Minimal Solvers for Single-View Lens-Distorted Camera Auto-Calibration

Yaroslava Lochman[1,2]   Oles Dobosevych[1]   Rostyslav Hryniv[1]   James Pritts[2]

lochman@ucu.edu.ua   dobosevych@ucu.edu.ua   rhryniv@ucu.edu.ua   jbpritts@fb.com

[1]Machine Learning Lab, Ukrainian Catholic University in Lviv
[2]Facebook Reality Labs in Pittsburgh

(a) Input   (b) Undistorted   (c) Manhattan Planes Rectified

Figure 1: (a) Input is an image. (b) Lens distortion corrected, and (c) Scene planes rectified. The method is fully automatic.

Abstract

This paper proposes minimal solvers that use combinations of imaged translational symmetries and parallel scene lines to jointly estimate lens undistortion with either affine rectification or focal length and absolute orientation. We use constraints provided by orthogonal scene planes to recover the focal length. We show that solvers using feature combinations can recover more accurate calibrations than solvers using only one feature type on scenes that have a balance of lines and texture. We also show that the proposed solvers are complementary and can be used together in a RANSAC-based estimator to improve auto-calibration accuracy. State-of-the-art performance is demonstrated on a standard dataset of lens-distorted urban images. The code is available at `https://github.com/ylochman/single-view-autocalib`.

1. Introduction

Imaged scene plane rectification and single-view camera auto-calibration are closely related computer-vision tasks. Both tasks are ill-posed single-view geometry estimation problems that are further complicated if the input image is distorted [1, 23]. In the presence of imaging noise, good feature coverage over large parts of the image is necessary to observe the joint effects of perspective warp and lens distortion. State-of-the-art techniques for auto-calibrating or rectifying lens-distorted images use either circular arcs fitted to edge detections or covariant region detections as inputs [1, 33, 22, 21]. Complementary features can provide measurements on image regions that lack either texture or lines. Each feature type—circular arc or covariant region—has distinct advantages. Circular arcs provide accurate measurements of imaged scene lines. However, it's difficult to group arcs that are the images of parallel scene lines by appearance. While regions do not give the accuracy of circular arcs, discriminative embeddings exist that can be used to cluster imaged coplanar repeats [2, 3, 9, 15]. Furthermore, multiple point correspondences can be extracted from one region correspondence. We propose solvers that combine the best of both worlds and demonstrate that solvers using complementary feature types can extend high-accuracy rectification and auto-calibration to challenging highly-distorted images with diverse scene content (see Fig. 1).

A solver whose inputs are sets of different geometric primitives is called a *hybrid solver* or *mixed solver* [5]. Minimal solvers constructed from points and lines were used in pose estimation [24, 14] and combinations of points and planes were explored in SLAM [27]. The proposed hybrid solvers are the first affine-rectifying and auto-calibrating solvers to admit complementary geometric primitives—combinations of point correspondences provided by a coplanar repeated region and circular arcs fitted to the linked edge extractions of imaged parallel lines. In addition, we propose two solvers whose inputs are circular arcs in a novel configuration, which provides an additional feature sampling flexibility.

The affine-rectifying solvers can be adapted for metric rectification and auto-calibration, which are the tasks evaluated in this paper. The relative angle between the preimage of input features is assumed known, which provides sufficient constraints for auto-calibration. Right angles are cho-

| | Coplanar Configuration | | | Manhattan Configuration | | |
| | Inputs | | Outputs | Inputs | | Outputs |
	PC	CA		PC	CA	
5CA [33]	0	3+2	$\{1\,\mathrm{VP}, \lambda\} \to 1\,\mathrm{VP} \to f$	0	3+1+1	$\{1\,\mathrm{VP}, \lambda\} \to \{2\,\mathrm{VP}, f\}$
7CA [1]	0	4+3	$2\,\mathrm{VP} \to \{\lambda, \mathbf{c}\} \to f$	-	-	-
4PC (EVP) [19]	2+2	0	$\{2\,\mathrm{VP}, \lambda, \mathbf{l}\} \to f$	-	-	-
4PC+2CA	2+2	2	$\{3\,\mathrm{VP}, \lambda, \mathbf{l}\} \to f$	3	1+1	$\{1\,\mathrm{VP}, \lambda\} \to \{2\,\mathrm{VP}, f\}$
2PC+4CA	2	2+2	$\{3\,\mathrm{VP}, \lambda, \mathbf{l}\} \to f$	2	2+2	$\{3\,\mathrm{VP}, \lambda, f\}$
5CA*	0	3+2	$\{1\,\mathrm{VP}, \lambda\} \to 1\,\mathrm{VP} \to f$	0	3+1+1	$\{1\,\mathrm{VP}, \lambda\} \to \{2\,\mathrm{VP}, f\}$
6CA	0	2+2+2	$\{3\,\mathrm{VP}, \lambda, \mathbf{l}\} \to f$	0	2+2+2	$\{3\,\mathrm{VP}, \lambda, f\}$

Table 1: Inputs and outputs of the state-of-the-art vs. the proposed solvers (shaded in grey). The inputs are counted by the number of corresponded features required for each estimated vanishing point. Denotations are PC for point correspondence, CA for circular arcs, VP for vanishing point, λ for the division model parameter, $\mathbf{c}$ for the distortion center, $\mathbf{l}$ for the vanishing line, and f for the focal length. A sets of outputs is jointly recovered by the solver, and right arrows indicate a chain of estimates. The output at each step depends on the configuration of VPs: either the VPs are coplanar, or they are oriented as a Manhattan frame in the scene.

sen because they are the most common in man-made scenes. Three mutually orthogonal directions in the scene are used to define a linear basis called the *Manhattan frame.* The presence of a Manhattan frame in the scene is assumed by the auto-calibrating solvers.

All proposed solvers are derived from common constraints and are related by a unified derivation. The derivation uses elementary techniques from projective geometry, which makes solver generation straightforward. The solvers are fast, stable and robust, which makes them good candidates to be used in RANSAC-based estimators [7].

The solvers are also used as an ensemble in a RANSAC estimator that samples combinations of arcs and point correspondences according to the required input of the invoked solver. Solvers are invoked with different frequencies depending on the content of the scene. The use of multiple solvers that have different input types is similar to the hybrid RANSAC approach introduced in [5, 13]. We show that sampling combinations of regions and arcs improves auto-calibration accuracy over using either only regions or arcs.

2. Related Work

The state of the art has extended scene-plane rectification and single-view auto-calibration to lens-distorted images. The common choice of parameterization for lens undistortion is the division model. It is preferred in ill-posed settings since it has only one parameter and can effectively model a wide range of radial lens undistortions [8].

Antunes *et al.* [1] and Wildenauer *et al.* [33] are two methods that recover vanishing points from distorted images by using the constraint that parallel scene lines are imaged as circles intersecting at their distorted vanishing point under the division model [4, 8, 26, 31]. The same

constraint is used for the proposed solvers. Sets of circular arcs whose preimages are parallel lines are used to induce constraints on the division-model parameter and vanishing point. Vanishing points are recovered, and auto-calibration is estimated by assuming that vanishing points correspond to imaged Manhattan frame directions. Antunes *et al.* [1] require a set of four circular arcs for the first vanishing point and three for the second, while Wildenauer *et al.* [33] requires three for the first and two for the second or three for the first, one for the second, and one for the third.

Pritts *et al.* [19, 20, 22, 21] proposed a suite of solvers that can jointly undistort and affinely-rectify from the imaged translation directions of coplanar repeated scene texture. The solvers directly estimate the vanishing line of the scene plane but also return the vanishing directions of the imaged translations that are consistent with the recovered vanishing line. The solver variant of [19, 21] requiring two correspondences of covariant regions can be directly used for auto-calibration, if the recovered vanishing points are imaged Manhattan frame directions.

See Table 1 for a comparison of the proposed solvers to the state of the art. The solver notation refers the input configurations, e.g., **4PC** means four point correspondences, while **6CA** means six circular arcs are required.

3. Preliminaries

We assume the division model [8] of lens undistortion

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

where λ encodes the magnitude and type (by sign) of radial distortion, $\tilde{\mathbf{x}} = \left(\tilde{x}, \tilde{y}, 1\right)^\top$ is a homogeneous image point with radius $\tilde{r}^2 = \tilde{x}^2 + \tilde{y}^2$, and the origin is at the distortion center. We assume an orthogonal raster with unit aspect ratio and fix the distortion center and the principal
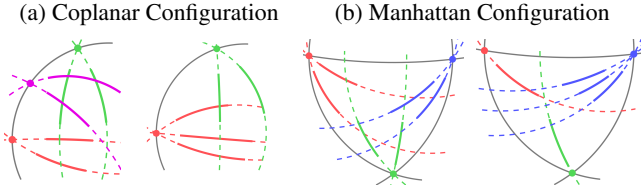
Figure 2: Input configurations of the proposed solvers. The vanishing points are constructed from either circular arcs or point correspondences. Inputs are color coded with respect to their corresponding vanishing points that are (a) coplanar (see Sec. 4.1) or (b) mutually orthogonal (see Sec. 4.2).

point at the image center. In this setting the camera's intrinsic matrix is $\mathtt{K} = \mathrm{diag}(f, f, 1)$, where f is the focal length in pixels. The projection of a homogeneous scene point $\mathbf{X} = (X, Y, Z, 1)^\top$ onto $\tilde{\mathbf{x}}$ is as following

$$\gamma g(\tilde{\mathbf{x}}, \lambda) = \mathtt{K}\left[\mathtt{R} \mid \mathbf{t}\right]\mathbf{X}, \tag{2}$$

where $\mathtt{R}$ is a 3×3 rotation matrix, $\mathbf{t}$ is a translation vector.

4. Minimal Solvers

All the proposed solvers use the invariant that imaged parallel scene plane lines intersect at the vanishing point of their imaged translation direction. A vanishing point $\mathbf{u}(\lambda)$ is constructed as the meet of the undistorted images $\mathbf{m}(\lambda)$ and $\mathbf{m}'(\lambda)$ of parallel lines,

$$\mathbf{u}(\lambda) = \mathbf{m}(\lambda) \times \mathbf{m}'(\lambda). \tag{3}$$

The undistorted images of parallel lines can be constructed from distorted measurements under (1). There are two different parametric forms of $\mathbf{m}(\lambda)$ and $\mathbf{m}'(\lambda)$ depending on the type of measurements used: either circular arcs or point correspondences.

The relative orientation of the detected vanishing points in not known apriori. The solvers handle two cases: either the vanishing points are coplanar, or they are the image of a Manhattan frame (see Fig. 2).

4.1. Coplanar Vanishing Points

The joint rectifying solvers share a common derivation based on the invariant that vanishing points are incident to the scene plane's vanishing line. The vanishing point-vanishing line incidence equation is $\mathbf{u}^\top\mathbf{l} = 0$. There are four unknowns to be recovered, 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, three scalar constraint equations of the form $\mathbf{u}_i(\lambda)^\top\mathbf{l} = 0$ are needed, where $\{\mathbf{u}_i\}_{i=1}^3$ are three distinct vanishing points constructed as in (3). With a matrix $\mathtt{U}(\lambda)$ formed
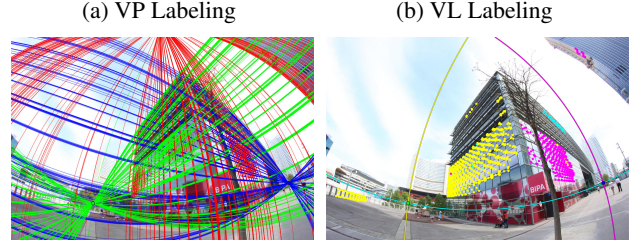

Figure 3: (a) The distorted vanishing points of the imaged Manhattan frame are colored RGB. Circular arcs and region correspondences are colored w.r.t. their assigned VPs. (b) Imaged vanishing lines (VL) are colored CMY Covariant regions are colored w.r.t. their assigned VLs.

by the vanishing points $\{\mathbf{u}_i\}$, the point-line incidence constraints can be concisely written as a homogeneous matrix-vector equation,

$$\mathtt{U}(\lambda)\mathbf{l} = \begin{bmatrix} (\mathbf{m}_1(\lambda) \times \mathbf{m}'_1(\lambda))^\top \\ (\mathbf{m}_2(\lambda) \times \mathbf{m}'_2(\lambda))^\top \\ (\mathbf{m}_3(\lambda) \times \mathbf{m}'_3(\lambda))^\top \end{bmatrix}\mathbf{l} = \mathbf{0}. \tag{4}$$

The system of equations (4) has a non-trivial solution $\mathbf{l}$ only if $\mathtt{U}(\lambda)$ is singular, which generates the scalar constraint equation $\det \mathtt{U}(\lambda) = 0$ on λ. The parameterization for lines $\mathbf{m}_i(\lambda)$ and $\mathbf{m}'_i(\lambda)$ used by all solver variants results in a quartic equation in λ, which can be solved in closed form. After recovering λ, the null space of $\mathtt{U}$ is computed, which gives the vanishing line $\mathbf{l}$. The system of equations in (4) is agnostic to the construction method for the $\mathbf{u}_i$, which gives a unified way to generate rectifying solvers that use different feature types.

Auto-Calibration Upgrade The Manhattan frame directions defined by the standard basis vectors $\{\mathbf{e}_i\}_{i=1}^3$ are imaged to the distorted vanishing points $\{\tilde{\mathbf{u}}_i\}_{i=1}^3$ by (2)

$$\mathbf{u}_i = \gamma g(\tilde{\mathbf{u}}_i, \lambda) = \mathtt{K}\mathtt{R}\mathbf{e}_i, \quad i \in \{1 \ldots 3\}. \tag{5}$$

See an example of detected imaged Manhattan frame in Fig. 3. Two finite vanishing points $\mathbf{u}_1 = (u_{x1}, u_{y1}, u_{w1})^\top$, $\mathbf{u}_2 = (u_{x2}, u_{y2}, u_{w2})^\top$ that are undistorted images of two Manhattan frame directions are sufficient to recover the focal length [32]

$$f = \sqrt{\frac{u_{x1}u_{x2} + u_{y1}u_{y2}}{-u_{w1}u_{w2}}}. \tag{6}$$

The relative orientation of the camera with respect to the Manhattan frame is then computed as following

$$\mathtt{R} = \left[\frac{\mathtt{K}^{-1}\mathbf{u}_1}{\|\mathtt{K}^{-1}\mathbf{u}_1\|} \quad \frac{\mathtt{K}^{-1}\mathbf{u}_2}{\|\mathtt{K}^{-1}\mathbf{u}_2\|} \quad \frac{\mathtt{K}^\top(\mathbf{u}_1 \times \mathbf{u}_2)}{\|\mathtt{K}^\top(\mathbf{u}_1 \times \mathbf{u}_2)\|} \right]. \tag{7}$$

A conjugate rotation $\mathtt{K}\mathtt{R}^\top\mathtt{K}^{-1}$ is used to metrically rectify the Manhattan planes (see Figs. 1, 4, and 7).
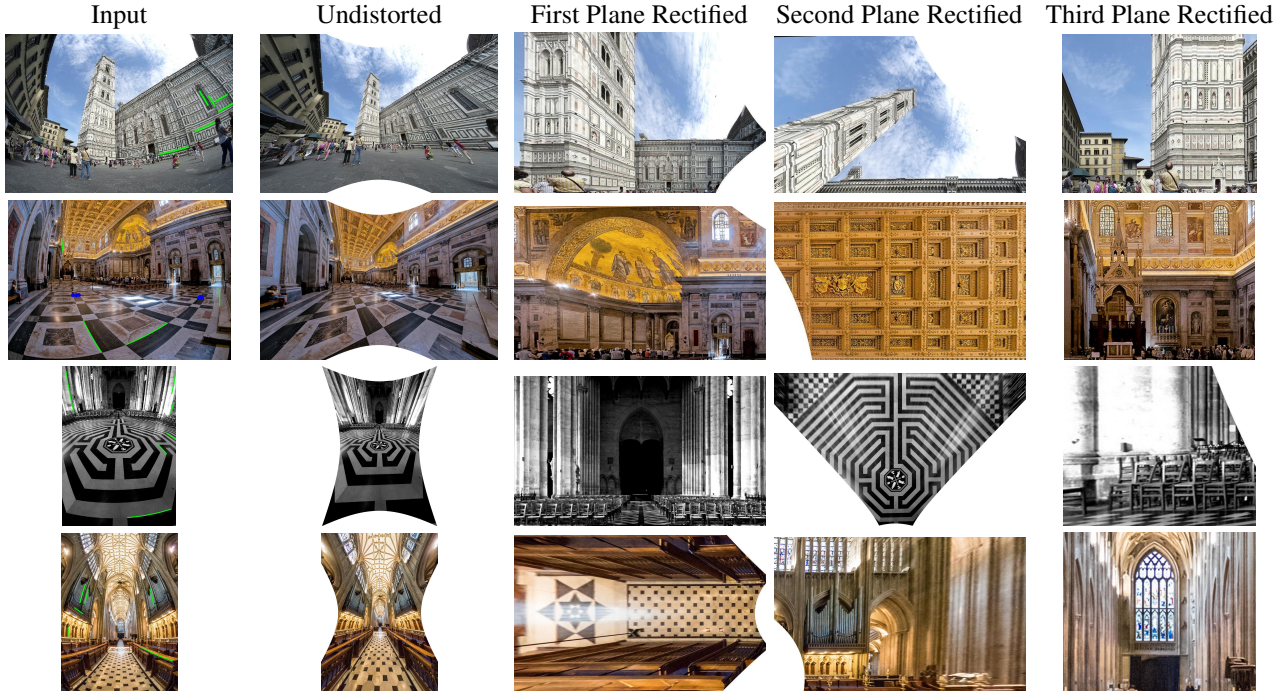
Figure 4: Auto-calibration results on wide-angle imagery. The scene planes oriented with the Manhattan frame are rectified. The minimal sample —green circles and blue regions—of the returned solution is depicted on the input image.

4.2. Orthogonal Vanishing Points

The joint auto-calibrating solvers use an invariant that the preimages of the vanishing points are mutually orthogonal scene directions. The orthogonality constraint is written as $\mathbf{u}_i^\top \omega \mathbf{u}_j = 0$, where $\omega = \mathrm{diag}(1/f^2, 1/f^2, 1)$ is the image of an absolute conic. The functional forms of the vanishing points $\mathbf{u}(\lambda)$ are constructed by (3). There are two unknowns to be recovered — λ and f. Thus, two scalar constraint equations of the form $\mathbf{u}_i^\top \omega \mathbf{u}_j = 0$ are needed, which means that three vanishing points $\{\,\mathbf{u}_i\,\}_{i=1}^3$ are needed. Rewriting the constraint equations in the matrix form gives

$$\begin{bmatrix} \mathbf{u}_1(\lambda) \odot \mathbf{u}_2(\lambda) \\ \mathbf{u}_1(\lambda) \odot \mathbf{u}_3(\lambda) \\ \mathbf{u}_2(\lambda) \odot \mathbf{u}_3(\lambda) \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ f^2 \end{pmatrix} = \mathbf{0}, \tag{8}$$

where $\odot$ is the entrywise product. By substituting f from the first row into the second and the third rows we obtain the system of equations only in λ. The coordinates of a vanishing point constructed from circular arcs or from point correspondences are either linear or quadratic in λ, which gives two polynomial equations in λ of order six. We solve the resulting system for λ, recover f from the linear system (8), and compute R by (7). Similarly to (4), the system of equations in (8) is agnostic to the construction method for the $\mathbf{u}_i$.

4.3. Coincident Vanishing Points

The system of equations (4) is trivially singular if two of the vanishing points from $\{\,\mathbf{u}_i\,\}_{i=1}^3$ are coincident. This occurs if two vanishing points are constructed by drawing two pairs of imaged lines that are mutually parallel in the scene. Similarly, the orthogonality constraint equations (8) are inconsistent if two of the vanishing points from $\{\,\mathbf{u}_i\,\}_{i=1}^3$ are coincident. Vanishing point coincidence can be used to place a constraint on the division model parameter

$$\mathbf{u}_i(\lambda) \times \mathbf{u}_j(\lambda) = \mathbf{0}. \tag{9}$$

The coordinates of a vanishing point are either linear or quadratic in λ, which gives two cubic equations and one quadratic equation in λ in the system of (9). After solving for λ we back-substitute and recover $\mathbf{l}$ from (4), if the vanishing points are collinear, or solve for f and R using (8), if they are the image of a Manhattan frame.

4.4. Solver Generation

The solvers can be generated by constructing vanishing points $\{\mathbf{u}_i\}$ from circular arcs or vanishing points $\{\mathbf{u}_j\}$ from imaged translated point correspondences, where $i \neq j$ and $i, j \in 1 \ldots 3$. We construct three solver variants: (i) the hybrid solver **4PC+2CA**, that uses a pair of circular arcs and two pairs of imaged translated point correspondences, (ii) the hybrid solver **2PC+4CA** admitting two pairs of circular arcs and a pair of imaged translated point correspon-
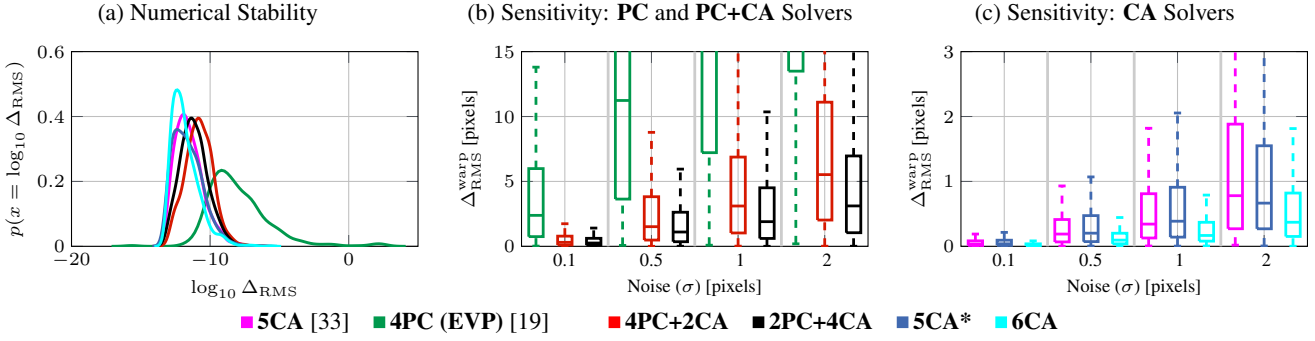
(a) Numerical Stability (b) Sensitivity: **PC** and **PC+CA** Solvers (c) Sensitivity: **CA** Solvers

■ **5CA** [33] ■ **4PC (EVP)** [19] ■ **4PC+2CA** ■ **2PC+4CA** ■ **5CA*** ■ **6CA**

Figure 5: *Numerical Stability and Noise Sensitivity of Solvers.* (a) Histogram of the $\log_{10}$ warp error for 1000 synthetic scenes with noiseless features. (b-c) RMS warp error Δ^{RMS} after 25 iterations of a simple RANSAC on 1000 synthetic scenes with increasing levels of noise σ added to the point correspondences and/or circular arcs. Results are shown for (b) the region based solver **4PC (EVP)** [19, 21] and proposed hybrid solvers **4PC+2CA** and **2PC+4CA**; (c) the arc-based solvers **5CA** [33], and proposed **5CA*** and **6CA**.

dences, (iii) and the solver **6CA** that uses three pairs of circular arcs. It is possible to construct a solver from only point correspondences. If coplanar vanishing points are assumed, then the solver is the same as the **6PC (EVL)** solver proposed in [21]. However, covariant region detections typically used to extract point correspondences are unlikely to provide orthogonal structures [21]. The case where two constructed vanishing points are coincident (see (9)) is handled by the fourth proposed **5CA*** solver. Three parallel scene lines are sufficient to provide constraints on two vanishing points that coincide. Thus the **5CA*** requires a triple and a pair of circular arcs. Inputs of the solvers are listed in Table 1 and input configurations are shown in Fig. 2.

Fig. 4 shows the qualitative performance of the solvers on challenging images from the Internet. See also Sec. D in Supplemental for more results.

4.5. Line Construction

The solvers use point correspondences and sets of circular arcs, where a set of arcs tentatively consists of the images of parallel scene lines under the division model. Point correspondences consist of points extracted from the image of translational symmetries. The preimages of the point correspondences must have the same translation direction and distance of translation in the scene plane. To reduce the expected number of RANSAC samples, we extract point correspondences from covariant region correspondences.

Imaged Translated Coplanar Repeats The corresponding points of two coplanar translated repeated regions [25] form parallel scene lines. Thus we can use the point correspondences extracted from distorted images of coplanar repeated texture [19, 22] to construct the undistorted images of parallel lines. Let $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ be two distorted points in correspondence. Then a line is constructed as a join of

their undistorted points,

$$\mathbf{t}(\lambda) = g(\tilde{\mathbf{x}}, \lambda) \times g(\tilde{\mathbf{x}}', \lambda). \tag{10}$$

Using (10), parallel lines $\mathbf{t}_i(\lambda)$ and $\mathbf{t}'_i(\lambda)$ can be constructed from points extracted from imaged translational symmetries to provide the constraints required by (4) to recover λ and $\mathbf{l}$. The coordinates of vanishing point $\mathbf{u}_i(\lambda) = \mathbf{t}_i(\lambda) \times \mathbf{t}'_i(\lambda)$ are either linear or quadratic in λ.

Imaged Scene Lines Lines are distorted to circles under the division model [4, 8, 26, 31], and the normals of a circle are mapped to the normals of the circle's undistorted image by the transposed inverse of the Jacobian of the division model. This gives the following form of the undistorted line $\mathbf{s}(\lambda)$

$$\mathbf{s}(\lambda) = \lambda \begin{pmatrix} \tilde{n}_x\tilde{x}^2 + 2\tilde{n}_y\tilde{x}\tilde{y} - \tilde{n}_x\tilde{y}^2 \\ \tilde{n}_y\tilde{y}^2 + 2\tilde{n}_x\tilde{x}\tilde{y} - \tilde{n}_y\tilde{x}^2 \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{n}_x \\ \tilde{n}_y \\ -\tilde{\mathbf{n}}^\top \tilde{\mathbf{x}} \end{pmatrix}, \tag{11}$$

where $\tilde{\mathbf{n}} = \left(\tilde{n}_x, \tilde{n}_y\right)^\top$ is a normal to a circle at the point $\tilde{\mathbf{x}} = \left(\tilde{x}, \tilde{y}\right)^\top$ [33]. Thus circles that are distorted images of parallel scene lines can generate pairs of undistorted lines $\mathbf{s}_i(\lambda)$ and $\mathbf{s}'_i(\lambda)$ that provide the constraints required by (4) to recover λ and $\mathbf{l}$. The coordinates of vanishing point $\mathbf{u}_i(\lambda) = \mathbf{s}_i(\lambda) \times \mathbf{s}'_i(\lambda)$ are either linear or quadratic in λ.

5. Features

Contours are constructed by linking sub-pixel Canny edge detections with morphological operations. The contours are decimated by the Ramer—Douglas—Peucker algorithm [6]. The maximum likelihood fit of a circle to the contour is estimated by nonlinear least squares, which is initialized by Taubin's bias-renormalization fit for conics [29].

Point correspondences are extracted from a covariant region correspondence [17, 22]. In particular, we use the

Solver	4PC (EVP)	5CA	4PC+2CA	2PC+4CA	5CA*	6CA	Hybrid	Arcs	All	6CA & 2PC+4CA
% of Top-1	1.5%	10.2%	15.5%	21.7%	25.4%	25.7%	-	-	-	-
Median λ Rel. Err.	14.91	2.27	2.81	2.38	2.24	2.29	2.33	2.27	**2.14**	2.16
Median f Rel. Err.	25.93	1.42	1.82	1.48	1.42	1.4	1.48	1.39	**1.38**	1.39
Median Δ^{RMS}	186.51	14.02	15.86	14.9	14.12	13.73	14.77	14.29	13.91	**13.35**

Table 2: *Performance on AIT Dataset.* The Top-1 solution is the calibration that has the lowest RMS warp error among the solvers. The proposed solvers are in grey. Solver combinations used in Hybrid RANSAC are in dark grey. "Hybrid" uses only the hybrid solvers *i.e.* **4PC+2CA** and **2PC+4CA**, "Arcs" uses only the arc solvers *i.e.* **5CA**, **5CA***, and **6CA**, and "All" uses every solver. The AIT dataset was run ten times for a total of 1020 calibrations for each solver or solver combination.

Maximally-Stable Extremal Region detector with the local affine-frame upgrade [30, 11, 12, 16], which provides three point correspondences. Affine-covariant regions are tentatively labeled as coplanar repeated texture if the regions are similar in appearance. Region appearance is embedded by the RootSIFT descriptor and clustered into tentative repeats similar to what was done in [21].

5.1. Rejecting Minimal Input Configurations

The configuration of the input sample is apriori unknown. The possible configurations are illustrated in Fig. 2. Implausible solutions from the invoked solvers that cannot handle the sampled input configuration are rejected by testing their geometric consistency with the minimal samples. Note that the minimal solution will exactly satisfy the algebraic constraints of the solver with the minimal sample as input. However, we test against additional properties of the minimal sample that are unused by the solver. Verification against the minimal sample incurs a negligible and justifiable computation cost since it can prevent unneeded consensus set evaluation by RANSAC, which is an expensive computation [10]. The rejection test is outlined for circular arcs and coplanar repeats in the next two paragraphs.

Consistency of Circular Arcs The minimal solution and the midpoint where the contour normal is measured are used to construct a circle through the distorted image of the vanishing point of the direction of the scene line that generates the contour. This construction is similar to a line through the vanishing point construction in the undistorted space proposed by Tardif in [28]. The contour midpoint is undistorted to $\bar{\mathbf{x}}$, and its join with the vanishing point $\mathbf{m} = \bar{\mathbf{x}} \times \mathbf{u}$ is distorted to the circle $\tilde{\mathbf{m}} = (a, b, c)$, where (a, b) is the circle center and c is the radius.

The consistency measure is the mean squared orthogonal distance between points of the contour $\{\tilde{\mathbf{x}}_k\}_{k=1}^{K}$ and the circle $\tilde{\mathbf{m}}$ through its midpoint

$$\mathcal{J} = \frac{1}{K} \sum_{k=1}^{K} (\sqrt{(\tilde{x}_k - a)^2 + (\tilde{y}_k - b)^2} - c)^2. \quad (12)$$

The construction is shown in Fig. A.1 of the Supplemental.

Consistency of Coplanar Repeats An affine-covariant region correspondence that is extracted from a translational symmetry can provide point correspondences that are translated in four directions in the scene [21]. See also Fig. A.2 of the Supplemental. The undistorted correspondences in each direction are used to estimate a vanishing point incident to the vanishing line recovered by the minimal solver. This is computed by solving the constrained least squares problem

$$\min_{\mathbf{u}} \|\mathbf{M}\mathbf{u}\|_2^2 \quad \text{subject to } \mathbf{C}\mathbf{u} = \mathbf{d}, \quad (13)$$

where $\mathbf{C} = \begin{bmatrix} \mathbf{l}^\top \\ 0 \quad 0 \quad 1 \end{bmatrix}, \mathbf{d} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mathbf{M} = \begin{bmatrix} \mathbf{t}_1 \cdots \mathbf{t}_k \end{bmatrix}^\top$, and $\mathbf{t}_1,...,\mathbf{t}_k$ are constructed by (10). The point correspondences are used to construct circles in the same way as contours, and the consistencies are measured by (12). Point correspondences used to compute the minimal solution will have zero error, but the minimal solution can be cross-validated by points along the unused translation directions provided by the region correspondence.

6. Experiments

The proposed solvers are quantitatively evaluated against the state-of-the-art solvers listed in Table 1 on synthetic and real data. Synthetic Manhattan scenes are used to assess the stabilities and noise sensitivities of the solvers. The standard AIT dataset of barrel-distorted images introduced by Wildenauer *et al.* in [33] is used to assess the solver accuracy for the auto-calibration task on challenging real images.

6.1. Warp Error

The warp error introduced in [23] is adapted to jointly assess the accuracy of the estimated calibrations. Ground truth absolute orientation of the camera with respect to the Manhattan frame is usually unknown for real images, thus the performance is assessed based on the intrinsics. The image is tessellated by an $N \times N$ grid of points $\{\tilde{\mathbf{x}}_i\}_{i=1}^{N^2}$. The tessellation ensures that the error is uniformly sampled. The image points $\tilde{\mathbf{x}}_i$ are back-projected to rays using ground-truth. The rays are projected by the estimated intrinsics and
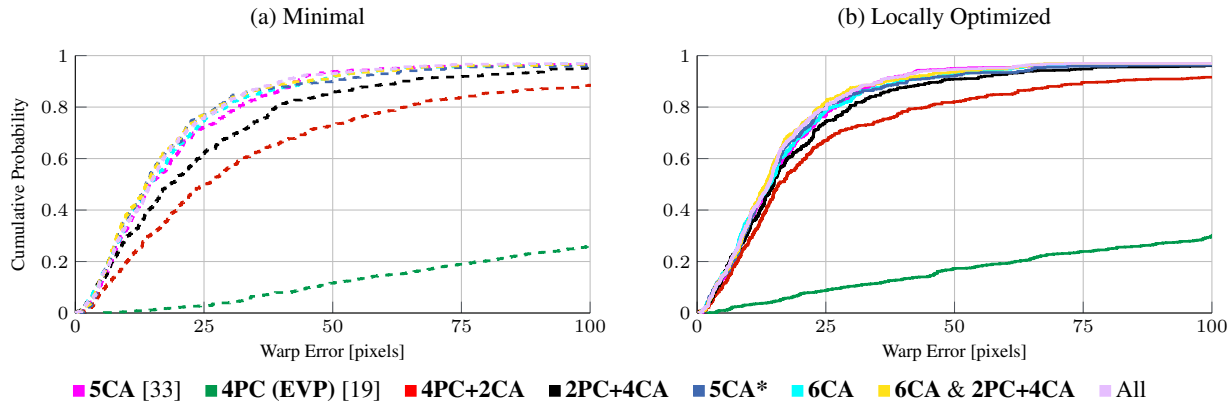
Figure 6: Cumulative distributions of the warp error on AIT dataset [33]. Results are shown for (a) the minimal *i.e.* initial solutions and (b) locally optimized *i.e.* final solutions. See also the distributions of the relative error of the division model parameter λ and the focal length f in Fig. C.1 of Supplemental.

the reprojection error between the projected rays and tessellation is used to define the warp error

$$\Delta_i = d(\tilde{\mathbf{x}}_i, g^d(\hat{\mathsf{K}}\mathsf{K}^{-1}g(\tilde{\mathbf{x}}_i, \lambda), \hat{\lambda})), \quad (14)$$

where $d(\cdot, \cdot)$ is the Euclidean distance and g^d is the distortion transformation. The root mean square warp error for $\{\tilde{\mathbf{x}}_i\}_{i=1}^{N^2}$ is reported and denoted as Δ^{RMS}. The warp error provides a geometric measure of calibration accuracy; however, an error in focal length can be compensated by an error in undistortion and vice-versa. Example warp errors are illustrated in Sec. B of the Supplemental.

6.2. Numerical Stability

The numerical stability measures the RMS warp error Δ^{RMS} of the solvers on noiseless features. Configurations of coplanar mutually orthogonal translated regions and parallel lines that are consistent with each solver's required inputs are generated for realistic scenes and camera configurations. Fig. 5a reports the distribution of $\log_{10} \Delta^{\mathrm{RMS}}$ on 1000 synthetic scenes. All of the proposed solvers demonstrate good numerical stability, which is consistent with the simple structure of the solvers. The arc-based solver **5CA** of [33] has similar structure to the proposed solvers and nearly as stable. The **4PC (EVP)** solver of [21] fails frequently. It is generated with the Gröbner bases method, which solves a complicated system of polynomial equations.

6.3. Noise Sensitivity

The proposed and state-of-the-art solvers are evaluated for their robustness to sensor noise (see Figs. 5c and 5b). White noise generated from $\mathcal{N}(\mathbf{0}_2, \sigma^2 \mathtt{I}_2)$ is added to the imaged translated symmetries and parallel lines. Solver sensitivity is measured at noise levels of $\sigma \in \{0.1, 0.5, 1, 2\}$. The solvers are used in a basic RANSAC estimator that minimizes the RMS warp error Δ^{RMS} over

25 minimal samples for 1000 scenes at each noise level. As expected, the solvers admitting arcs give superior performance since the entire contour is used to regress the arc. The proposed **6CA** is the most robust, while the proposed **5CA*** and state-of-the-art **5CA** solver [33] are both competitive. The proposed **4PC+2CA** is significantly more robust than the state-of-the-art **4PC (EVP)** solver [19], and the proposed **2PC+4CA** is competitive with the arc-based solvers.

6.4. Real Data

The solvers are evaluated on the AIT dataset introduced in [33], which consists of 102 barrel-distorted images taken by Canon EOS 500D camera with a Walimex Pro 8mm $170°$ HFOV fisheye lens. An offline calibration is provided with the dataset and used as the ground truth. The image center, distortion center, and principal point are assumed to be the same, and lens undistortion is modeled with the division model. The calibrated focal length is 1126.3 pixels and division model parameter is -2.4951×10^{-7} pixels for the AIT images at a resolution of 3000×2000.

The number of iterations was set to 1000 for the region based **4PC (EVP)** solver, and 500 for the other solvers. The reprojection threshold is 5.05 pixels for points of covariant regions and 1.26 pixels for contour points. The consistency of the features with an auto-calibration used for evaluation is similar to (12), and the models with maximal consensus sets are locally optimized by a method similar to [18].

Auto-calibrations on the AIT dataset are accumulated over ten runs. Table 2 reports the percentage of Top-1 solutions achieved by each solver, where a Top-1 solution is the auto-calibration that has the minimal RMS warp error for an image. We also report the median relative focal length and lens undistortion errors and the median RMS warp error. The best performing arc-based solver is the proposed **6CA** with twice as many Top-1 solutions and a 2% reduc-
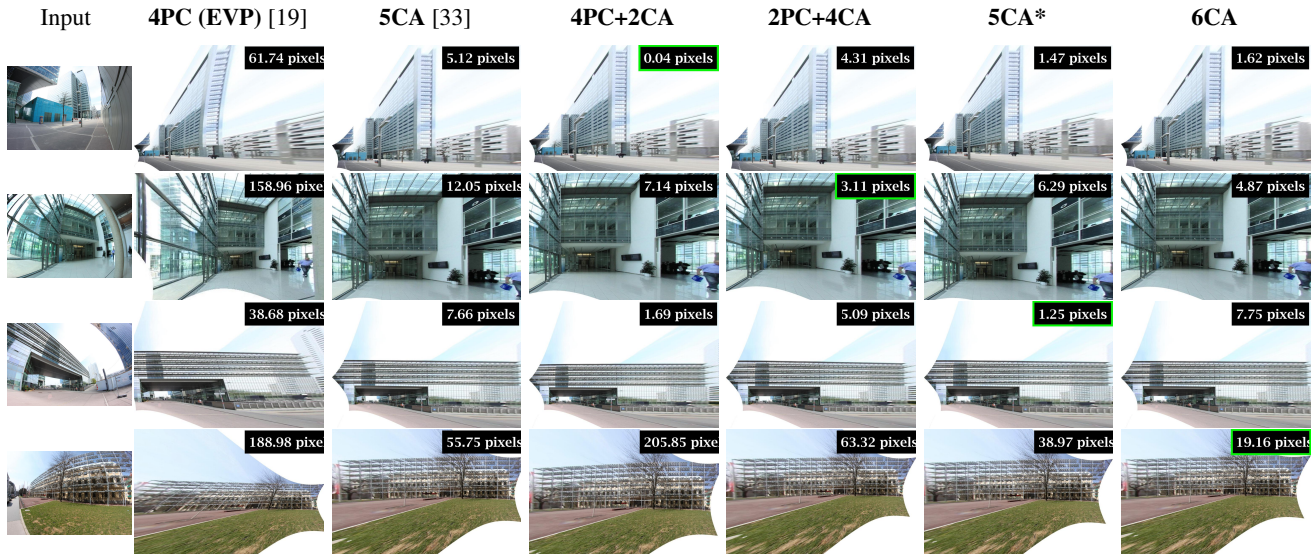
Figure 7: Example scenes for which either the arc-based or hybrid solvers have an advantage. The input image is on the left, and the dominant plane metrically-rectified for each solver is on the right. RMS warp errors are reported in the top right corners. Best results are in green.

tion in the warp error compared to the state-of-the-art **5CA** solver. The **2PC+4CA** hybrid solver is the best performing for solvers using points, and it significantly outperforms the state-of-the-art **4PC (EVP)**. As shown by Table 2, subsets of the AIT dataset are best solved by particular solvers, which suggests that using combinations of solvers is necessary to recover the best calibration across the dataset. Hybrid RANSAC was used to sample arcs and points according to the input types listed in Table 1 [5]. The solver admitting the sampled input is invoked to hypothesize an auto-calibration. We used the following solvers together: (i) all hybrid solvers, (ii) all arc-based solvers (iii) all solvers, and (iv) only the **6CA** and **2PC+4CA** solvers. The configuration using all solvers gave the most accurate estimates of focal length and division model parameter, while the combination of **6CA** and **2PC+4CA** gave the lowest warp error.

Fig. 6 reports the cumulative distributions of errors reported in Table 2 for the individual solvers and for the combination of all solvers, as well as using the **6CA** and **2PC+4CA** in combination. The proposed **5CA***, **6CA**, **2PC+4CA** and the state-of-the-art **5CA** of [33] give comparably good performance. The region-based **4PC (EVP)** solver of [19] performs poorly, due to noisy covariant region detections. The solver combinations are better than all individual results. Notably, the performance of the **6CA** and **2PC+4CA** combination matches the use of all solvers and exceeds the performance of the individual solvers. This suggests that the dataset is saturated if only arc-based solvers are used, and the inclusion of a point-based solver is necessary to improve calibration accuracy. Inclusion of more solvers gives only a small increase in accuracy across the dataset, which may show up in Top-1 solutions, but is not

impactful enough to justify adding extra solvers. Choosing the two best individual arc and hybrid solvers—**6CA** and **2PC+4CA**—works well.

Fig. 7 shows diverse scene content for which either the arc-based or hybrid solvers have an advantage. Note that the scenes where the **4PC+2CA** and **2PC+4CA** hybrid solvers perform best have high-contrast lines in only one direction, which suggests that translational symmetries are needed to constrain the second Manhattan frame direction.

7. Conclusions

We propose rectifying and calibrating minimal solvers that use combinations of circular arcs and point correspondences, which are images of parallel lines and translational symmetries in the scene, respectively. The proposed solvers extend accurate rectification and auto-calibration to distorted images of scenes that lack either coplanar texture or parallel scene lines. No individual solver emerged as a clear winner for the task of auto-calibration on the standard AIT dataset. Instead, we found that the solvers are complementary. Each solver works best on scenes with content that reflect its particular input configuration. There is a benefit to including the constraints from imaged translational symmetries even though the point correspondences extracted from covariant regions are more noisy than the circular arcs. However, experiments suggest that a significant improvement could be achieved by refining translational symmetries detected from covariant region correspondences. We achieve state-of-the-art performance on the AIT dataset by using multiple solvers in a RANSAC variant that samples different combinations of point correspondences and arcs as input for each of the solvers.

References

[1] Michel Antunes, Joao P. Barreto, Djamila Aouada, and Bjorn Ottersten. Unsupervised vanishing point detection and camera calibration from a single manhattan image with radial distortion. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6691–6699, 2017.

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

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *In ECCV*, pages 404–417, 2006.

[4] Faisal Bukhari and Matthew N Dailey. Automatic radial distortion estimation from a single image. *Journal of mathematical imaging and vision*, 45(1):31–45, 2013.

[5] Federico Camposeco, Andrea Cohen, Marc Pollefeys, and Torsten Sattler. Hybrid camera pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 136–144, 2018.

[6] David H Douglas and Thomas K Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2):112–122, 1973.

[7] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[8] Andrew W Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

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

[10] Jiří Matas and Ondrej Chum. Randomized ransac with sequential probability ratio test. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1727–1732 Vol. 2, 2005.

[11] Jiří Matas, Ondřej Chum, Martin Urban, and Tomás Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, 2002.

[12] Jiří Matas, Štěpán Obdržálek, and Ondřej Chum. Local affine frames for wide-baseline stereo. In *ICPR*, 2002.

[13] Andre Mateus, Srikumar Ramalingam, and Pedro Miraldo. Minimal solvers for 3d scan alignment with pairs of intersecting lines. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7234–7244, 2020.

[14] Pedro Miraldo, Tiago Dias, and Srikumar Ramalingam. A minimal closed-form solution for multi-perspective pose estimation using points and lines. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 490–507, Cham, 2018. Springer International Publishing.

[15] Anastasiia Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. In *Advances in Neural Information Processing Systems 30*, pages 4826–4837. Curran Associates, Inc., 2017.

[16] Štěpán Obdržálek and Jiří Matas. Object recognition using local affine frames on distinguished regions. In *BMVC*, 2002.

[17] Michal Perdoch, Juan E. Sala Matas, and Ondřej Chum. Epipolar geometry from two correspondences. *18th International Conference on Pattern Recognition (ICPR'06)*, 4:215–219, 2006.

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

[19] James Pritts, Zuzana Kukelova, Viktor Larsson, and Ondřej Chum. Radially-distorted conjugate translations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1993–2001, 2018.

[20] James Pritts, Zuzana Kukelova, Viktor Larsson, and Ondřej Chum. Rectification from radially-distorted scales. In *Asian Conference on Computer Vision*, pages 36–52. Springer, 2018.

[21] James Pritts, Zuzana Kukelova, Viktor Larsson, Yaroslava Lochman, and Ondřej Chum. Minimal solvers for rectifying from radially-distorted conjugate translations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[22] James Pritts, Zuzana Kukelova, Viktor Larsson, Yaroslava Lochman, and Ondrej Chum. Minimal solvers for rectifying from radially-distorted scales and change of scales. *Int. J. Comput. Vis.*, 128(4):950–968, 2020.

[23] James Pritts, Denys Rozumnyi, M Pawan Kumar, and Ondrej Chum. Coplanar repeats by energy minimization. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 107.1–107.12. BMVA Press, September 2016.

[24] S. Ramalingam, S. Bouaziz, and P. Sturm. Pose estimation using both points and lines for geo-localization. In *2011 IEEE International Conference on Robotics and Automation*, pages 4716–4723, 2011.

[25] Frederik Schaffalitzky and Andrew Zisserman. Geometric grouping of repeated elements within images. In *BMVC*, 1998.

[26] Rickard Strand and Eric Hayman. Correcting radial distortion by circle fitting. In *Proceedings of the British Machine Vision Conference*, pages 9.1–9.10. BMVA Press, 2005.

[27] Y. Taguchi, Y. Jian, S. Ramalingam, and C. Feng. Point-plane slam for hand-held 3d sensors. In *2013 IEEE International Conference on Robotics and Automation*, pages 5182–5189, 2013.

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

[29] Gabriel Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(11):1115–1138, 1991.

[30] Andrea Vedaldi and Brian Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.

[31] Aiqi Wang, Tianshuang Qiu, and Longtan Shao. A simple method of radial distortion correction with centre of distortion estimation. *Journal of Mathematical Imaging and Vision*, 35(3):165–172, 2009.

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

[33] Horst Wildenauer and Branislav Micusik. Closed form solution for radial distortion estimation from a single vanishing point. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 106.1–106.11. BMVA Press, 2013.