tilde{\mathbf{x}}_j, \lambda) \right) \times \left(f(\tilde{\mathbf{x}}_i', \lambda) \times f(\tilde{\mathbf{x}}_j', \lambda) \right)^\top \mathbf{l} = 0. \tag{20}$$

The skew-symmetric operator, denoted $[\cdot]_\times$, is used to transform (20) into the homogeneous matrix-vector equation

$$\left([[f(\tilde{\mathbf{x}}_i, \lambda)]_\times f(\tilde{\mathbf{x}}_j, \lambda)]_\times [f(\tilde{\mathbf{x}}_i', \lambda)]_\times f(\tilde{\mathbf{x}}_j', \lambda) \right)^\top \mathbf{l} = 0, \tag{21}$$

where where $i < j$ and $i, j \in \{\, 1 \dots 3 \,\}$. Independent scalar constraint equations of the form (21) can be stacked to add the necessary number of constraints for jointly estimating $\mathbf{l}$ and λ.

5.2.1 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 (21), namely $\mathbf{l} = (l_1, l_2, l_3)^\top$ and the division model parameter λ. The vanishing line $\mathbf{l}$ is homogeneous, so it has only two degrees of freedom. Thus 3 scalar constraint equations of the form (21) 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 (21), can be concisely encoded in the matrix $\mathtt{M}(\lambda) \in \mathbb{R}^{3 \times 3}$ as

$$\mathtt{M}(\lambda) \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \mathbf{0}. \tag{22}$$

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 $i, j, k \in \{\, 1 \dots 3 \,\}$ 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 Sec. 5.2.2.

The division model parameter λ is hidden in (22) using the hidden-variable trick in the entries of coefficient matrix $\mathtt{M}$, which are polynomials only in λ [34]. 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 $\mathtt{H}_2 \mathbf{l} \lambda$.

5.2.2 Best Minimal Solution Selection

The EVL geometry of Fig. 6 has 10 meets that can be used to generate scalar constraint equations in (21). However, only 3 meets are needed to jointly estimate $\mathbf{l}$ and λ. Since the time to solution
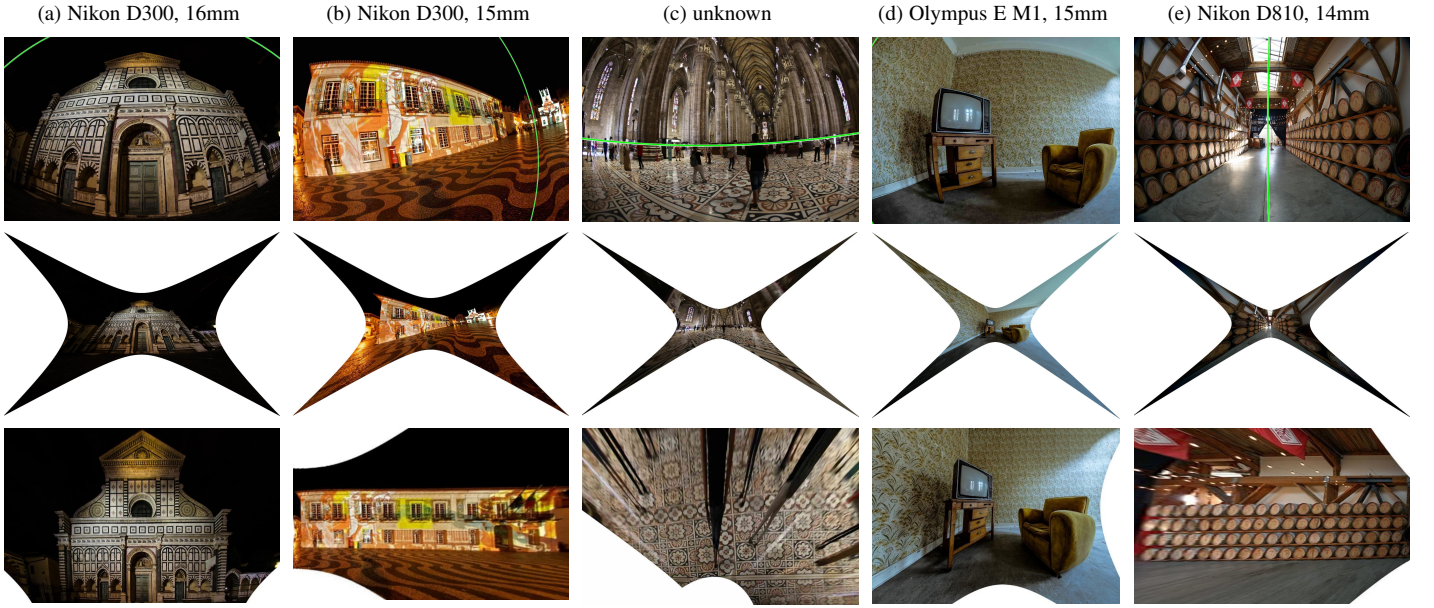
Fig. 7. EVL Solver Results on Fisheye Images. The distorted image of the vanishing line is rendered in green in the input images on the top row. Results were produced using the $\text{H}_2\text{l}\lambda$ with 1-correspondence sampling in a RANSAC framework. The $\text{H}_2\text{l}\lambda$ solver runs 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.

for the $\text{H}_2\text{l}\lambda$ is only $\mathbf{0.5}$ **µs**, 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

$$\sum_i d(\tilde{\mathbf{x}}_i, f^d(\text{H}^{-1}f(\tilde{\mathbf{x}}'_i, \lambda), \lambda))^2 + d(f^d(\text{H}f(\tilde{\mathbf{x}}_i, \lambda), \lambda), \tilde{\mathbf{x}}'_i)^2, \quad (23)$$

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 (23) has several benefits: (i) Near degenerate correspondence configurations can be rejected, (ii) Correspondences with geometric properties that are more robust to noise will be preferred, *e.g.*, regions that are further apart, (iii) 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 (23).

Best minimal solution selection is evaluated in the sensitivity studies in Sec. 8. The solver incorporating best minimal solution selection is denoted in the standard way, $\text{H}_2\text{l}\lambda$. For comparison we introduce a baseline solver, denoted $\text{H}_2^{\text{RND}}\text{l}\lambda$, which randomly selects from the 10 possible constraint configurations associated with the EVL geometry (see Fig. 6). As expected, the $\text{H}_2\text{l}\lambda$ performs better than $\text{H}_2^{\text{RND}}\text{l}\lambda$ on all sensitivity measures.

5.2.3 Optimal Estimate of the Vanishing Point

Unlike the EVP solvers in Sec. 5.1, which jointly estimate the vanishing point $\mathbf{u}$ (shown in Fig. 6) using all constraints from the set of conjugate translations $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}_{i=1}^3$ (see (14)), the $\text{H}_2\text{l}\lambda$ 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 Fig. 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 $\text{H}_2\text{l}\lambda$) by solving a constrained least squares system that includes all constraints induced by $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}_{i=1}^3$ (see Fig. 6). The incidence of $\mathbf{u}$ with $\mathbf{l}$ is explicitly enforced by including (11) 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,

$$\alpha\mathbf{x}' = \text{H}_{\mathbf{u}}\mathbf{x} = \begin{bmatrix} \mathbf{h}_{\mathbf{u}}^1 & \mathbf{h}_{\mathbf{u}}^2 & \mathbf{h}_{\mathbf{u}}^3 \end{bmatrix}^\top \mathbf{x} = \begin{bmatrix} \text{I}_3 + \mathbf{u}\mathbf{l}^\top \end{bmatrix} \mathbf{x}. \quad (24)$$

The homogeneous scale in (24) can be eliminated by substituting $\mathbf{h}_{\mathbf{u}}^{3\top}\mathbf{x}$ for α, and the system can be rearranged such that

$$\begin{aligned} \mathbf{x}^\top \mathbf{h}_{\mathbf{u}}^1 &= (x'\mathbf{x}^\top)\mathbf{h}_{\mathbf{u}}^3 \\ \mathbf{x}^\top \mathbf{h}_{\mathbf{u}}^2 &= (y'\mathbf{x}^\top)\mathbf{h}_{\mathbf{u}}^3. \end{aligned} \quad (25)$$

Collecting the terms of vanishing point after expanding the dot products in (25) 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

$$\underset{\mathbf{u}}{\text{minimize}} \ \|\text{M}\mathbf{u} - \mathbf{y}\|^2$$

$$\text{subject to} \ \mathbf{l}^\top \mathbf{u} = 0,$$

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

Since the matrix $\begin{bmatrix} \text{M}^\top & \mathbf{l} \end{bmatrix}^\top$ has linearly independent columns,

and $\mathbf{l}^\top$ is trivially row independent, $\mathbf{u}$ is recovered by solving

$$\begin{bmatrix} \mathtt{M}^\top \mathtt{M} & \mathbf{l} \\ \mathbf{l}^\top & 0 \end{bmatrix} \begin{pmatrix} \mathbf{u} \\ z \end{pmatrix} = \begin{pmatrix} \mathtt{M}^\top \mathbf{y} \\ 0 \end{pmatrix}, \qquad (26)$$

where z is a nuisance variable [44]. Surprisingly, a superior estimation of the vanishing point $\mathbf{u}$ is given by using (26) 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 Fig. 10a).

6 DEGENERACIES

We identified three important degeneracies for the solvers: Two geometric configurations of features such that there exists either a subspace of rectifications or no valid solution, and a modeling degeneracy introduced by the use of the expression (4) for the affine-rectifying homography, which requires $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$ such that $l_3 \neq 0$ [10]. The proposed solvers and the state-of-the-art solvers of Pritts et al. in [14], [26] 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.

6.1 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 of lens undistortion uniformly scales the points about the image center. Given $\lambda \neq 0$, for a transformation by $f(\cdot, \lambda)$ defined in (5) 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 $\mathtt{H}\mathtt{S}(\lambda)$ 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 Fig. 5. 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 Fig. 8). Regardless, the degeneracy can be avoided by using different meets.

6.2 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 (6) is canceled. In this case, it is not possible to recover the division model parameter λ from the systems of equations (16), (17) or (22) 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 [32]. The derivations in this paper assume that image center, distortion center and image origin are coincident. The proposed and state-of-the-art solvers of Pritts et al. [14], [26] formulate joint undistortion and rectification in terms of (6), which leaves the distortion center stationary.

Directional cameras see only points in front of the camera [45], 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.*, Figs. 4c, 7c, and 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 4). 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.

7 ROBUST ESTIMATION

The solvers are used in a LO-RANSAC-based robust-estimation framework [11], [46]. 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 [11]. The metric-rectifications are presented in the results.

(a) MSER detection   (b) Normalized Frame   (c) LAF representation
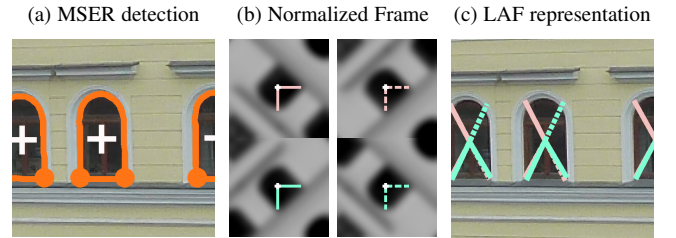

Fig. 8. Repeat Detection, Description, and Representation. (a) Center of gravity (white cross) and curvature extrema (orange circles) of a detected MSER (orange contour [40]). Patches are normalized to a square and oriented to define an affine frame as in [41], (b) Bases are reflected for detecting axial symmetries. The RootSIFT transform embeds the local texture [47], [48]. (c) Affine frames are mapped to the image.

7.1 Local Features and Descriptors

Covariant region detectors are highly repeatable on the same imaged scene texture with respect to significant changes of viewpoint and illumination [49], [50]. Their proven robustness in the multi-view matching task makes them good candidates for representing

the local geometry of repeated textures. In particular, we use the Maximally-Stable Extremal Region and Hessian-Affine detectors [23], [40]. The affine-covariant regions are given by an affine basis (see Sec. 3.2), equivalently three distinct points, in the image space [42]. The image patch local to the affine frame is embedded into a descriptor vector by the RootSIFT transform [47], [48]. See Fig. 8 for a visualization.

7.2 Detection, Description, and Clustering

Affine frames are tentatively labeled as repeated texture by their appearance. The appearance of an affine frame is given by the RootSIFT embedding of the image patch local to the affine frame [47]. Affine-covariant regions are also extracted and embedded in the reflected image, where the detections in are transformed into the original image space and have a left-handed representation.

The RootSIFT descriptors are agglomeratively clustered, which establishes pair-wise tentative correspondences amongst connected components. Since the proposed $H_2 lu\lambda$, $H_2 lu s^u \lambda$, $H_2 l\lambda$, and $H_{22} luv\lambda$ 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 $H_{22} luv s^v \lambda$, which admits reflections of similarity-covariant regions. 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*.

7.3 Sampling

Sample configurations for the proposed minimal solvers are illustrated in Figs. 1, 3, and 6 as well as detailed in Secs. 5.1 and 5.2. 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.

7.4 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 Sec. 7.3. The affine rectification estimated by the minimal solver is used to build an affine-rectified scale consensus set. The scale consensus set is 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 [10], [14], [21], [26], [29]. Note that if clustered left and right-handed regions were partitioned for sampling with the $H_2 lu\lambda$, $H_2 lu s^u \lambda$, $H_2 l\lambda$, and $H_{22} luv\lambda$ 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. [11], 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], [21], [26]. 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 [11]. 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 [11].

8 EXPERIMENTS

We compare the proposed solvers to the bench of state-of-the-art solvers listed in Table 3. We apply the denotations for the solvers introduced in Sec. 2.3 to all tested solvers. Included is the state-of-the-art joint undistorting and rectifying solver $H_{222}^{DES} l\lambda$ of Pritts et al. [14], [26], 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 [14], [26], we test only the $H_{222}^{DES} l\lambda$ solver since all variants are reported to have similar noise sensitivities. Also included is the $H_2 l$ solver of Schaffalitzky et al. [20], which incorporates similar constraints from conjugate translations that are used to derive the proposed solvers. Two full-homography and radial-undistortion solvers are included: the $H_{22} \lambda$ solver of Fitzgibbon et al. [32] and the $H_{22} \lambda_1 \lambda_2$ solver of Kukelova et al. [12], 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 [12], [32]. The solvers are evaluated on synthetic scenes and challenging real images.

8.1 Synthetic Data

The sensitivity studies evaluate the solvers on noisy measurements over 3 task-related performance metrics: (i) the transfer error, which measures the accuracy of radially-distorted conjugate translation estimation (ii) the warp error which measures rectification accuracy, and (iii) the relative error of the division-model parameter estimate, which reports the accuracy of the lens undistortion estimate. A solver stability study evaluates the proposed solvers by the warp error on noiseless measurements. The study demonstrates the benefit of constraint simplification by the hidden-variable trick, 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 [34].

These studies are evaluated on 1000 synthetic images of 3D scenes with known ground-truth parameters. 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 imaged translated coplanar repeats. 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. The image is distorted according to the
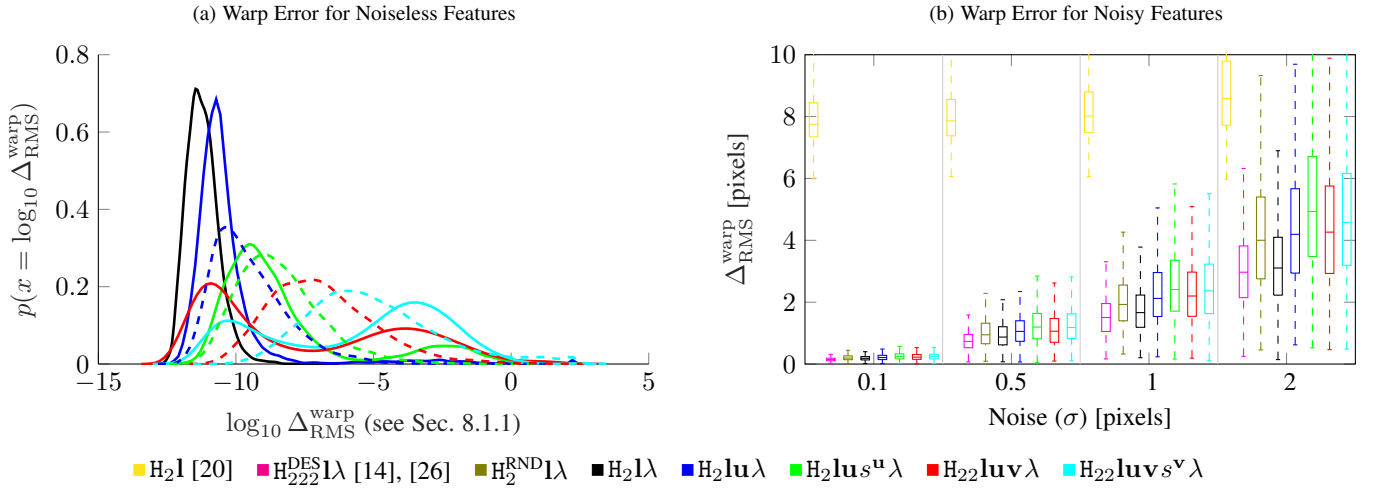
(a) Warp Error for Noiseless Features

(b) Warp Error for Noisy Features

$\blacksquare \mathrm{H}_2\mathbf{l}$ [20] $\blacksquare \mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ [14], [26] $\blacksquare \mathrm{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$ $\blacksquare \mathrm{H}_2\mathbf{l}\lambda$ $\blacksquare \mathrm{H}_2\mathbf{l}\mathbf{u}\lambda$ $\blacksquare \mathrm{H}_2\mathbf{l}\mathbf{u}s^{\mathbf{u}}\lambda$ $\blacksquare \mathrm{H}_{22}\mathbf{l}\mathbf{u}\mathbf{v}\lambda$ $\blacksquare \mathrm{H}_{22}\mathbf{l}\mathbf{u}\mathbf{v}s^{\mathbf{v}}\lambda$

Fig. 9. (a) The $\log_{10}$ RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ is reported for noiseless scenes generated as described in Secs. 8.1.1 and 8.1.3. Hidden variable trick solvers are solid; solvers generated without simplified constraints equations are dashed. The hidden-variable trick increases stability. The EVL $\mathrm{H}_2\lambda$ solver is the most stable since it does not require solving a complicated polynomial system of equations. (b) Reports the RMS error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ 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, 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 $\mathrm{H}_2\lambda$ is competitive with $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$, which requires two more correspondences. The $\mathrm{H}_2\mathbf{l}\lambda$ solver uses best minimal solution selection, which improves its performance compared to $\mathrm{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$, which randomly selects a solution.

division model. For the sensitivity experiments, isotropic white noise is added to the distorted affine frames at increasing levels.

In addition, the convergence speed of RANSAC is compared with respect to the use of each of the proposed and state-of-the-art solvers to hypothesize models. In particular, convergence is evaluated by plotting rectification accuracy with respect to trial number.

8.1.1 Warp Error

Since the accuracy of scene-plane rectification is a primary concern, an extended version of the warp error for rectifying homographies proposed by Pritts et al. [5] that incorporates the division model for radial lens distortion of Fitzgibbon [32] is used to evaluate a solver's stability and robustness to noise. A scene plane is tessellated by a 10x10 square grid of points $\{\mathbf{X}_i\}_{i=1}^{100}$ and imaged as $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$ by the lens-distorted ground-truth camera. The tessellation ensures that error is uniformly measured over the scene plane. A round trip between the image space and rectified space is made by affine-rectifying $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$ using the estimated division model parameter $\hat{\lambda}$ and rectifying homography $\mathrm{H}(\hat{\mathbf{l}})$ (see (4)) and then imaging the rectified plane by the ground-truth camera P. Ideally, the ground-truth camera P images the rectified points $\{\underline{\mathbf{x}}_i\}_{i=1}^{100}$ onto the distorted points $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$. There is an affine ambiguity, denoted A, between $\mathrm{H}(\hat{\mathbf{l}})$ and the ground-truth camera matrix P. The ambiguity is estimated during computation of the warp error,

$$\Delta^{\mathrm{warp}} = \min_{\mathtt{A}} \sum_i d^2(\tilde{\mathbf{x}}_i, f^d(\mathtt{PAH}(\hat{\mathbf{l}})f(\tilde{\mathbf{x}}_i, \hat{\lambda})), \lambda), \quad (27)$$

where $d(\cdot, \cdot)$ is the Euclidean distance, f^d is the inverse of the division model (the inverse of (5)). The root mean square warp error for $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$ is reported and denoted as $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$.

Note that the $\mathrm{H}_{22}\lambda$ solver of [32] and the $\mathrm{H}_{22}\lambda_1\lambda_2$ solver of [12] are omitted from the warp error since the vanishing line is not directly estimated.

8.1.2 Transfer Error

The geometric transfer error measures the accuracy of the estimated radially-distorted conjugate translation. The scene plane is tessellated by a 10x10 grid of points spaced one unit apart. The tessellation ensures that the geometric transfer error is uniformly sampled across the image. Denote the tessellation as $\{\mathbf{X}_i\}_{i=1}^{100}$. Suppose that $\mathbf{x} \leftrightarrow \mathbf{x}'$ are conjugately-translated points that are consistent with $\mathrm{H}_{\mathbf{u}} = [\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top]$. The preimage $\mathbf{U}$ of the vanishing translation direction $\mathbf{u}$ is recovered as $\beta\mathbf{U} = \mathtt{P}^{-1}\mathbf{u} = \beta(u_x, u_y, 0)^\top$. The tessellation is translated by $\mathbf{U}/\|\mathbf{U}\|$ on the scene plane. Let $\mathtt{T}(\mathbf{U}/\|\mathbf{U}\|)$ be a homogeneous translation matrix as defined in (7). Then by (8) the image of the unit-magnitude translation on the scene plane in the direction of correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ is

$$\mathrm{H}_{\mathbf{u}/\|\mathbf{U}\|} = [\mathtt{I}_3 + \frac{\mathbf{u}}{\|\mathbf{U}\|}\mathbf{l}^\top]. \quad (28)$$

The unit conjugate translation $\mathrm{H}_{\mathbf{u}/\|\mathbf{U}\|}$ can be written in terms of $\mathrm{H}_{\mathbf{u}}$ as

$$\begin{aligned} \mathtt{I}_3 + \frac{\mathbf{u}}{\|\mathbf{U}\|}\mathbf{l}^\top &= \mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}[\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top - \mathtt{I}_3] \\ &= \mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}[\mathrm{H}_{\mathbf{u}} - \mathtt{I}_3]. \end{aligned} \quad (29)$$

The points are distorted with the ground-truth division-model parameter λ as $\tilde{\mathbf{x}}_i = f^d(\mathbf{x}_i, \lambda)$ and $\tilde{\mathbf{x}}_i' = f^d(\mathbf{x}_i', \lambda)$, where f^d transforms from pinhole points to radially-distorted points. Then the geometric transfer error is defined as

$$\Delta^{\mathrm{xfer}} = d(f^d([\mathtt{I}_3 + \frac{1}{\|\mathbf{U}\|}(\hat{\mathrm{H}}_{\mathbf{u}} - \mathtt{I}_3)]f(\tilde{\mathbf{x}}, \hat{\lambda}_1), \hat{\lambda}_2), \tilde{\mathbf{x}}'), \quad (30)$$

where $d(\cdot, \cdot)$ is the Euclidean distance and $\hat{\mathrm{H}}_{\mathbf{u}}$ and $\hat{\lambda}_1, \hat{\lambda}_2$ are the estimated conjugate translations and division model parameters, respectively. All solvers except $\mathrm{H}_{22}\lambda_1\lambda_2$ have the constraint that $\hat{\lambda}_1 = \hat{\lambda}_2$ [12].
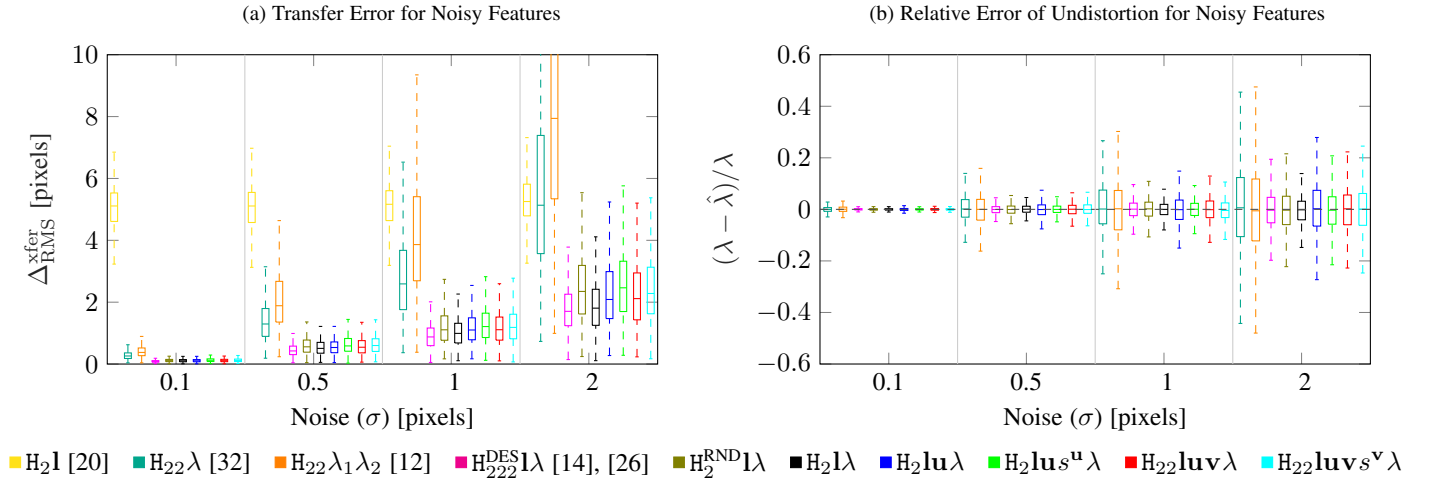
Fig. 10. Comparison of two error measures after 25 iterations of a simple RANSAC for different solvers with increasing levels of white noise added to the affine covariant region correspondences, where the normalized division model parameter is set to -4 (see Sec. 3.1), which is similar to the distortion of a GoPro Hero 4. Results are for imaged translated coplanar repeats: (a) Reports the root mean square transfer error $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$. With the exception of the $\mathtt{H}_{222}^{\mathrm{DES}}\mathtt{l}\lambda$ solver, the proposed solvers are significantly more robust for both types of repeats on both error measures; however $\mathtt{H}_{222}^{\mathrm{DES}}\mathtt{l}\lambda$ requires the most correspondences, and (b) reports the relative error of the estimated division model parameter. The $\mathtt{H}_2\mathtt{l}\lambda$ solver uses best minimal solution selection, which improves its performance compared to $\mathtt{H}_2^{\mathrm{RND}}\mathtt{l}\lambda$, which randomly selects a solution.

The root mean square transfer error $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$ 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 $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$. The transfer error is used in the sensitivity study, where the solvers are tested over varying noise levels with a fixed division model parameter.

8.1.3 Numerical Stability

The stability study measures the RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ of solvers 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 parameter λ 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 $1/(\text{width} + \text{height})$. Fig. 9a reports the histogram of $\log_{10}$ warp errors $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$.

For the proposed EVP solvers we evaluate a solver generated from constraints derived with (solid histogram) and without (dashed histogram) the hidden-variable trick. The hidden-variable trick significantly improves the stability of the proposed solvers. The increased stabilities of the hidden-variable solvers 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 $\mathtt{H}_2\mathtt{l}\lambda$ is derived with the hidden-variable trick as well, which results in a quartic. The superior stability of the $\mathtt{H}_2\mathtt{l}\lambda$ solver demonstrates the benefits of the elementary formulation.

8.1.4 Noise Sensitivity

The accuracy of the proposed and state-of-the-art solvers is measured by the warp error, transfer error, and relative error of lens undistortion with respect to increasing levels of white noise added to radially-distorted conjugately-translated point correspondences.

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 Fig. 9b), the RMS transfer error (see Fig. 10a) $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$, or the relative error of lens distortion (see Fig. 10b) 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—$\mathtt{H}_2\mathtt{lu}\lambda, \mathtt{H}_2\mathtt{lu}s^{\mathbf{u}}\lambda, \mathtt{H}_{22}\mathtt{luv}\lambda, \mathtt{H}_{22}\mathtt{luv}s^{\mathbf{v}}\lambda$, and $\mathtt{H}_2\mathtt{l}\lambda$—demonstrate excellent robustness to noisy features across all three error measures. In particular, the $\mathtt{H}_2\mathtt{l}\lambda$ solver is the least sensitive to noise of the proposed solvers and gives the best undistortion estimates of any solver in the bench (see Fig. 10b). All proposed solvers estimate the correct lens distortion parameter more than half the time. Fig. 9b 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. Fig. 10a 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.

For both the warp error and transfer error studies, the $\mathtt{H}_2\mathtt{l}$ solver of Schaffalitzky et al. [20] 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 overparameterized radial-distortion homography solvers of $\mathtt{H}_{22}\lambda$ [32] Fitzgibbon and $\mathtt{H}_{22}\lambda_1\lambda_2$ [12] 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
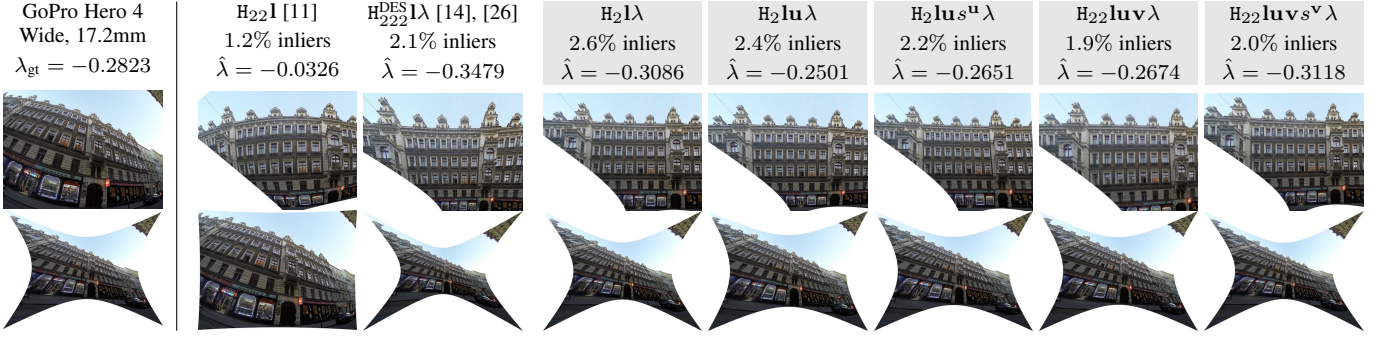
| GoPro Hero 4 Wide, 17.2mm $\lambda_{gt} = -0.2823$ | $\mathrm{H}_{22}\mathbf{l}$ [11] 1.2% inliers $\hat{\lambda} = -0.0326$ | $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ [14], [26] 2.1% inliers $\hat{\lambda} = -0.3479$ | $\mathrm{H}_2\mathbf{l}\lambda$ 2.6% inliers $\hat{\lambda} = -0.3086$ | $\mathrm{H}_2\mathbf{lu}\lambda$ 2.4% inliers $\hat{\lambda} = -0.2501$ | $\mathrm{H}_2\mathbf{lu}s^{\mathbf{u}}\lambda$ 2.2% inliers $\hat{\lambda} = -0.2651$ | $\mathrm{H}_{22}\mathbf{luv}\lambda$ 1.9% inliers $\hat{\lambda} = -0.2674$ | $\mathrm{H}_{22}\mathbf{luv}s^{\mathbf{v}}\lambda$ 2.0% inliers $\hat{\lambda} = -0.3118$ |

Fig. 11. The bench of solvers is evaluated on an image with poorly corresponded repeated regions. The leftmost column shows the original image (top) and its undistortion using ground truth (bottom) for comparison. The proposed solvers are highlighted in gray. The estimated rectifications and undistorted images are synthesized, and the normalized values of the division-model parameter and ratio of inliers are reported. The $\mathrm{H}_2\mathbf{l}\lambda$ solver gives the best result, and all proposed solvers perform better than the state-of-the-art.

Fig. 10a, the performance of these solvers is worse than the $\mathrm{H}_2\mathbf{l}$ solver, which does not model radial lens distortion.

The state-of-the art solver $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ of Pritts et al. [14], [26] 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 $\mathrm{H}_2\mathbf{l}\lambda$ reaches near parity with the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver and requires only one region correspondence versus three required by the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver. As is shown in Sec. 8.2, 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 $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver for images with radially-distorted conjugate translations.

Each of the $\mathrm{H}_2\mathbf{l}\lambda$ and $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solvers requires the ex-post estimation of vanishing point of the translation direction, which is computed by the method proposed in Sec. 5.2.3. Surprisingly, the sequential estimation used by the proposed $\mathrm{H}_2\mathbf{l}\lambda$ and the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver of [14], [26] achieve the best performances on the transfer error $\Delta_{\mathrm{RMS}}^{\mathrm{xfer}}$. This is explainable by the improved performance of the $\mathrm{H}_2\mathbf{l}\lambda$ EVL solver with respect to the EVP solvers on all measures, and the fact that the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver uses three correspondences, the most of any in the bench of solvers (see Table 3).

The benefit of best minimal solution selection as proposed in (5.2.2) can be seen by comparing the $\mathrm{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$ and $\mathrm{H}_2\mathbf{l}\lambda$ solvers in all sensitivity studies. To quickly recap, The $\mathrm{H}_2^{\mathrm{RND}}\mathbf{l}\lambda$ solver randomly selects a minimal solution from 10 possible solutions given by the EVL geometry shown in Fig. 6, while the $\mathrm{H}_2\mathbf{l}\lambda$ chooses the solution that minimizes a geometric error on the unused constraints. The sensitivity improvements using best minimal solution selection are considerable: at the 2 pixel noise levels, the RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ (Fig. 9b) and RMS transfer error (Fig. 10a) 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 $\mathrm{H}_2\mathbf{l}\lambda$ solver on par with the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver, which requires two more region correspondences.

8.1.5 RANSAC convergence study

The speed of convergence of RANSAC is evaluated using each of the proposed and state-of-the-art rectifying solvers to hypothesize models. Each RANSAC variant is run on a set of thirty synthetic scenes generated as described in Sec. 8.1 with one-pixel white noise and with corresponded sets of regions corrupted by 50%

outliers. Fig. 12 reports the mean RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ over all scenes at each iteration for the bench of solvers. The proposed $\mathrm{H}_2\mathbf{l}\lambda$ solver converges fastest, which demonstrates its robustness to noise and the advantage of one-correspondence sampling combined with best minimal-solution selection. The $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ performs second best, despite requiring three region correspondences; however, it not so surprising since its robustness is known (e.g., see Fig. 10 and [14], [26]). As expected, the solver of [20] performs poorly since it does not estimate lens undistortion. The solvers of [12], [32] are omitted from this study since they do not directly recover the vanishing line. In theory, it is possible to use a rank-one decomposition to recover the estimated vanishing line from a full homography for warp error computation, but we found this to give poor results.

8.2 Computational Complexity

Table 4 lists the wall-clock time to solution for the optimized C++ implementations of the proposed solvers and the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver [14], [26], which was the only competitive solver from the sensitivity experiments reported in Figs. 10a, 9b, and 10b. Also reported for easy comparison are the relative speeds with respect to the $\mathrm{H}_2\mathbf{l}\lambda$ solver and the elimination template sizes, where applicable. The proposed EVL $\mathrm{H}_2\mathbf{l}\lambda$ solver is an astounding $2153.6\times$ faster than the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver and significantly faster than all EVP solvers ($\mathrm{H}_2\mathbf{lu}\lambda, \mathrm{H}_2\mathbf{lu}s^{\mathbf{u}}\lambda, \mathrm{H}_{22}\mathbf{luv}\lambda$, and $\mathrm{H}_{22}\mathbf{luv}s^{\mathbf{v}}\lambda$), which require the Gröbner basis method to solve polynomial systems of equations. All of the proposed solvers are much faster than the $\mathrm{H}_{222}^{\mathrm{DES}}\mathbf{l}\lambda$ solver, making them more suitable for fast sampling in RANSAC for scenes containing translational symmetries.

TABLE 4
Runtime Analysis

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

Wall-clock times are reported for optimized C++ implementations of the proposed solvers versus $\mathrm{H}_{222}^{DES}\mathbf{l}\lambda$ of [14], [26], which was the only competitive solver from the noise sensitivity experiments. The EVL solver is $2153.6\times$ faster than $\mathrm{H}_{222}^{DES}\mathbf{l}\lambda$, and the other proposed variants are orders of magnitude faster.
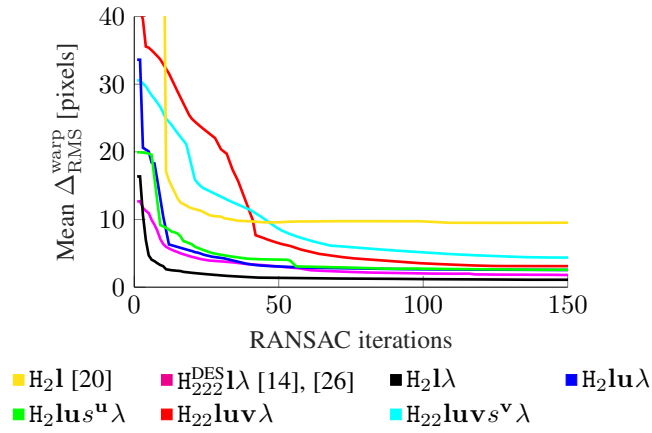
Fig. 12. Each solver is used to generate models for RANSAC on a sets of synthetic noisy region correspondences with 50% outliers. The mean RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ over all scenes at each iteration is shown. The $\mathrm{H}_2\mathrm{l}\lambda$ gives the most accurate rectifications the fastest.

8.3 Real Images

In the experiments on real images shown in Figs. 1 and 4, 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 Fig. 4, 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 Fig. 4a and the undistortions between the wide GoPro images in Figs. 1 and 4b. Regardless of the 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, an excellent rectification is achieved for the fisheye distorted image taken with the Samyang 7.5mm lens in Fig. 4c, and the horizon line is perfectly estimated.

Fig. 7 shows results obtained with 1-correspondence sampling using the proposed $\mathrm{H}_2\mathrm{l}\lambda$ EVL solver on very challenging fisheye images. Images from five distinct fisheye lenses are evaluated with Figs. 7b, 7c, and 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.*, Figs. 7c and 7d exhibit some mustache distortion, which cannot be modeled with 1 parameter. However, the local optimizer of [11] could be modified to regress a higher-order distortion model using the results of Fig. 7 as an initial guess. We leave this for future work.

Figs. 4c, 7c, and 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 Sec. 6). 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 distorted image evaluated in Fig. 11 has a very low inlier ratio of corresponded coplanar regions, which is typical for repeated content that is clustered by appearance. The low-inlier example is rectified with the five proposed solvers and the solvers $\mathrm{H}_{222}^{\mathrm{DES}}\mathrm{l}\lambda$ of [14], [26] and $\mathrm{H}_{22}\mathrm{l}$ of [21]. All solvers were used within an extension to the coplanar repeat detection and rectification framework of Pritts et al. [11]. The ground truth division-model parameter λ_{gt} was obtained using calibration software and chessboard images. The proposed $\mathrm{H}_2\mathrm{l}\lambda$ gives the best undistortion and rectification both quantitatively and qualitatively. The estimation framework using the $\mathrm{H}_2\mathrm{l}$ solver, which does not solve for lens undistortion, is unable to recover a reasonable lens undistortion. In general, the solvers requiring more correspondences and having more degrees of freedom give less accurate results. 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 [11].

The narrow field of view and diverse content experiment of Fig. 13 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.

9 CONCLUSIONS

This paper 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 $\mathrm{H}_2\mathrm{l}\lambda$ solver admits the same point configuration as the one-direction EVP solver $\mathrm{H}_2\mathrm{lu}\lambda$, but is much simpler (*i.e.*, does not require the Gröbner bases method), more stable, and is $7.4\times$ 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.*, Pritts et al. in [14], [26] were unable to reduce the degree of their constraint equations used for the $\mathrm{H}_{222}^{\mathrm{DES}}\mathrm{l}\lambda$ solver, which resulted in slow solver (see Table 4). Furthermore, stability sampling was required to generate useful solvers [17].

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. [12], [32]. 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 $\mathrm{H}_{222}^{\mathrm{DES}}\mathrm{l}\lambda$ solver of [14], [26] despite the fact that the $\mathrm{H}_{222}^{\mathrm{DES}}\mathrm{l}\lambda$ solver requires two more region correspondences as input
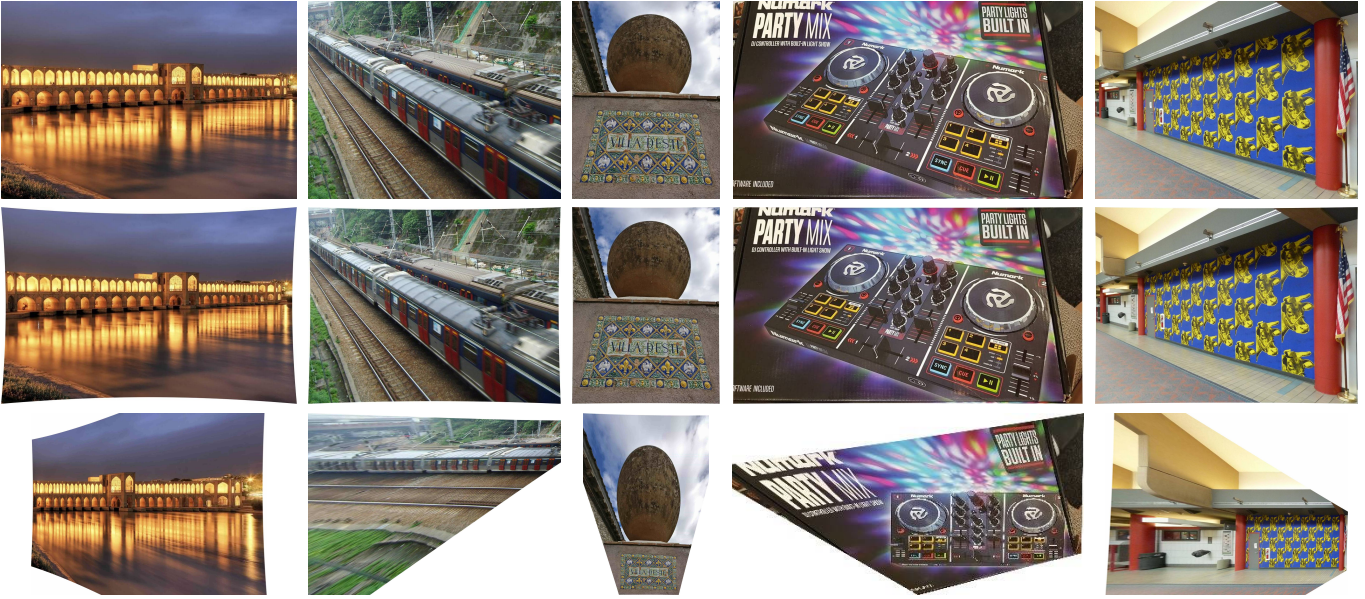
Fig. 13. The proposed solvers works well on images with small lens distortions. Input images are on the top row; undistorted images are on the middle row, and the rectified images are on the bottom. Results were generated with the $\mathrm{H}_2\mathbf{lu}\lambda$ solver.

(compared to $\mathrm{H}_2\mathbf{l}\lambda$, $\mathrm{H}_2\mathbf{lu}\lambda$, and $\mathrm{H}_2\mathbf{lu}s^{\mathbf{u}}\lambda$). 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.

ACKNOWLEDGMENTS

REFERENCES

[1] J. Hays, M. Leordeanu, A. A. Efros, and Y. Liu, "Discovering texture regularity as a higher-order correspondence problem," in *ECCV*, 2006.

[2] 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.

[3] 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.*, vol. 36, no. 4, pp. 108:1–108:11, Jul. 2017.

[4] C. Wu, J. M. Frahm, and M. Pollefeys, "Repetition-based dense single-view reconstruction," in *CVPR*, 2011.

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

[6] G. Zeng and L. Van Gool, "Multi-label image segmentation via point-wise repetition," in *CVPR*, 2008, pp. 1–8.

[7] H. Wildenauer and A. Hanbury, "Robust camera self-calibration from monocular images of manhattan worlds," in *CVPR*. IEEE, 2012, pp. 2831–2838.

[8] H. Wildenauer and B. Micusík, "Closed form solution for radial distortion estimation from a single vanishing point," in *BMVC*, 2013.

[9] 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.

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

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

[12] Z. Kukelova, J. Heller, B. M., and T. Pajdla, "Radial distortion homography," in *CVPR*, 2015.

[13] F. Bukhari and M. N. Dailey, "Automatic radial distortion estimation from a single image," *Journal of Mathematical Imaging and Vision*, vol. 45, no. 1, pp. 31–45, Jan 2013.

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

[15] Z. Kukelova, M. Bujnak, and T. Pajdla, "Automatic generator of minimal problem solvers," in *ECCV*, 2008.

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

[17] V. Larsson, M. Oskarsson, K. Åström, A. Wallis, Z. Kukelova, and T. Pajdla, "Beyond grobner bases: Basis selection for minimal solvers," in *CVPR*, 2018.

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

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *CACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[20] F. Schaffalitzky and A. Zisserman, "Geometric grouping of repeated elements within images," in *BMVC*, 1998.

[21] O. Chum and J. Matas, "Planar affine rectification from change of scale," in *ACCV*, 2010.

[22] D. Barath, T. Toth, and L. Hajder, "A minimal solution for two-view focal-length estimation using two affine correspondences," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[23] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.

[24] M. Perd'och, J. Matas, and O. Chum, "Epipolar geometry from two correspondences," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4, 2006, pp. 215–219.

[25] C. Raposo and J. P. Barreto, "Theory and practice of structure-from-motion using affine correspondences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5470–5478.

[26] J. Pritts, Z. Kukelova, V. Larsson, Y. Lochman, and O. Chum, "Minimal solvers for rectifying from radially-distorted scales and change of scales," *International journal of computer vision*, pp. 1–19, 2020.

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

[28] D. Aiger, D. Cohen-Or, and N. J. Mitra, "Repetition maximization based texture rectification," *Computer Graphics Forum*, vol. 31, no. 2pt2, pp. 439–448, 2012.

[29] A. Criminisi and A. Zisserman, "Shape from texture: homogeneity revisited," in *BMVC*, 2000.

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

[31] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: transform invariant low-rank textures," *International Journal of Computer Vision*, vol. 99, no. 1, pp. 1–24, 2012.

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

[33] H. Li and R. Hartley, "A non-iterative method for correcting lens distortion from nine-point correspondences," in *In Proc. OmniVision'05, ICCV-workshop,2005*, 2005.

[34] D. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*, 2nd ed. Springer, 2004.

[35] V. Larsson, K. Astrom, and M. Oskarsson, "Polynomial solvers for saturated ideals," in *ICCV*, 2017.

[36] R. G. Willson and S. Shafer, "What is the center of the image?" *Journal of the Optical Society of America A*, vol. 11, no. 11, pp. 2946–2955, Nov 1994.

[37] R. Strand and E. Hayman, "Correcting radial distortion by circle fitting," in *BMVC*, 2005.

[38] 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*, vol. 35, no. 3, pp. 165–172, 2009.

[39] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

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

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

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

[43] D. R. Grayson and M. E. Stillman, "Macaulay 2, a software system for research in algebraic geometry," 2002.

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

[45] R. I. Hartley, "Chirality," *International Journal of Computer Vision*, vol. 26, no. 1, pp. 41–61, Jan 1998.

[46] O. Chum, J. Matas, and v. Obdržálek, "Enhancing RANSAC by generalized model optimization," in *ACCV*, 2004.

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

[48] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[49] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[50] D. Mishkin, F. Radenovic, and J. Matas, "Repeatability is not enough: Learning affine regions via discriminability," in *ECCV*, 2018.

James Pritts is a Post-Doctoral Research Scientist at Facebook Reality Labs. Previously, Mr. Pritts was a researcher at the Czech Institute of Informatics, Robotics and Cybernetics at Czech Technical University in Prague. Mr. Pritts also serves as an adviser for The Machine Learning Lab at Ukrainian Catholic University in Lviv. He is the recipient of the ACCV 2018 Saburo Tsuji Best Paper Award and the ICVNZ 2013 Best Paper Award. He works mostly on minimal solvers and multi-model estimation.

Zuzana Kukelova is a research fellow at the Visual Recognition Group at the Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague. She was a postdoctoral researcher at Microsoft Research Cambridge, United Kingdom. Her research interests include algebraic geometry and computer vision, where she focuses on minimal problems and methods for solving systems of polynomial equations. She is the recipient of the 2015 Cor Baayen Award, given to a promising young researcher in computer science and applied mathematics from countries associated with the European Research Consortium for Informatics and Mathematics.
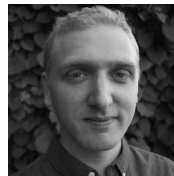
Viktor Larsson received his PhD degree from the Centre for Mathematical Sciences at Lund University in Sweden. He is currently a postdoctoral researcher at the Computer Vision and Geometry Group at ETH Zurich led by Prof. Marc Pollefeys. His work is mostly focused on minimal problems in geometric computer vision.

Yaroslava Lochman is a research intern at Facebook Reality Labs. She received her MSc in data science from Ukrainian Catholic University in Lviv, where she also conducted research in computer vision at The Machine Learning Lab. She is currently focused on robust methods for camera calibration.

Ondřej Chum is an associate professor at the Czech Technical University in Prague, where he leads a team within the Visual Recognition Group at the Department of Cybernetics, Faculty of Electrical Engineering. He received the MSc degree in computer science from Charles University, Prague, in 2001 and the PhD degree from the Czech Technical University in Prague, in 2005. From 2006 to 2007, he was a postdoctoral researcher at the Visual Geometry Group, University of Oxford, United Kingdom. The research interests include large-scale image and particular object retrieval, object recognition, and robust estimation of geometric models. He is a member of Image and Vision Computing editorial board, and has served in various roles at major international conferences (e.g., ICCV, ECCV, CVPR, and BMVC). He was the recipient of the Best Paper Prize at the BMVC in 2002, the Best Science Paper Honorable Mention at BMVC 2017, Longuet-Higgins Prize at CVPR 2017, and the Saburo Tsuji Best Paper Award at ACCV 2018. Ondřej was awarded the 2012 Outstanding Young Researcher in Image and Vision Computing runner up for researchers within seven years of their PhD.