Rectification from Radially-Distorted Scales

James Pritts[1,2], Zuzana Kukelova[2], Viktor Larsson[3,*], and Ondřej Chum[2]

[1] Czech Institute of Informatics, Robotics and Cybernetics (CIIRC), CTU in Prague
[2] Visual Recognition Group (VRG), FEE, CTU in Prague
[3] Department of Computer Science, ETH Zürich, Switzerland

Abstract. This paper introduces the first minimal solvers that jointly estimate lens distortion and affine rectification from repetitions of rigidly-transformed coplanar local features. The proposed solvers incorporate lens distortion into the camera model and extend accurate rectification to wide-angle images that contain nearly any type of coplanar repeated content. We demonstrate a principled approach to generating stable minimal solvers by the Gröbner basis method, which is accomplished by sampling feasible monomial bases to maximize numerical stability. Synthetic and real-image experiments confirm that the solvers give accurate rectifications from noisy measurements if used in a RANSAC-based estimator. The proposed solvers demonstrate superior robustness to noise compared to the state of the art. The solvers work on scenes without straight lines and, in general, relax strong assumptions about scene content made by the state of the art. Accurate rectifications on imagery taken with narrow focal length to fisheye lenses demonstrate the wide applicability of the proposed method. The method is automatic, and the code is published at https://github.com/prittjam/repeats.

Keywords: rectification · radial lens distortion · repeated patterns.

1 Introduction

This paper proposes minimal solvers that jointly estimate affine rectification and lens distortion from local features extracted from arbitrarily repeating coplanar texture. Wide-angle lenses with significant radial lens distortion are common in consumer cameras like the GoPro series of cameras. In the case of Internet imagery, the camera and its metadata are often unavailable for use with off-line calibration techniques. The state of the art has several approaches for rectifying (or partially calibrating) a distorted image, but these methods make restrictive assumptions about scene content by assuming, *e.g.*, the presence of sets of parallel lines [1, 2]. The proposed solvers relax the need for special scene structure to unknown repeated structures (see Table 1). The solvers are fast and robust to noisy feature detections, so they work well in robust estimation frameworks like RANSAC [3]. The proposed work is applicable for several important computer vision tasks including symmetry detection [4], inpainting [5], and single-view 3D reconstruction [6].

The proposed solvers enforce the affine invariant that rectified repeated regions have the same scale. We introduce three variants (see Fig. 1) that use different configurations

---

*This work was done while Viktor Larsson was at Lund University.
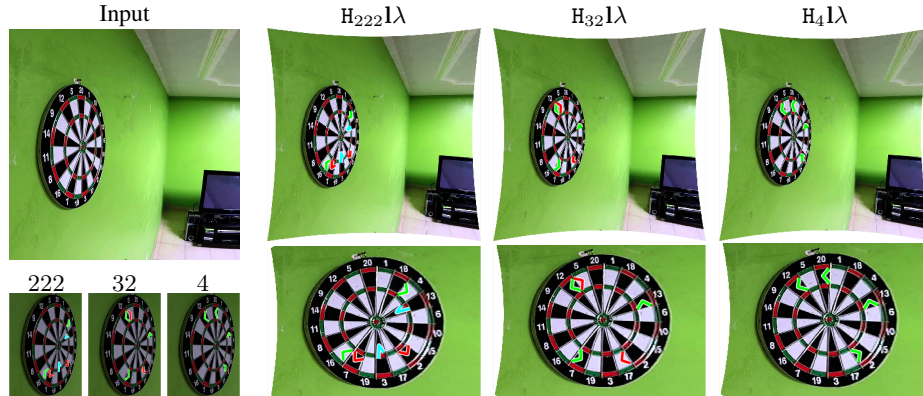
Fig. 1. *Solver Variants.* (top-left image) The input to the method is a single image. (bottom-left triptych, contrast enhanced) The three configurations—$222, 32, 4$—of affine frames that are inputs to the proposed solvers variants. Corresponded frames have the same color. (top row, right) Undistorted outputs of the proposed solver variants. (bottom row, right) Cutouts of the dartboard rectified by the proposed solver variants. The affine frame configurations—$222, 32, 4$—are transformed to the undistorted and rectified images.

of coplanar repeated features as input, which allows for flexible sampling during robust estimation. Lens distortion is parameterized by the division model, which Fitzgibbon [7] first used to model lens distortion and showed that it accurately models significant distortions. The use of the division model is crucial because other typical distortion models result in unsolvable constraint equations (see Sec. 2.1). A fourth solver variant is proposed that assumes the pinhole camera model. The pinhole variant is also novel because it does not linearize the rectifying transformation, which is the approach of the state of the art [8–10].

The polynomial system of equations encoding the rectifying constraints is solved using an algebraic method based on Gröbner bases. Automated solver generators using the Gröbner basis method [11, 12] have been used to generate solvers for several camera geometry estimation problems [11–15]. However, straightforward application of automated solver generators to the proposed constraints resulted in unstable solvers (see Sec. 5). Recently, Larsson et al. [16] sampled feasible monomial bases, which can be used in the action-matrix method. In [16] basis sampling was used to minimize the size of the solver. We modified the objective of [16] to maximize for solver stability. Stability sampling generated significantly more numerically stable solvers (see Fig. 3).

Several state-of-the-art methods can rectify from imaged coplanar repeated texture, but these methods assume the pinhole camera model [5, 8–10, 17–19]. A subset of these methods use algebraic constraints induced by the equal-scale invariant of affine-rectified space [8–10]. These methods linearize the rectifying transformation and use the Jacobian determinant to induce local constraints on the imaged scene plane's vanishing line. The Jacobian determinant measures the local change-of-scale of the rectifying transformation. In contrast, the proposed solvers are the first to directly encode the unknown scale of a rectified region as the joint function of the measured region, vanishing line,

	Wildenauer et al. [1]	Antunes et al. [2]	Pritts et al. [15]	Proposed
Feature Type	fitted circles	fitted circles	affine-covariant	affine-covariant
Assumption	3 & 3 parallel lines	3 & 4 parallel lines	2 trans. repeats	4 repeats
Rectification	multi-model	multi-model	direct	direct

Table 1. *Scene Assumptions.* Solvers [1, 2] require distinct sets of parallel scene lines as input and multi-model estimation for rectification. Pritts et al. [15] is restricted to scenes with translational symmetries. The proposed solvers directly rectify from as few as 4 rigidly transformed repeats.

and undistortion parameter. The direct approach eliminates the need for iterative refinement due to the linearization of the rectifying homography.
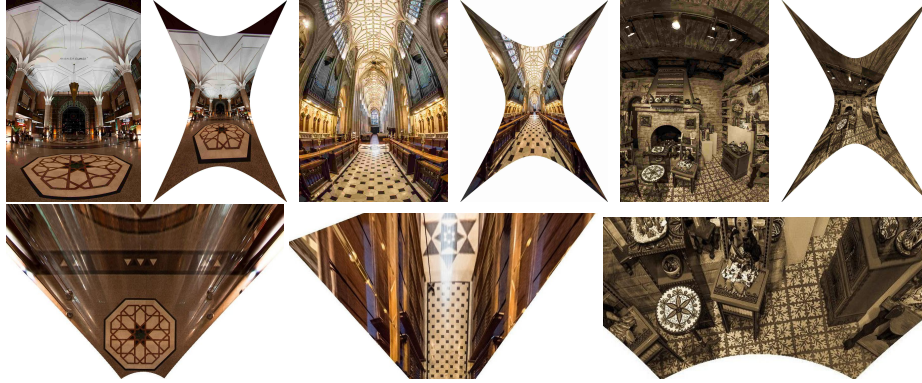
Pritts et al. [20] recover rectification with distortion using a two-step approach: a rectification estimated from a minimal sample using the pinhole assumption is refined by a nonlinear program that incorporates lens distortion. However, even with relaxed thresholds, a robust estimator like RANSAC [3] 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 segmentation of good and bad measurements, as done during the verification step of RANSAC, can give fits that are biased to barrel distortion [14], which degrades rectification accuracy. Pritts et al. [15] were the first to propose minimal solvers that jointly estimate affine rectification and lens distortion, but this method is restricted to scene content with translational symmetries (see Table 1). Furthermore, we show that the conjugate translation solvers of [15] are more noise sensitive than the proposed scale-based solvers (see Figs. 3 and 4).

There are two recent methods that affine-rectify lens distorted images by enforcing the constraint that scene lines are imaged as circles with the division model [1, 2]. The input into these solvers are circles fitted to contours extracted from the image. Sets of circles whose preimages are coplanar parallel lines are used to induce constraints on the division model parameter and vanishing points. These methods require two distinct sets of imaged parallel lines (6 total lines for [1] and 7 for [2]) to estimate rectification, which is a strong scene-content assumption. In addition, these methods must perform a multi-model estimation to label distinct vanishing points as pairwise consistent with vanishing lines. In contrast, the proposed solvers can undistorty and rectify from as few as 4 coplanar repeated local features (see Table 1).

2   Problem Formulation

An affine-rectifying homography $\mathtt{H}$ transforms the image of the scene plane's vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$ to the line at infinity $\mathbf{l}_\infty = \left(0, 0, 1\right)^\top$ [21]. Thus any homography $\mathtt{H}$ satisfying the constraint

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

Fig. 2. *Wide-Angle Results.* Input (top left) is an image of a scene plane. Outputs include the undistorted image (top right) and rectified scene planes (bottom row). The method is automatic.

where $\mathbf{l}$ is an imaged scene plane's vanishing line, is an affine-rectifying homography. Constraint (1) implies that $\mathbf{h}_3 = \mathbf{l}$, and that the line at infinity is invariant to rows $\mathbf{h}_1^\top$ and $\mathbf{h}_2^\top$ of $\mathtt{H}$. Thus the affine-rectification of image point $\mathbf{x}$ to the affine-rectified point $\mathbf{x}'$ is defined as

$$\alpha\mathbf{x}' = \mathtt{H}\mathbf{x} \quad \text{such that } \mathtt{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ & \mathbf{l}^\top & \end{bmatrix} \text{ and } \alpha \neq 0. \tag{2}$$

2.1   Radial Lens Distortion

Rectification as given in (2) 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 $\mathbf{x}$. A common parameterization for radial lens distortion is the one-parameter division model of Fitzgibbon et al. [7], which has the form

$$\mathbf{x} = f(\tilde{\mathbf{x}}, \lambda) = \left(\tilde{x}, \tilde{y}, 1 + \lambda(\tilde{x}^2 + \tilde{y}^2)\right)^\top, \tag{3}$$

where $\tilde{\mathbf{x}} = \left(\tilde{x}, \tilde{y}, 1\right)^\top$ is a feature point with the distortion center subtracted. Substituting (3) into (2) gives

$$\alpha\mathbf{x}' = \left(\alpha x', \alpha y', \alpha\right)^\top = \mathtt{H}\mathbf{x} = \mathtt{H}f(\tilde{\mathbf{x}}, \lambda) =$$
$$\left(\tilde{x}, \tilde{y}, l_1\tilde{x} + l_2\tilde{y} + l_3(1 + \lambda(\tilde{x}^2 + \tilde{y}^2))\right)^\top. \tag{4}$$

The unknown division model parameter λ and vanishing line $\mathbf{l}$ appear only in the homogeneous coordinate α. This property simplifies the solvers derived in Section 3. We also generated a solver using the standard second-order Brown-Conrady model [21–23]; however, these constraints generated a very larger solver with 85 solutions because the radial distortion coefficients appear in the inhomogeneous coordinates.

3   Solvers

The proposed solvers use the invariant that rectified coplanar repeats have equal scales. In Sections 3.1 and 3.2 the equal-scale invariant is used to formulate a system of polynomial constraint equations on rectified coplanar repeats with the vanishing line and radial distortion parameters as unknowns. Radial lens distortion is modeled with the one-parameter division model as defined in Section 2.1. Affine-covariant region detections are used to model repeats since they encode the necessary geometry for scale estimation (see Fig. 1 and Section 4.1). The solvers require 3 points from each detected region to measure the region's scale in the image space. The geometry of an affine-covariant region is uniquely given by an affine frame (see Section 3.1). Three minimal cases exist for the joint estimation of the vanishing line and division-model parameter (see Fig. 1 and Section 3.2). These cases differ by the number of affine-covariant regions needed for each detected repetition. The method for generating the minimal solvers for the three variants is described in Section 3.3. Finally, in Section 3.4, we show that if the distortion parameter is given, then the constraint equations simplify, which results in a small solver for estimating rectification under the pinhole camera assumption.

3.1   Equal Scales Constraint from Rectified Affine-Covariant Regions

The geometry of an oriented affine-covariant region $\mathcal{R}$ is given by an affine frame with its origin at the midpoint of the affine-covariant region detection [24, 25]. The affine frame is typically given as the orientation-preserving homogeneous transformation $\mathtt{A}$ that maps from the right-handed orthonormal frame (normalized descriptor space) to the image space as

$$\begin{bmatrix} \mathbf{y} \ \mathbf{o} \ \mathbf{x} \end{bmatrix} = \mathtt{A} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix},$$

where $\mathbf{o}$ is the origin of the linear basis defined by $\mathbf{x}$ and $\mathbf{y}$ in the image coordinate system [24, 25]. Thus the matrix $\begin{bmatrix} \mathbf{y} \ \mathbf{o} \ \mathbf{x} \end{bmatrix}$ is a parameterization of affine-covariant region $\mathcal{R}$, which we call its *point-parameterization*.

Let $\begin{bmatrix} \tilde{\mathbf{x}}_{i,1} \ \tilde{\mathbf{x}}_{i,2} \ \tilde{\mathbf{x}}_{i,3} \end{bmatrix}$ be the point parameterization of an affine-covariant region $\tilde{\mathcal{R}}_i$ detected in a radially-distorted image. Then by (4) the affine-rectified point parameterization of $\tilde{\mathcal{R}}$ is $\begin{bmatrix} \mathtt{H}f(\tilde{\mathbf{x}}_{i,1}, \lambda) \ \mathtt{H}f(\tilde{\mathbf{x}}_{i,2}, \lambda) \ \mathtt{H}f(\tilde{\mathbf{x}}_{i,3}, \lambda) \end{bmatrix} = \begin{bmatrix} \alpha_{i,1}\mathbf{x}'_{i,1} \ \alpha_{i,2}\mathbf{x}'_{i,2} \ \alpha_{i,3}\mathbf{x}'_{i,3} \end{bmatrix}$, where $\alpha_{i,j} = \mathbf{l}^\top f(\tilde{\mathbf{x}}_{i,j}, \lambda)$. Thus the affine-rectified scale s_i of $\tilde{\mathcal{R}}$ is

$$s_i = \frac{\det\left(\begin{bmatrix} \alpha_{i,1}\mathbf{x}'_{i,1} \ \alpha_{i,2}\mathbf{x}'_{i,2} \ \alpha_{i,3}\mathbf{x}'_{i,3} \end{bmatrix}\right)}{\alpha_{i,1}\alpha_{i,2}\alpha_{i,3}} = \frac{1}{\alpha_{i,1}\alpha_{i,2}\alpha_{i,3}} \cdot \begin{vmatrix} \tilde{x}_{i,1} & \tilde{x}_{i,2} & \tilde{x}_{i,3} \\ \tilde{y}_{i,1} & \tilde{y}_{i,2} & \tilde{y}_{i,3} \\ \alpha_{i,1} & \alpha_{i,2} & \alpha_{i,3} \end{vmatrix} =$$

$$\frac{\alpha_{i,1} \cdot \begin{vmatrix} \tilde{x}_{i,2} & \tilde{x}_{i,3} \\ \tilde{y}_{i,2} & \tilde{y}_{i,3} \end{vmatrix} - \alpha_{i,2} \cdot \begin{vmatrix} \tilde{x}_{i,1} & \tilde{x}_{i,3} \\ \tilde{y}_{i,1} & \tilde{y}_{i,3} \end{vmatrix} + \alpha_{i,3} \cdot \begin{vmatrix} \tilde{x}_{i,1} & \tilde{x}_{i,2} \\ \tilde{y}_{i,1} & \tilde{y}_{i,2} \end{vmatrix}}{\alpha_{i,1}\alpha_{i,2}\alpha_{i,3}}. \tag{5}$$

The numerator in (5) depends only on the distortion parameter λ and l_3 due to cancellations in the determinant. The sign of s_i depends on the handedness of the detected affine-covariant region. See Section 3.6 for a method to use reflected affine-covariant regions with the proposed solvers.

3.2   Eliminating the Rectified Scales

The affine-rectified scale in s_i (5) is a function of the unknown undistortion parameter λ and vanishing line $\mathbf{l} = \left(l_1, l_2, l_3\right)^\top$. A unique solution to (5) can be defined by restricting the vanishing line to the affine subspace $l_3 = 1$ or by fixing a rectified scale, *e.g.*, $s_1 = 1$. The inhomogenous representation for the vanishing line is used since it results in degree 4 constraints in the unknowns λ, l_1, l_2 and s_i as opposed to fixing a rectified scale, which results in complicated equations of degree 7.

Let $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ be repeated affine-covariant region detections. Then the affine-rectified scales of $\tilde{\mathcal{R}}_i$ and $\tilde{\mathcal{R}}_j$ are equal, namely $s_i = s_j$. Thus the unknown rectified scales of a corresponded set of n affine-covariant repeated regions $s_1, s_2, \ldots, s_n$ can be eliminated in pairs, which gives $n-1$ algebraically independent constraints and $\binom{n}{2}$ linearly independent equations. After eliminating the rectified scales, 3 unknowns remain, $\mathbf{l} = \left(l_1, l_2, 1\right)^\top$ and λ, so 3 constraints are needed. There are 3 minimal configurations for which we derive 3 solver variants: (i) 3 affine-covariant region correspondences, which we denote as the 222-configuration; (ii) 1 corresponded set of 3 affine-covariant regions and 1 affine-covariant region correspondence, denoted the 32-configuration; (iii) and 1 corresponded set of 4 affine-covariant regions, denoted the 4-configuration. The notational convention introduced for the input configurations—$(222, 32, 4)$—is extended to the bench of state-of-the-art solvers evaluated in the experiments (see Section 5) to make comparisons between the inputs of the solvers easy. See Fig. 1 for examples of all input configurations and results from their corresponding solver variant, and see Table 2 for a summary of all tested solvers.

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

$$\alpha_{j,1}\alpha_{j,2}\alpha_{j,3} \sum_{k=1}^{3}(-1)^k M_{3,k}\alpha_{i,k} = \alpha_{i,1}\alpha_{i,2}\alpha_{i,3} \sum_{k=1}^{3}(-1)^k M_{3,k}\alpha_{j,k}, \qquad (6)$$

where $M_{i,j}$ is the (i, j)-minor of matrix $\left[\alpha_{i,1}\mathbf{x}'_{i,1}\ \alpha_{i,2}\mathbf{x}'_{i,2}\ \alpha_{i,3}\mathbf{x}'_{i,3}\right]$. Note that the minors $M_{3,\cdot}$ of (6) are constant coefficients. The 222-configuration results in a system of 3 polynomial equations of degree 4 in three unknowns l_1, l_2 and λ; the 32-configuration results in 4 equations of degree 4, and the 4-configuration gives 6 equations of degree 4. Only 3 equations are needed, but we found that for the 32- and 4-configurations that all $\binom{n}{2}$ equations must be used to avoid spurious solutions that arise from vanishing $\alpha_{i,j}$ when the rectified scales are eliminated. For example, if only equations $s_1 = s_2$, $s_1 = s_3$, $s_1 = s_4$ are used for the 4-configuration

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

then λ can be chosen such that $\sum_{k=1}^{3}(-1)^k M_{3,k}\alpha_{1,k} = 0$, and the remaining unknowns l_1 or l_2 chosen such that $\alpha_{1,1}\alpha_{1,2}\alpha_{1,3} = 0$, which gives a 1-dimensional family of solutions. Thus, adding two additional equations removes all spurious solutions. Furthermore, including all equations simplified the elimination template construction.

In principle a solver for the 222-configuration can be used to solve the 32- and 4-configurations. However, this decouples the scales within each group of regions, and there will exist additional solutions that do not satisfy the original constraints.

3.3   Creating the Solvers

We used the automatic generator from Larsson et al. [12] to make the polynomial solvers for the three input configurations $(222, 32, 4)$. The solver corresponding to each input configuration is denoted $\mathtt{H}_{222}\mathtt{l}\lambda$, $\mathtt{H}_{32}\mathtt{l}\lambda$, and $\mathtt{H}_4\mathtt{l}\lambda$, respectively. The resulting elimination templates were of sizes 101×155 (54 solutions), 107×152 (45 solutions), and 115×151 (36 solutions). The equations have coefficients of very different magnitude (e.g. both image coordinates $x_i, y_i \approx 10^3$ and their squares in $x_i^2 + y_i^2 \approx 10^6$). To improve numerical conditioning, we re-scaled both the image coordinates and the squared distances by their average magnitudes. Note that this corresponds to a simple re-scaling of the variables in (λ, l_1, l_2), which is reversed once the solutions are obtained.

Experiments on synthetic data (see Section 5.1) revealed that using the standard GRevLex bases in the generator of [12] gave solvers with poor numerical stability. To generate stable solvers we used the recently proposed basis sampling technique from Larsson et al. [16]. In [16] the authors propose a method for randomly sampling feasible monomial bases, which can be used to construct polynomial solvers. We generated (with [12]) 1,000 solvers with different monomial bases for each of the three variants using the heuristic from [16]. Following the method from Kuang et al. [26], the sampled solvers were evaluated on a test set of 1,000 synthetic instances, and the solvers with the smallest median equation residual were kept. The resulting solvers have slightly larger elimination templates (133×187, 154×199, and 115×151); however, they are significantly more stable. See Section 5.1 for a comparison between the solvers using the sampled bases and the standard GRevLex bases (default in [12]).

3.4   The Fixed Lens Distortion Variant

Finally, we consider the case of known division-model parameter λ. Fixing λ in (6) yields degree 3 constraints in only 2 unknowns l_1 and l_2. Thus only 2 correspondences of 2 repeated affine-covariant regions are needed. The generator of [12] found a stable solver (denoted $\mathtt{H}_{22}\mathtt{l}$) with an elimination template of size 12×21, which has 9 solutions. Basis sampling was unneeded in this case. There is second minimal problem for 3 repeated affine-covariant regions; however, unlike the case of unknown distortion, this minimal problem is equivalent to the $\mathtt{H}_{22}$ variant. It also has 9 solutions and can be solved with the $\mathtt{H}_{22}$ solver by duplicating a region in the input. The $\mathtt{H}_{22}\mathtt{l}$ solver contrasts to the solvers proposed in [8–10] in that it is generated from constraints directly induced by the rectifying homography rather than its linearization.

3.5   Degeneracies

The solvers have two degeneracies. If the vanishing line passes through the image origin $\mathtt{l} = \left(l_1, l_2, 0\right)^\top$, then the radial term in the homogeneous coordinate of (4) vanishes.

If the scene plane is fronto-parallel to the camera and corresponding points from the affine-covariant regions fall on circles centered at the image origin, then the radial distortion is unobservable. The proposed solvers do not have the degeneracy of the $\mathrm{H}_{22}\mathbf{l}s_i$ solver of [10], which occurs if the centroids of the sampled affine-covariant regions are colinear.

3.6   Reflections

In the derivation of (6), the rectified scales s_i were eliminated with the assumption that they had equal signs (see Sec. 3.3). Reflections will have oppositely signed rectified scales; however, reversing the orientation of left-handed affine frames in a simple pre-processing step that admits the use of reflections. Suppose that $\det\left(\left[\tilde{\mathbf{x}}_{i,1}\,\tilde{\mathbf{x}}_{i,2}\,\tilde{\mathbf{x}}_{i,3}\right]\right) < 0$, where $(\tilde{\mathbf{x}}_{i,1}, \tilde{\mathbf{x}}_{i,2}, \tilde{\mathbf{x}}_{i,3})$ is a distorted left-handed point parameterization of an affine-covariant region. Then reordering the point parameterization as $(\tilde{\mathbf{x}}_{i,3}, \tilde{\mathbf{x}}_{i,2}, \tilde{\mathbf{x}}_{i,1})$ results in a right-handed point-parameterization such that $\det\left(\left[\tilde{\mathbf{x}}_{i,3}\,\tilde{\mathbf{x}}_{i,2}\,\tilde{\mathbf{x}}_{i,1}\right]\right) > 0$, and the scales of corresponded rectified reflections will be equal.

4   Robust Estimation

The solvers are used in a LO-RANSAC based robust-estimation framework [27]. Minimal samples are drawn according to the solver variant's requirements (see Table 2 and Fig. 1). Affine rectifications and undistortions are jointly hypothesized by the solver. A metric upgrade is directly attempted using the minimal sample (see [20]), and the consensus set is estimated in the metric-rectified space by verifying the congruence of the basis vectors of the corresponded affine frames. Models with the maximal consensus set are locally optimized in a method similar to [20]. The metric-rectified images are presented in the results.

4.1   Local Features and Descriptors

Affine-covariant region detectors are highly repeatable on the same imaged scene texture with respect to significant changes of viewpoint and illumination [28, 29]. 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 Hesssian-Affine detectors [30, 24]. The affine-covariant regions are given by an affine transform (see Sec. 3), equivalently 3 distinct points, which defines an affine frame in the image space [31]. The image patch local to the affine frame is embedded into a descriptor vector by the RootSIFT transform [32, 33].

4.2   Appearance Clustering and Robust Estimation

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. The RootSIFT descriptors are agglomeratively clustered, which establishes pair-wise tentative correspondences among connected components. Each

	$H_2lus\lambda$	$H_{22}luvs\lambda$	$H_{22}ls_i$	$H_{22}\lambda$	$H_{22}l$	$H_{222}l\lambda$	$H_{32}l\lambda$	$H_4l\lambda$
Reference	[15]	[15]	[10]	[7]				
Rectifies	✓	✓	✓		✓	✓	✓	✓
Undistorts	✓	✓		✓		✓	✓	✓
Motion	trans.	trans.	rigid	rigid[4]	rigid	rigid	rigid	rigid
# regions	2	4	4	4	4	6	5	4
# sols.	2	4	1	18	9	54	45	36
Size	24x26	76x80	4x4	18x18	12x21	133x187	154x199	115x151

Table 2. State-of-the-art vs. proposed (shaded in grey) solvers. The proposed solvers return more solutions, but typically only 1 solution is feasible (see Fig. 5).

appearance cluster has some proportion of its indices corresponding to affine frames that represent the same repeated scene content, which are the *inliers* of that appearance cluster. The remaining affine frames are the *outliers*.

Samples for the minimal solvers are either 2 correspondences of 2 covariant regions (the $H_{22}\cdot$ solvers), a corresponded set of 3 covariant regions (the $H_{32}l\lambda$ solver) and a correspondence of 2 covariant regions, and a corresponded set of 4 covariant regions (the $H_4l\lambda$ solver). An appearance cluster is selected with the probability given by its relative size to the other appearance clusters. The consensus is measured by the number of pairs of affine frames that are mutually consistent with a rigid transform within the appearance group, normalized by the size of each respective group. A non-linear optimizer following [20] is used as the local optimization step.

5   Experiments

The stabilities and noise sensitivities of the proposed solvers are evaluated on synthetic data. We compare the proposed solvers to a bench of 4 state-of-the-art solvers (see Table 2). We apply the denotations for the solvers introduced in Section 3 to all the solvers in the benchmark, *e.g.*, a solver requiring 2 correspondences of 2 affine-covariant regions will be prefixed by H_{22}, while the proposed solver requiring 1 corresponded set of 4 affine-covariant regions is prefixed H_4.

Included are two state-of-the-art single-view solvers for radially-distorted conjugate translations, denoted $H_2lus\lambda$ and $H_{22}luvs\lambda$ [15]; a full-homography and radial distortion solver, denoted $H_{22}\lambda$ [7]; and the change-of-scale solver for affine rectification of [10], denoted $H_{22}ls_i$. The sensitivity benchmarks measure the performance of rectification accuracy by the warp error (see Section 5.1) and the relative error of the division parameter estimate. Stability is measured with respect to the estimated division-model parameter. The $H_{22}\lambda$ solver is omitted from the warp error since the vanishing line is not estimated, and the $H_{22}ls_i$ and $H_{22}l$ solvers are omitted from benchmarks involving lens distortion since the solvers assume a pinhole camera.

---

[4]Correspondences must induce the same rigid transform in the scene plane.
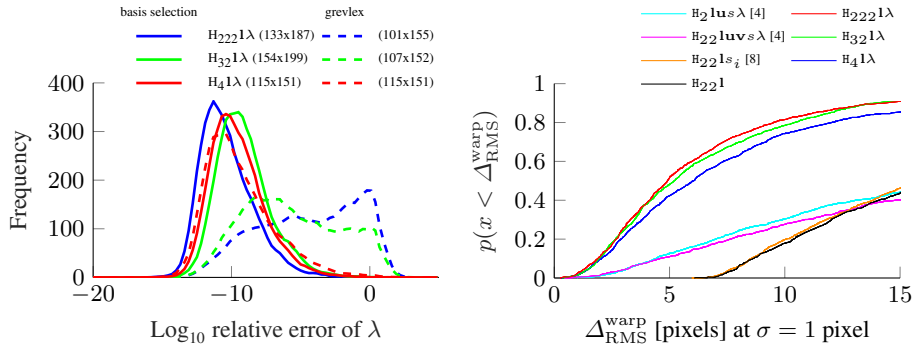
Fig. 3. (left) *Stability study.* Shows that the basis selection method of [16] is essential for stable solver generation. (right) *Proposal Study.* Reports the cumulative distributions of raw warp errors (see Sec. 5.1) for the bench of solvers on 1000 synthetic scenes 1-σ pixel of imaging white noise. The proposed solvers (with distortion) give significantly better proposals than the state of the art.

5.1   Synthetic Data

The performance of the proposed solvers on 1000 synthetic images of 3D scenes with known ground-truth parameters is evaluated. A camera with a random but realistic focal length is randomly placed with respect to a scene plane such that it is mostly in the camera's field-of-view. The image resolution is set to 1000x1000 pixels. Conjugately translated affine frames are generated on the scene plane such that their scale with respect to the scene plane is realistic. The motion is restricted to conjugate translations so that [15] can be included in the benchmark. Fig. A.1 of the supplemental includes experiments for rigidly transformed affine frames. The modeling choice reflects the use of affine-covariant region detectors on real images (see Section 3). The image is distorted according to the division model. For the sensitivity experiments, isotropic white noise is added to the distorted affine frames at increasing levels. Performance is characterized by the relative error of the estimated distortion parameter and by the warp error, which measures the accuracy of the affine-rectification.

Warp Error Since the accuracy of scene-plane rectification is a primary concern, the warp error for rectifying homographies proposed by Pritts et al. [34] is reported, which we extend to incorporate the division model for radial lens distortion [7]. 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 $\hat{\mathbf{H}}$ and then imaging the rectified plane by the ground-truth camera. Ideally, the ground-truth camera images the rectified points onto $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$. There is an affine ambiguity, denoted $\mathbf{A}$, between $\hat{\mathbf{H}}$ and the ground-truth camera matrix
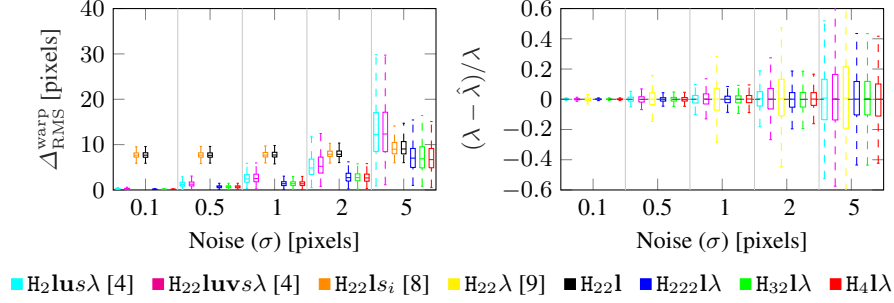
Fig. 4. *Sensitivity Benchmark.* Comparison of two error measures after 25 iterations of a simple RANSAC for different solvers with increasing levels of white noise added to the affine frame correspondences. (left) Reports the warp error as $\Delta^{\mathrm{warp}}_{\mathrm{RMS}}$ and (right) Reports the relative error of the estimated division model parameter. The proposed solvers are significantly more robust.

P. The ambiguity is estimated during computation of the warp error,

$$\Delta^{\mathrm{warp}} = \min_{\mathtt{A}} \sum_i d^2(\tilde{\mathbf{x}}, f^d(\mathtt{PA}\hat{\mathtt{H}} f(\tilde{\mathbf{x}}, \hat{\lambda})), \hat{\lambda}), \tag{7}$$

where $d(\cdot, \cdot)$ is the Euclidean distance, f^d is the inverse of the division model (the inverse of (3)), and $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$ are the imaged grid points of the scene-plane tessellation. The root mean square warp error for $\{\tilde{\mathbf{x}}_i\}_{i=1}^{100}$ is reported and denoted as $\Delta^{\mathrm{warp}}_{\mathrm{RMS}}$. Note that the vanishing line is not directly estimated by the solver $\mathtt{H}_{22}\lambda$ of [7], so it is not reported.

Numerical Stability The stability study compares the solver variants generated using the standard GRevLex bases versus solvers generated with bases chosen by basis sampling using [16] (see Section 3.3). The generator of Larsson et al. [12] et al. was used to generate both sets of solvers. Stability is measured by the relative error of the estimated division model parameter for noiseless affine-frame correspondences across realistic synthetic scenes, which are generated as described in the introduction of Section 5.1. The ground-truth parameter of the division model λ is drawn uniformly from the interval $[-8, 0.5]$. As a reference, the normalized division parameter $\lambda = -4$ is typical for wide field-of-view cameras like the GoPro, where the image is normalized by $1/(\text{width} + \text{height})$. Fig. 3 (left) reports the histogram of $\log_{10}$ relative error of the estimates of the division model parameter, and Fig. 3 shows that the basis selection method of [16] significantly improves the stability of the generated solvers. The basis-sampled solvers are used for the remainder of the experiments.

Noise Sensitivity The proposed and state-of-the-art solvers are tested with increasing levels of white noise added to the point parameterizations (see Section 3.1) of the affine-covariant region correspondences, which are conjugately translated (see [15, 35]). The amount of white noise is given by the standard-deviation of a zero-mean isotropic Gaussian distribution, and the solvers are tested at noise levels of $\sigma \in \{0.1, 0.5, 1, 2, 5\}$. The
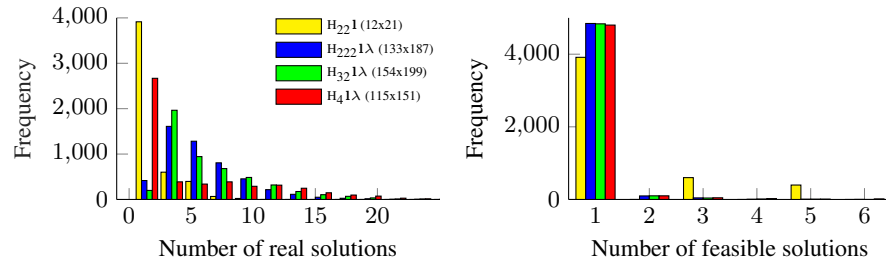
Fig. 5. (left) *Real Solutions.* The histogram of real solutions for the proposed solvers. (right) *Feasible Solutions.* There is typically only 1 feasible real solution.

ground-truth normalized division model parameter is set to $\lambda = -4$, which is typical for GoPro-type imagery in normalized image coordinates.

The proposal study in the right panel of Fig. 3 shows that for 1-pixel white noise, the proposed solvers—$\mathrm{H}_{222}\mathrm{l}\lambda, \mathrm{H}_{32}\mathrm{l}\lambda$ and $\mathrm{H}_4\mathrm{l}\lambda$—give significantly more accurate estimates than the state-of-the-art conjugate translation solvers of [15]. If 5 pixel RMS warp error is fixed as a threshold for a good model proposal, then 50% of the models given by the proposed solvers are good versus less than 20% by [15]. The proposed $\mathrm{H}_{22}\mathrm{l}$ solver and $\mathrm{H}_{22}\mathrm{l}s_i$ [10] both give biased proposals since they don't estimate lens distortion.

For the sensitivity study in Fig. 4, the solvers are wrapped by a basic RANSAC estimator, which minimizes the RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ 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. As shown in Fig. 4, the proposed solvers— $\mathrm{H}_{222}\mathrm{l}\lambda, \mathrm{H}_{32}\mathrm{l}\lambda$ and $\mathrm{H}_4\mathrm{l}\lambda$—give the most accurate lens distortion and rectification estimates. The proposed solvers are superior to the state of the art at all noise levels. The proposed distortion-estimating solvers give solutions with less than 5-pixel RMS warp error $\Delta_{\mathrm{RMS}}^{\mathrm{warp}}$ 75% of the time and estimate the correct division model parameter more than half the time at the 2-pixel noise level. The fixed-lens distortion solvers $\mathrm{H}_{22}\mathrm{l}$ and $\mathrm{H}_{22}\mathrm{l}s_i$ of [10] give biased solutions since they assume the pinhole camera model.

Feasible Solutions and Runtime Fig. 5 (left) shows the number of real solutions given by the proposed solvers for 5000 synthetic scenes, and Fig. 5 (right) shows the subset of feasible solutions as defined by the estimated normalized division-model parameter solution falling in the interval $[-8, 0.5]$. All solutions are considered feasible for the $\mathrm{H}_{22}\mathrm{l}$ solver. Fig. 5 (right) shows that in 97% of the scenes only 1 solution is feasible, which means that nearly all incorrect solutions can be quickly discarded. The runtimes of the MATLAB implementation of the solvers on a standard desktop are 2 ms for $\mathrm{H}_{222}\mathrm{l}\lambda$, 2.2 ms for $\mathrm{H}_{32}\mathrm{l}\lambda$, 1.7 ms for $\mathrm{H}_4\mathrm{l}\lambda$, and 0.2 ms for $\mathrm{H}_{22}\mathrm{l}$.

5.2   Real Images

The field-of-view experiment of Fig. 7 evaluates the proposed $\mathrm{H}_{222}\mathrm{l}\lambda$ solver on real images taken with narrow, medium, and wide-angle lenses as well as a fisheye lens.
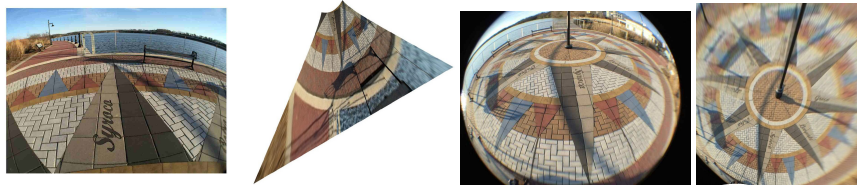
Fig. 6. (left pair) The pre-image of the waterfront is a circle, which violates the plumb-line assumption. (right pair) A reasonable rectification is estmiated by the proposed method even with a fish-eye lens.

Images with diverse scene content were chosen. Fig. 7 shows that the $H_{222}l\lambda$ gives accurate rectifications for all lens types. Additional results for wide-angle lenses are included in Section A of the supplemental. Fig. 8 compares the proposed $H_{222}l\lambda$ and $H_{22}l$ solvers to the state-of-the-art solvers on images with increasing levels of radial lens distortion (top to bottom) that contain either translated or rigidly-transformed coplanar repeated patterns. Only the proposed $H_{222}l\lambda$ accurately rectifies on both pattern types and at all levels of distortion. The results are after a local optimization and demonstrate that the method of Pritts et al. [20] is unable to accurately rectify without a good initial guess at the lens distortion. The proposed fixed-distortion solver $H_{22}l$ gave a better rectification than the change-of-scale solver of Chum et al. [10]. Fig. 6 shows the rectifications of a deceiving picture of a landmark taken by wide-angle and fisheye lenses. From the wide-angle image it is not obvious which lines are really straight in the scene making undistortion with the plumb-line constraint difficult.
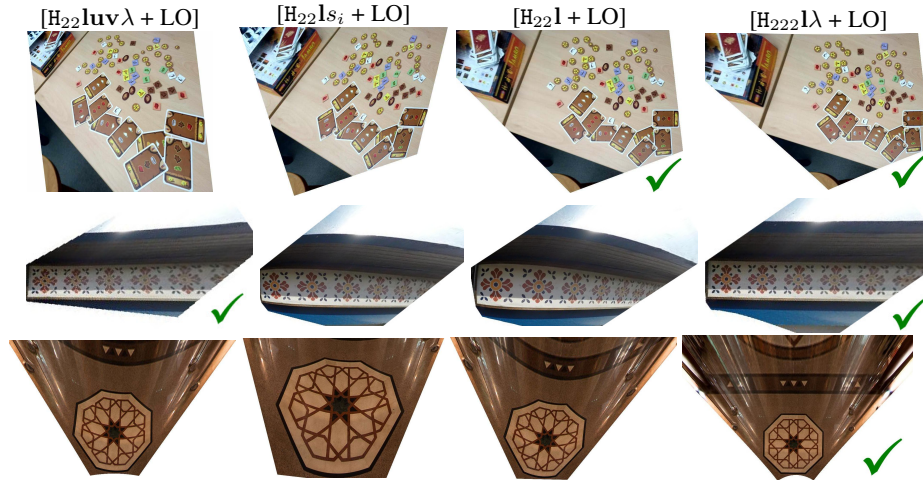
6   Conclusion

This paper proposes solvers that extend affine-rectification to radially-distorted images that contain essentially arbitrarily repeating coplanar patterns. Synthetic experiments show that the proposed solvers are more robust to noise with respect to the state of the art while being applicable to a broader set of image content. The paper demonstrates that robust solvers can be generated with by the basis selection method of [16] by maximizing for numerical stability. Experiments on difficult images with large radial distortions confirm that the solvers give high-accuracy rectifications if used inside a robust estimator. By jointly estimating rectification and radial distortion, the proposed minimal solvers eliminate the need for sampling lens distortion parameters in RANSAC.

<table>
<tr><td>[narrow]</td><td>[medium]</td><td>[wide]</td><td>[fisheye]</td></tr>
</table>
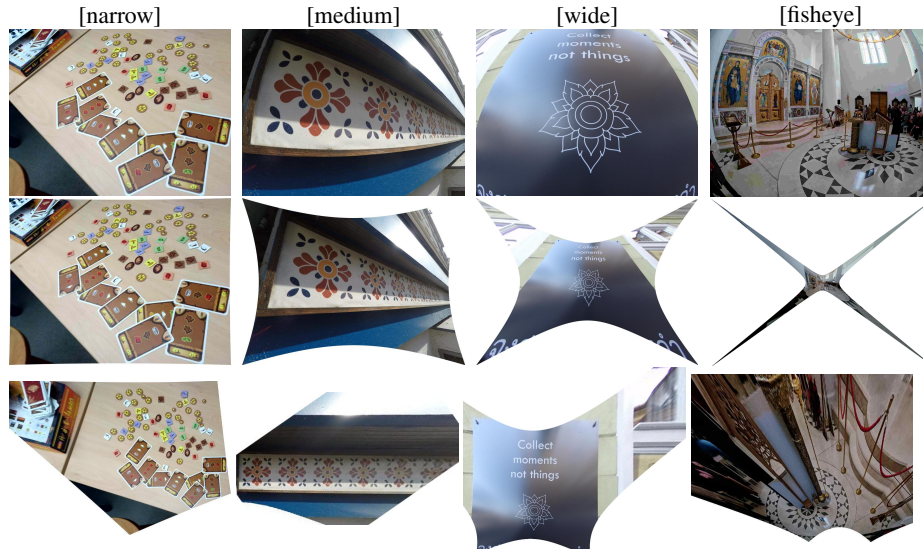
Fig. 7. *Field-of-View Study.* The proposed solver $\mathtt{H}_{222}\mathbf{l}\lambda$ gives accurate rectifications across all fields-of-view: (left-to-right) Android phone, GoPro Hero 4 at the medium- and wide-FOV settings, and a Panasonic DMC-GM5 with a Samyang 7.5mm fisheye lens. The outputs are the undistorted (middle row) and rectified images (bottom row).

<table>
<tr><td>[$\mathtt{H}_{22}\mathbf{luv}\lambda$ + LO]</td><td>[$\mathtt{H}_{22}\mathbf{l}s_i$ + LO]</td><td>[$\mathtt{H}_{22}\mathbf{l}$ + LO]</td><td>[$\mathtt{H}_{222}\mathbf{l}\lambda$ + LO]</td></tr>
</table>

Fig. 8. *Solver Comparison.* The state-of-the art solvers $\mathtt{H}_{22}\mathbf{luv}\lambda$ and $\mathtt{H}_{22}\mathbf{l}s_i$ [10, 15] are compared with the proposed solvers $\mathtt{H}_{222}\mathbf{l}\lambda$ and $\mathtt{H}_{22}\mathbf{l}$ on images containing either translated or rigidly-transformed coplanar repeated patterns with increasing amounts of lens distortion. (top) small distortion, rigidly-transformed; (middle) medium distortion, translated; (bottom) large distortion, rigidly-transformed. Accurate rectifications for all images is only given by the proposed $\mathtt{H}_{222}\mathbf{l}\lambda$.

References

1. Wildenauer, H., Micusík, B.: Closed form solution for radial distortion estimation from a single vanishing point. In: BMVC. (2013)
2. Antunes, M., Barreto, J.P., Aouada, D., Ottersten, B.: Unsupervised vanishing point detection and camera calibration from a single manhattan image with radial distortion. In: CVPR. (2017)
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. CACM **24** (1981) 381–395
4. Funk, C., Lee, S., Oswald, M.R., Tsogkas, S., Shen, W., Cohen, A., Dickinson, S., Y., L.: 2017 ICCV challenge: Detecting symmetry in the wild. In: ICCV Workshop. (2017)
5. Lukáč, M., Sýkora, D., Sunkavalli, K., Shechtman, E., Jamriška, O., Carr, N., Pajdla, T.: Nautilus: Recovering regional symmetry transformations for image editing. ACM Trans. Graph. **36** (2017) 108:1–108:11
6. Wu, C., Frahm, J.M., Pollefeys, M.: Repetition-based dense single-view reconstruction. In: CVPR. (2011)
7. Fitzgibbon, A.W.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: CVPR. (2001)
8. Ohta, T., Maenobu, K., Sakai, T.: Obtaining surface orientation from texels under perspective projection. In: IJCAI. (1981)
9. Criminisi, A., Zisserman, A.: Shape from texture: homogeneity revisited. In: BMVC. (2000)
10. Chum, O., Matas, J.: Planar affine rectification from change of scale. In: ACCV. (2010)
11. Kukelova, Z., Bujnak, M., Pajdla, T.: Automatic generator of minimal problem solvers. In: ECCV. (2008)
12. Larsson, V., Åström, K., Oskarsson, M.: Efficient solvers for minimal problems by syzygy-based reduction. In: CVPR. (2017)
13. Larsson, V., Åström, K., Oskarsson, M.: Polynomial solvers for saturated ideals. In: ICCV. (2017)
14. Kukelova, Z., Heller, J., M., B., Pajdla, T.: Radial distortion homography. In: CVPR. (2015)
15. Pritts, J., Kukelova, Z., Larsson, V., Chum, O.: Radially-distorted conjugate translations. In: CVPR. (2018)
16. Larsson, V., Oskarsson, M., Astrom, K., Wallis, A., Kukelova, Z., Pajdla, T.: Beyond grobner bases: Basis selection for minimal solvers. In: CVPR. (2018)
17. Ahmad, S., Cheong, L.F.: Robust detection and affine rectification of planar homogeneous texture for scene understanding. International Journal of Computer Vision **126** (2018) 822–854
18. Aiger, D., Cohen-Or, D., Mitra, N.J.: Repetition maximization based texture rectification. Computer Graphics Forum **31** (2012) 439–448
19. Zhang, Z., Ganesh, A., Liang, X., Ma, Y.: TILT: transform invariant low-rank textures. International Journal of Computer Vision **99** (2012) 1–24
20. Pritts, J., Chum, O., Matas, J.: Detection, rectification and segmentation of coplanar repeated patterns. In: CVPR. (2014)
21. Hartley, R.I., Zisserman, A.W.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)
22. Brown, D.C.: Decentering distortion of lenses. Photometric Engineering **32** (1966) 444–462
23. Conrady, A.: Decentering lens systems. Monthly Notices of the Royal Astronomical Society **79** (1919) 384–390
24. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International journal of computer vision **60** (2004) 63–86

25. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2008)
26. Kuang, Y., Åström, K.: Numerically stable optimization of polynomial solvers for minimal problems. In: ECCV. (2012)
27. Chum, O., Matas, J., Obdržálek, v.: Enhancing RANSAC by generalized model optimization. In: ACCV. (2004)
28. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence **27** (2005) 1615–1630
29. Mishkin, D., Radenovic, F., Matas, J.: Repeatability is not enough: Learning affine regions via discriminability. In: Proceedings of ECCV. (2018)
30. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC. (2002)
31. Obdržálek, Š., Matas, J.: Object recognition using local affine frames on distinguished regions. In: BMVC. (2002)
32. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: CVPR. (2012)
33. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60** (2004) 91–110
34. Pritts, J., Rozumnyi, D., Kumar, M.P., Chum, O.: Coplanar repeats by energy minimization. In: BMVC. (2016)
35. Schaffalitzky, F., Zisserman, A.: Geometric grouping of repeated elements within images. In: BMVC. (1998)
36. Kannala, J., Brandt, S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE transactions on pattern analysis and machine intelligence **28** (2006) 1335–1340