Algebraic Constraint for Preserving Convexity of Planar Homography

Gaku Nakano

Central Research Labs, NEC Corporation

1753 Shimonumabe, Kawasaki, Japan, 211-8666

g-nakano@nec.com

Abstract

This paper proposes a new algebraic constraint for the planar homography estimation to ensure transformations between two convex quadrilaterals. The new constraint is derived by utilizing a projective invariance of an ellipse, i.e. *an ellipse is projected as an ellipse in other views under a physically plausible homography. The invariance is expressed by a quadratic inequality about a homography matrix, therefore, the quadratic constraint can be incorporated with a direct linear method that can be solved as a generalized eigenvalue problem. We demonstrate by experiments that both LO-RANSAC and M-estimator with the proposed constraint are more accurate and robust to outliers than LO-RANSAC with the standard 4-point DLT method.*

1. Introduction

The planar homography is a projective transformation of a 3D plane onto images. Determining a homography between two views is the basic step in various applications of computer vision: camera calibration [38], image stitching [32], visual odometry [4, 39], and visual-SLAM [31, 34]. Therefore, the planar homography has been investigated in decades for its properties [22] and improvement of estimation on both accuracy and robustness.

The planar homography is expressed as a 3×3 full rank matrix. The traditional method for estimating a homography matrix is the direct linear transform (DLT) method [15], which uses at least four point correspondences obtained by feature point detection and matching on two images. For obtaining point correspondences, scale or affine invariant features are widely used, such as SIFT [21], ORB [30], Hessian-Affine [24] *etc*. By utilizing scale and orientation information from feature points, a homography matrix can be estimated using at least two points [2]. Moreover, using five points, radial lens distortion is jointly obtainable [13, 19]. Not only points, lines [9] and conics [18] are also available for the homography matrix estimation.

Since the DLT method gives a biased solution, higher accurate direct methods [16, 29] have been proposed by analyzing algebraic error to remove the statistical bias. These methods are based on the minimization of independent and isotropic Gaussian noise, therefore, they assume that the point correspondences are not contaminated by outliers.

To deal with outliers for robust estimation, the standard approach is to use a variant of random sample consensus (RANSAC) [5, 6, 11], which iteratively performs hypothesis generation and validation by sampling four points from contaminated point correspondences. Optimization-based methods, which are not the mainstream though, have been also proposed as alternative to RANSAC: ICP-based [35] and M-estimator-based [26]. Recently, learning-based approaches [8, 27], which directly estimate a homography matrix from images, have been expected to achieve both efficiency and robustness.

While many methods for handling outliers have been extensively discussed, it has received less attention on theories or methods for suppressing implausible homographies that lead to an unrealistic transformation. Hartley and Zisserman [15] described the projective orientation of a transformed point, called the cheiral inequalities. Márquez-Neila *et al.* [23] introduced a geometric constraint that four point correspondences must satisfy if they are inliers to accelerate the sampling process in RANSAC. Kanazawa and Kanatani [17] proposed a multi-step RANSAC which starts by fitting a rigid transformation with a large tolerance, then applies affine and homography transformations with tightening the tolerance to obtain a valid homography.

This paper introduces a new algebraic constraint of the homography matrix, which contributes to improve accuracy and robustness in the presence of outliers. As shown in Fig. 1, the traditional approach fails to estimate a geometrically plausible homography transformation if point correspondences are contaminated by slight outliers that survived from RANSAC. Using the proposed constraint, we can prevent such homographies that transform a convex quadrilateral to a non-convex quadrilateral. The new constraint is formulated as a simple quadratic inequality, therefore, it can be incorporated to a normalizing factor of the DLT method.

(a) Point pairs and ground-truth homograpy

(b) DLT

(c) Proposed

Figure 1: Homography estimation from matched point pairs including outliers. (a) 20 correct matches in green and 2 outliers in purple. Cyan quadrilaterals are corresponding areas from the left image to the right transformed by the homography matrix using the inliers in green. (b) The rectangle turns to a crossed quadrilateral by the traditional DLT method using all 22 points. (c) Convexity is preserved by the proposed method with the new constraint.

2. Planar Homography

The planar homography $\mathbf{H}$ is a 3×3 matrix that relates a projective transformation of a plane between two views. Given a pair of corresponding 2D points $\mathbf{x} \leftrightarrow \mathbf{x}'$ in the first and the second images, the homography transformation can be written by

$$\mathbf{x}' = \frac{1}{\mathbf{h}_3^{\mathsf{T}} \mathbf{x}} \mathbf{H} \mathbf{x} \quad \longrightarrow \quad \mathbf{x}' \propto \mathbf{H} \mathbf{x}, \qquad (1)$$

where $\mathbf{h}_i^{\mathsf{T}}$ represents the i-th row of $\mathbf{H}$ and $\propto$ denotes equality up to scale. Note that 2D points are represented by their homogeneous coordinates, *i.e.* $\mathbf{x} = [x, y, 1]^{\mathsf{T}}$ and $\mathbf{x}' = [x', y', 1]^{\mathsf{T}}$.

A homography matrix $\mathbf{H}$ can be expressed as a chain of three transformations: a similarity $\mathbf{H}_s$, an affinity $\mathbf{H}_a$, and a projectivity $\mathbf{H}_p$. The decomposition can be given in the form

$$\mathbf{H} = \mathbf{H}_s \mathbf{H}_a \mathbf{H}_p = \begin{bmatrix} s\mathbf{R}\mathbf{U} + \mathbf{t}\mathbf{v}^{\mathsf{T}} & \mathbf{t} \\ \mathbf{v}^{\mathsf{T}} & \nu \end{bmatrix}, \qquad (2)$$

where s is a scaling factor, $\mathbf{R}$ is a 2×2 rotation matrix, $\mathbf{t}$ is a 2×1 translation vector, and $\mathbf{U}$ is an upper triangular matrix with $\det(\mathbf{U}) = 1$. For a valid and unique decomposition, $s > 0$ and $\nu \neq 0$.

Although $\mathbf{H}$ consists of 9 elements, the homography actually has 8 degrees of freedom due to the scale ambiguity. Equation (1) gives two constraints for a pair of corresponding points $\{\mathbf{x}, \mathbf{x}'\}$, therefore, we can find $\mathbf{H}$ by at least four points.

Given $n \geq 4$ points, the homography estimation can be formulated by

$$\min_{\mathbf{H}} \quad \sum_{i=1}^{n} \|\mathbf{x}_i' \times \mathbf{H}\mathbf{x}_i\|^2 = \|\mathbf{A}\mathbf{h}\|^2 \qquad (3)$$
$$\text{s.t.} \qquad \|\mathbf{h}\|^2 = 1$$

where $\mathbf{h}^{\mathsf{T}} = [\mathbf{h}_1^{\mathsf{T}}, \mathbf{h}_2^{\mathsf{T}}, \mathbf{h}_3^{\mathsf{T}}]$ and $\mathbf{A}$ is an $n \times 9$ design matrix. The optimal solution of Eq. (3) can be obtained by the DLT method [15], which formulates Eq. (3) as $\mathbf{B}\mathbf{h} = \lambda\mathbf{h}$, where $\mathbf{B} = \mathbf{A}^{\mathsf{T}}\mathbf{A}$ and a Lagrange multiplier λ. The DLT method determines $\mathbf{h}$ as the eigenvector corresponding to the smallest eigenvalue of $\mathbf{B}$.

The DLT method minimizes algebraic distance that is not geometrically meaningful. Thus, a solution obtained by the DLT method can be further optimized by performing an iterative non-linear optimization

$$\min_{\mathbf{H}} \quad \sum_{i=1}^{n} d(\mathbf{x}_i', \mathbf{H}\mathbf{x}_i)^2 \qquad (4)$$
$$\text{s.t.} \qquad \|\mathbf{h}\|^2 = 1$$

where $d(\cdot, \cdot)$ denotes the Euclidean distance between two points in the image coordinates.

We can obtain the statistically optimal solution $\mathbf{H}$ by using the DLT and the non-linear methods as long as the point correspondences are not contaminated by outliers. To deal with outliers, RANSAC [5, 6, 11] and M-estimators [26, 36] have been used in the literature.

3. Previous Work

The standard methods mentioned in Sec. 2 do not consider that a homography matrix expresses a physically meaningful transformation. Therefore, a homography matrix estimated by these methods may lead to an undesirable transformation that cannot occur in the real world, *e.g.* from a rectangle to a butterfly quadrilateral, as shown in Fig. 1b. In this section, we will briefly review mathematical conditions investigated by the previous studies for preserving the convexity of a homography matrix.

3.1. Cheiral inequalities

It has been known in a textbook (Chapter.21.6–21.7 in [15]) that we can obtain the projective orientation of a point (x', y') by validating the sign of the determinant of its Jacobian. The Jacobian $\mathbf{J}$, the partial derivatives of (x', y')

at (x, y), can be written by

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{bmatrix}$$
$$= \frac{1}{\mathbf{h}_3^\mathsf{T} \mathbf{x}} \begin{bmatrix} H_{11} - H_{31}x' & H_{12} - H_{32}x' \\ H_{21} - H_{31}y' & H_{22} - H_{32}y' \end{bmatrix}, \quad (5)$$

where H_{ij} denotes the (i, j) element of $\mathbf{H}$. It is called orientation-preserving if the determinant of $\mathbf{J}$ is positive:

$$\begin{aligned} \det(\mathbf{J}) \propto & (H_{32}H_{21} - H_{31}H_{22})x' \\ & + (H_{31}H_{12} - H_{32}H_{11})y' \\ & + (H_{11}H_{22} - H_{12}H_{21}) > 0. \end{aligned} \quad (6)$$

Considering the inverse-mapping $\mathbf{H}^{-1}$ [25], Eq.(6) results in

$$(H_{31}x + H_{32}y + H_{33})/\det(\mathbf{H}) > 0. \quad (7)$$

We can validate the plausibility of a homography matrix if four corner points surrounding other points satisfy Eq. (7). Moreover, Eq. (7) can also be used to remove orientation-reversing points when a homography matrix is given.

The above constraint, also known as the cheiral inequalities, can be incorporated into a nonlinear optimization to enforce that a homography matrix preserves the convexity of the transformation:

$$\min_{\mathbf{H}} \quad \sum_{i=1}^{n} d(\mathbf{x}'_i, \mathbf{H}\mathbf{x}_i)^2 \quad (8)$$
$$\text{s.t.} \quad \det(\mathbf{H}) = 1, \ \mathbf{h}_3^\mathsf{T}\mathbf{x}_j \geq 0, \forall j \in \{1, \ldots, 4\}.$$

To the best of our knowledge, the constrained problem Eq. (8) has not been discussed well in the literature.

3.2. Signed area constraint

The cheiral inequalities cannot be applied to validate whether the input point correspondences can produce a physically meaningful homography because Eq. (7) requires a homography matrix. To deal with this issue, Márquez-Neila *et al.* [23] proposed another constraint based on the oriented projective geometry [20].

If given four point correspondences are all correct pairs, the area of a triangle formed by any three of the four points must have the same sign. This can be formulated by

$$\begin{aligned} \text{sign}(\mathbf{x}_i^\mathsf{T}(\mathbf{x}_j \times \mathbf{x}_k)) = \text{sign}(\mathbf{x}'^\mathsf{T}_i(\mathbf{x}'_j \times \mathbf{x}'_k)), \\ \forall i, j, k \in \{1, 2, 3, 4\}. \end{aligned} \quad (9)$$

The signed area constraint Eq. (9) can be easily calculated, therefore, it is suitable for discarding outliers in each RANSAC iteration before computing a homography matrix. Introducing this step, RANSAC for homography estimation can be accelerated by several times without any performance loss [28]. In contrast to the cheiral inequalities, the signed area constraint has a drawback that is difficult to use for nonlinear optimization because of a non-smooth function and the combinatorial explosion.
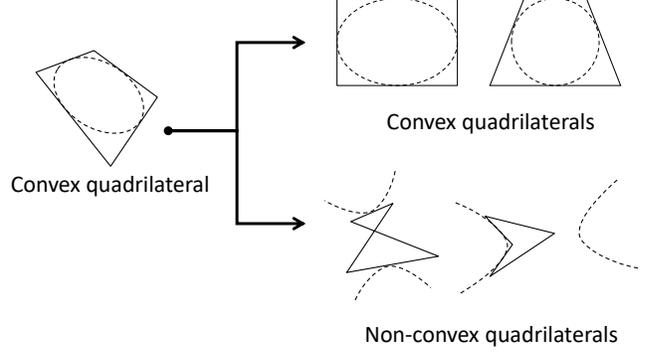

Figure 2: Concept of the proposed constraint. An ellipse inside a convex quadrilateral is still an ellipse under a physically meaningful homography transformation. If the transformation leads to a non-convex quadrilateral, the ellipse turns to be a hyperbola.

4. New Algebraic Constraint

In this section, we will address a projective invariance of a homography transformation between two convex quadrilaterals. The basic idea, illustrated in Fig. 2, is that an ellipse in the first image should be observed as an ellipse in the second image under a homography preserving the convexity of a quadrilateral. We aim to find a mathematical condition that guarantees *"an ellipse has to be an ellipse"* before and after a homography transformation.

4.1. Projective invariance of ellipse

An ellipse on a plane can be written by

$$ax^2 + bxy + cy^2 + dx + ey + f = \mathbf{x}^\mathsf{T}\mathbf{C}\mathbf{x} = 0, \quad (10)$$

where

$$\mathbf{C} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}, \quad (11)$$

$$\Delta = ac - \frac{b^2}{4} > 0. \quad (12)$$

Note that $\mathbf{x}$ in Eq. (10) represents a point on the ellipse $\mathbf{C}$, which is not a point correspondence for estimating a homography matrix $\mathbf{H}$. Equation (12) is the necessary and sufficient condition for a conic to be an ellipse if $\mathbf{C}$ is full ranked, *i.e.* $\det(\mathbf{C}) \neq 0$. The coefficients $a, \ldots, f$ can be expressed by using ellipse parameters:

$$\begin{aligned} a &= r_a^2 \sin^2 \theta + r_b^2 \cos^2 \theta, \\ b &= 2(r_b^2 - r_a^2) \sin \theta \cos \theta, \\ c &= r_a^2 \cos^2 \theta + r_b^2 \sin^2 \theta, \\ d &= -2ax_c - by_c, \\ e &= -bx_c - 2cy_c, \\ f &= ax_c^2 + bx_cy_c + cy_c^2 - r_a^2r_b^2, \end{aligned} \quad (13)$$

where semi-major axis r_a, semi-minor axis r_b, center coordinate (x_c, y_c), and rotation angle θ.

Under a homography transformation $\mathbf{x}' \propto \mathtt{H}\mathbf{x}$, the ellipse $\mathtt{C}$ transforms to a conic $\mathtt{C}'$ in the second image:

$$\mathbf{x}^\mathsf{T}\mathtt{C}\mathbf{x} \propto \mathbf{x}'^\mathsf{T}\mathtt{H}^{-\mathsf{T}}\mathtt{C}\mathtt{H}^{-1}\mathbf{x}' \qquad (14)$$
$$\propto \mathbf{x}'^\mathsf{T}\mathtt{C}'\mathbf{x}'.$$

Thus, we obtain

$$\mathtt{C}' \propto \mathtt{H}^{-\mathsf{T}}\mathtt{C}\mathtt{H}^{-1}. \qquad (15)$$

The transformed conic $\mathtt{C}'$ has to be an ellipse if the homography matrix $\mathtt{H}$ represents a transformation between two convex quadrilaterals. Let a', b', c' be the elements of $\mathtt{C}'$ similarly to Eq. (11). We can formulate an ellipse constraint Δ' of $\mathtt{C}'$ as follows:

$$
\begin{aligned}
\Delta' &= a'c' - \frac{b'^2}{4} \\
&= \frac{\begin{aligned}(4cf - e^2)H_{31}^2 + (2de - 4bf)H_{31}H_{32} \\ +(2be - 4cd)H_{31}H_{33} + (4af - d^2)H_{32}^2 \\ + (2bd - 4ae)H_{32}H_{33} + (4ac - b^2)H_{33}^2\end{aligned}}{\begin{aligned}4(H_{11}H_{22}H_{33} - H_{11}H_{23}H_{32} \\ - H_{12}H_{21}H_{33} + H_{12}H_{23}H_{31} \\ + H_{13}H_{21}H_{32} - H_{13}H_{22}H_{31})^2\end{aligned}} \\
&= \frac{\mathbf{h}^\mathsf{T}\mathtt{S}\mathbf{h}}{\det(\mathtt{H})^2},
\end{aligned}
\qquad (16)
$$

where

$$\mathtt{S} = \begin{bmatrix} \mathbf{0}_{6\times6} & \mathbf{0}_{6\times3} \\ \mathbf{0}_{3\times6} & \mathtt{S}_1 \end{bmatrix},$$

$$\mathtt{S}_1 = \frac{1}{4}\begin{bmatrix} 4cf - e^2 & de - 2bf & be - 2cd \\ de - 2bf & 4af - d^2 & bd - 2ae \\ be - 2cd & bd - 2ae & 4ac - b^2 \end{bmatrix}. \qquad (17)$$

The denominator of Eq. (16) is obviously positive for non-degenerate homography matrices, therefore, the conic $\mathtt{C}'$ can be an ellipse if and only if the numerator is greater than zero, *i.e.*

$$\mathbf{h}^\mathsf{T}\mathtt{S}\mathbf{h} > 0. \qquad (18)$$

The inequality Eq. (18) is the new constraint that ensures a homography transformation to be a mapping between two convex quadrilaterals. Comparing Eq. (18) with Eq. (2), we can see that the constraint relates to only the projective term $[\mathbf{v}^\mathsf{T}, \nu] = [H_{31}, H_{32}, H_{33}]$. Thus, the new constraint can be considered reasonable.

4.2. Ellipse to be fit

We derived the new constraint for homography transformations in the previous section. A question now arises: Arbitrary ellipses can be defined on the target plane, what is the appropriate one? To answer this question, we discuss ellipse fitting based on computational geometry in this section. Figure 3 illustrates three types of ellipse fitting we will describe.
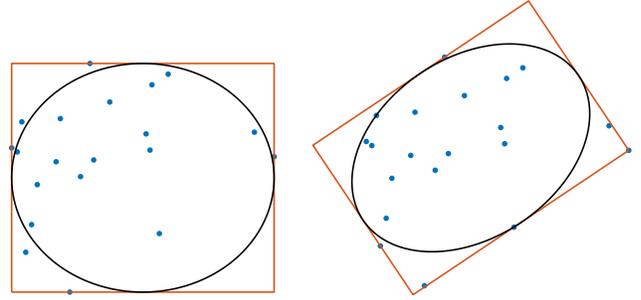
(a) Ellipse inscribed in a bounding box.

(b) Ellipse inscribed in the minimum rotated bounding rectangle.
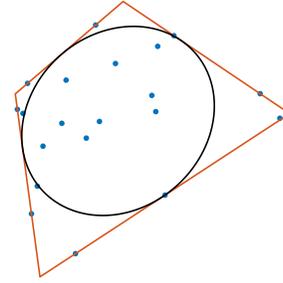
(c) Ellipse inscribed in the minimum bounding quadrilateral.

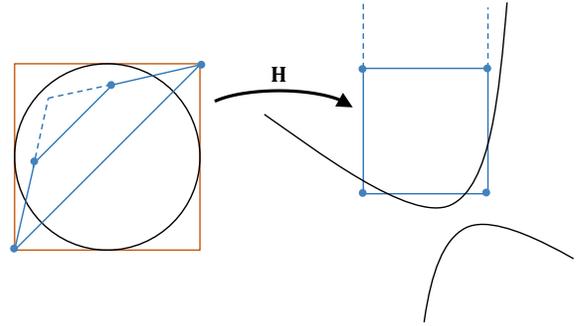Figure 3: Three types of an ellipse fitted on a 2D point cloud.

Figure 4: Degenerate case of the bounding box based approach, shown in Fig. 3a. The intersection of the two sides of the trapezoid is inside the ellipse in the first image. A homography mapping from the trapezoid to a square transfers the intersection to a point at infinity in the second image. The ellipse becomes a hyperbola, therefore, the convexity cannot be preserved using the proposed constraint.

4.2.1 Ellipse inscribed in the bounding box

One of the simplest ways is shown in Fig. 3a, where an ellipse is inscribed in a bounding box surrounding the 2D points. The ellipse parameters can be easily computed from the four corner points: r_a and r_b correspond to the half of the length of the sides, (x_c, y_c) is the center of the mass, and $\theta = 0$.

This idea seems to be intuitively correct; however, it is not always true. Let us consider an example shown in Fig. 4. Here, an inclined trapezoid in the first image gets mapped by H to a square in the second image. In this case, the intersections of the opposite side pairs of the trapezoid should map to a point at infinity after the transformation. If one of these intersections is inside the ellipse inscribed in a bounding box, the ellipse cannot be transformed to an ellipse in the second image. Therefore, this approach is too simple to be appropriate.

4.2.2 Ellipse inscribed in the rotated bounding rectangle

Figure 3b represents another approach to fit an ellipse, which is inscribed in the minimum rotated bounding rectangle. It is expected that the degenerate case such as Fig. 4 can be avoidable. Algorithms for finding the rotated rectangle have been well studied [33], and public implementations are available, *e.g.* the `minAreaRect` function in OpenCV. Parameters r_a, r_b, x_c, y_c can be calculated similarly to the bounding box approach. Rotation angle θ is generally non-zero, which is the angle between the semi-major axis and the horizontal axis.

4.2.3 Ellipse inscribed in the minimum bounding quadrilateral

More sophisticated approach than the previous ones is to fit the minimum bounding quadrilateral, which is illustrated in Fig. 3c. Generally, orthogonality of a quadrilateral is not preserved in this case. Although the minimum bounding quadrilateral is not often used in the computer vision community, some algorithms [7, 10] have been proposed for solving this problem in computational geometry.

There are infinite ellipses inscribed in the minimum bounding quadrilateral. We uniquely determine a single ellipse tangent to four edges of the quadrilateral as follows:

1. Assume that the center (x_c, y_c) is at the centroid of the four corners of the quadrilateral.
2. Shift the four points by applying a translation T so that their origin is at (x_c, y_c). Thus, $d = e = 0$.
3. Calculate four edge lines $\mathbf{l}_i$ of the quadrilateral.
4. Find a dual conic $\mathtt{C}^* = \begin{bmatrix} a^* & b^*/2 & 0 \\ b^*/2 & c^* & 0 \\ 0 & 0 & f^* \end{bmatrix}$ such that $\mathbf{l}_i^\mathsf{T} \mathtt{C}^* \mathbf{l}_i = 0, \forall i \in \{1, \dots, 4\}$.
5. Determine an ellipse by $\mathtt{C} = \mathtt{T}^\mathsf{T} \mathtt{C}^{*-1} \mathtt{T}$.

The minimum bounding quadrilateral is convex, therefore, a conic calculated at the step 5 is always an ellipse.

5. Proposed Solution

5.1. Point normalization

Using the ellipse fitting described in Sec. 4.2, we normalize 2D points so that the center of an ellipse is at the origin $(x_c, y_c) = (0, 0)$ and the semi-minor and semi-major axes are r_b/r_a and $r_a = 1$, respectively. According to Eq. (13), the shifting leads to $d = e = 0$. Thus, the constrained matrix $\mathtt{S}_1$ can be rewritten by

$$\mathtt{S}_1 = \begin{bmatrix} cf & -bf/2 & 0 \\ -bf/2 & af & 0 \\ 0 & 0 & ac - b^2/4 \end{bmatrix}. \quad (19)$$

Lemma 1. *Under the above normalization,* $\mathtt{S}_1$ *has a single positive eigenvalue* λ_+ *and two negative eigenvalues.*

Proof. One of the three eigenvalues is obviously $ac - b^2/4 > 0$, which coincides with the ellipse condition, Eq. (12). By inserting $r_a = 1$ into Eq. (13), the positive eigenvalue can be expressed by $\lambda_+ = ac - b^2/4 = r_b^2$. From the top-left 2×2 submatrix of $\mathtt{S}_1$, we can obtain the other two eigenvalues in a closed form:

$$\begin{aligned} f(a + c - \sqrt{(a-c)^2 + b^2})/2 &= -r_b^4, \\ f(a + c + \sqrt{(a-c)^2 + b^2})/2 &= -r_b^2. \quad \square \end{aligned} \quad (20)$$

In the following section, we will propose a linear solution for finding a homography matrix using $\mathtt{S}_1$ in Eq. (19).

5.2. Generalized eigenvalue solver

Although the proposed constraint Eq. (18) is an inequality, we can use a quadratic constraint $\mathbf{h}^\mathsf{T} \mathtt{S} \mathbf{h} = 1$ by incorporating a proper scaling to remove the scale ambiguity. This is a similar technique used in the ellipse-specific fitting problem [12, 14].

We can build a homography estimation problem with the new constraint as follows:

$$\begin{aligned} \min_{\mathtt{H}} \quad & \|\mathtt{A}\mathbf{h}\|^2 = \mathbf{h}^\mathsf{T} \mathtt{B} \mathbf{h} \\ \text{s.t.} \quad & \mathbf{h}^\mathsf{T} \mathtt{S} \mathbf{h} = 1 \end{aligned} \quad (21)$$

where $\mathtt{B} = \mathtt{A}^\mathsf{T} \mathtt{A}$. Introducing a Lagrange multiplier λ, we obtain the Karush-Kuhn-Tucker condition for the optimal homography as a system of linear equations

$$\begin{aligned} &\mathtt{B}\mathbf{h} = \lambda \mathtt{S}\mathbf{h} \\ \longrightarrow &\begin{bmatrix} \mathtt{B}_1 & \mathtt{B}_2 \\ \mathtt{B}_2^\mathsf{T} & \mathtt{B}_3 \end{bmatrix} \begin{bmatrix} \mathbf{h}_{1,2} \\ \mathbf{h}_3 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{0}_{6\times 6} & \mathbf{0}_{6\times 3} \\ \mathbf{0}_{3\times 6} & \mathtt{S}_1 \end{bmatrix} \begin{bmatrix} \mathbf{h}_{1,2} \\ \mathbf{h}_3 \end{bmatrix} \\ \longrightarrow &\begin{cases} \mathtt{B}_1 \mathbf{h}_{1,2} + \mathtt{B}_2 \mathbf{h}_3 = \mathbf{0} \\ \mathtt{B}_2^\mathsf{T} \mathbf{h}_{1,2} + \mathtt{B}_3 \mathbf{h}_3 = \lambda \mathtt{S}_1 \mathbf{h}_3 \end{cases} \end{aligned} \quad (22)$$

where $\mathtt{B}_1$, $\mathtt{B}_2$, and $\mathtt{B}_3$ denote submatrices of $\mathtt{B}$ which are of size 6×6, 6×3, and 3×3, respectively, and $\mathbf{h}_{1,2}^\mathsf{T} = [\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}]$.

Rewriting Eq. (22) with respect to $\mathbf{h}_{1,2}$ and $\mathbf{h}_3$, we obtain

$$\mathbf{h}_{1,2} = -\mathbf{B}_1^{-1}\mathbf{B}_2\mathbf{h}_3, \qquad (23)$$

$$\mathbf{M}\mathbf{h}_3 = \lambda\mathbf{h}_3, \qquad (24)$$

where $\mathbf{M} = \mathbf{S}_1^{-1}(\mathbf{B}_3 - \mathbf{B}_2^\mathsf{T}\mathbf{B}_1^{-1}\mathbf{B}_2)$. According to Eq. (24), we can obtain $\mathbf{h}_3$ as the eigenvector of $\mathbf{M}$ corresponding to its eigenvalue λ. The 3×3 matrix $\mathbf{M}$ has three eigenvalues, however, we can uniquely determine the optimal $\mathbf{h}_3$.

Generally, $\mathbf{B}$ is symmetric positive definite for noisy inputs. From Lemma 1, the sign of the eigenvalues of $\mathbf{S}$ are always $[+, -, -, 0, 0, 0, 0, 0, 0]$ and the generalized eigenvalue problem $\mathbf{B}\mathbf{h} = \lambda\mathbf{S}\mathbf{h}$ has a single positive eigenvalue that corresponds to λ_+ of $\mathbf{M}$. Thus, we can determine the optimal $\mathbf{h}_3$ as the eigenvector of $\mathbf{M}$ associated with the single positive eigenvalue. Substituting $\mathbf{h}_3$ into Eq. (23), we obtain $\mathbf{h}_{1,2}$. Finally, we can recover the homography matrix $\mathbf{H}$ in the image coordinates by performing de-normalization.

When we have $n = 4$ points, the solution of Eq. (21) coincides with that of the conventional DLT method, Eq. (3), because $\mathtt{rank}(\mathbf{B}) = 8$ and $\lambda_+ = 0$. In this case, the generalized eigenvalue problem $\mathbf{B}\mathbf{h} = \lambda\mathbf{S}\mathbf{h}$ results in $\mathbf{B}\mathbf{h} = \mathbf{0}$, where the solution is given by $\mathbf{h} = \mathtt{null}(\mathbf{B})$. From this, we cannot enforce the proposed constraint for $n = 4$ points. It makes sense because $n = 4$ points correspond to the four vertices of a quadrilateral, which is not always convex.

6. Experiment

This section reports experimental results on both synthetic and real data evaluations. We refer to methods compared in this section as:

DLT The standard 4-point DLT method [15].

NL Nonlinear optimization minimizing the single transfer error, Eq. (4), by the Levenberg-Marquardt method.

NLχ Nonlinear optimization with the cheiral inequalities, Eq. (8), by the interior-point method. Four corner points for the inequality constraint was given by the minimum bounding quadrilateral in Sec. 4.2.3.

Sign The signed area constraint, Eq. (9), for RANSAC.

ConvexDLT The proposed method, Eq. (21). Ellipse fitting is denoted by **+bbox/+minRect/+minQuad**.

A MATLAB code [7][1] was used for solving the minimum bounding quadrilateral problem required in **NLχ** and **ConvexDLT+minQuad**. We wrote all methods in MATLAB and ran the experiments on a PC with Core i9-7920X.

To evaluate the accuracy of estimated homography matrices by a geometrically meaningful metric, we measured the normalized symmetric pixel transfer error (NSPT), which is inspired by [3, 37]. The NSPT error is calculated as follows:

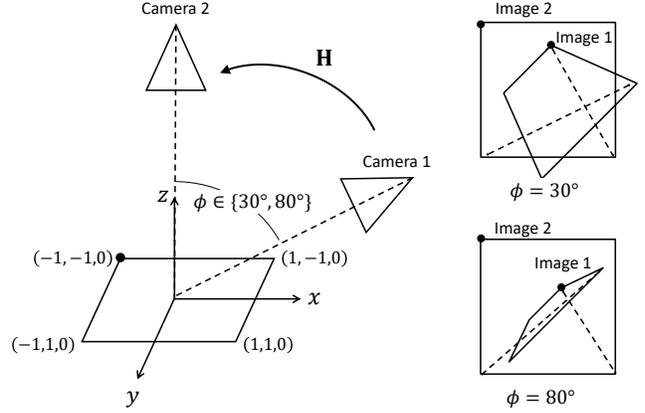[1]https://www.mathworks.com/matlabcentral/fileexchange/34767-a-suite-of-minimal-bounding-objects

Figure 5: Synthetic data generation. *Left*: Experimental configuration. *Right*: Examples of synthesized images. Two diagonal corners connected by dashed lines are actually outliers but regarded as correct matches.

1. Sample pixels uniformly in the first image, and project them onto the second image by $\mathtt{H}_\mathrm{gt}$ and $\mathtt{H}_\mathrm{est}$.
2. Find projected pixels in the overwrapped visible area in the second image.
3. Calculate the Euclidean distance of the projected pixels between the first and the second images.
4. Average the distance by the number of the visible pixels.
5. Normalize the mean by $1/\sqrt{w_\mathrm{img}^2 + h_\mathrm{img}^2}$. w_img: the image width, h_img: the image height.
6. Calculate backward errors by switching the first and the second images together with $\mathtt{H}_\mathrm{gt}^{-1}$ and $\mathtt{H}_\mathrm{est}^{-1}$.

The NSPT error has an advantage on interpretability because the error range is within $[0, 1]$ regardless of image resolutions. Since the perspective terms of a homography matrix, *i.e.* H_{31} and H_{32}, are relatively much smaller than the other elements, the NSPT error is a better metric than the difference of the Frobenius norm $\|\mathtt{H}_\mathrm{gt} - \mathtt{H}_\mathrm{est}\|_\mathrm{F}$.

6.1. Synthetic data evaluation

We evaluated the robustness of **DLT**, **NL**, **NLχ**, and **ConvexDLT** in the presence of unexpected outliers. Namely, we assume that some outliers survived RANSAC and are treated as inliers. **DLT** was used as the initial guess for **NL** and **NLχ** in this experiment.

Figure 5 illustrates configurations of the simulation. The first camera rotates around the origin with polar angle $\phi \in \{30°, 80°\}$ so that a square $[-1, 1, 0] \times [-1, 1, 0] \times [-1, 1, 0]$ is observed as a trapezoid. The four corner points of the square is projected onto a 1000×1000 image in each camera. We set two different diagonal corners of the two images as inliers to simulate implausible homographies that lead to a non-convex transformations between the two images. We added *true* inliers by varying the number of the
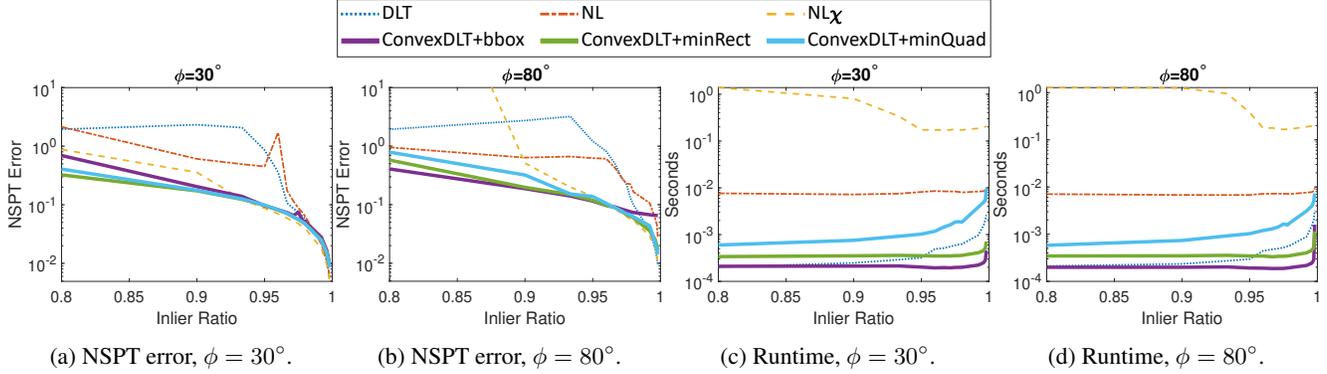
Figure 6: Quantitative results on synthetic data evaluation. All plots are semi-log graphs with respect to the y-axis.

points, $8 \leq n \leq 998$. Thus, the inlier ratio changes from $8/10 = 80\%$ to $998/1000 = 99.8\%$. *True* inliers were randomly generated in each trial and added by Gaussian noise of zero-mean with $\sigma = 2$ pixels. We conducted 1000 independent trials for each n, and measured the NSPT error.

Figure 6 shows the average of the NSPT error and computational time. First of all, the NSPT error evaluation in Figs. 6a and 6b indicates a notable result that the proposed **ConvexDLT** methods outperform the two nonlinear methods, **NL** and **NLχ**, for low inlier ratios and become comparable for high inlier ratios. The proposed methods are more accurate than **DLT** by more than an order of magnitude for the inlier ratio less than 98%. On the other hand, **ConvexDLT+bbox** for $\phi = 80°$ is not as robust as the other two methods for the inlier ratio close to 1. The reason can be considered to be the occurrence of cases where **ConvexDLT+bbox** degenerates, as shown in Fig. 4.

In the view of computational time, **ConvexDLT+bbox** is fastest among all the methods. Note that the number of the points n increases as the inlier ratio becomes higher in Figs. 6c and 6d, *e.g.* $n = 10$ at 80%, $n = 1000$ at 99.8%. Since the proposed generalized eigenvalue solver divides a 9×9 eigenvalue problem into a two partitioned matrix decomposition as described in Sec. 5.2, **ConvexDLT+bbox** is faster than **DLT**. **ConvexDLT+minRect** and **ConvexDLT+minQuad** are not as efficient as **ConvexDLT+bbox**. The two methods require finding a convex hull of the input points, of which computational complexity is $\mathcal{O}(n \log n)$. In addition to that, the complexity of the minimum bounding quadrilateral algorithm [7] is $\mathcal{O}(N^4)$, where N is the number of distinct edges of a convex hull. Due to this, **ConvexDLT+minQuad** is slower by an order of magnitude than **ConvexDLT+bbox**. On the other hand, the computational time of **NLχ** decreases as the inlier ratio becomes high. It seems to be strange, however, we infer that the increase in the inlier ratio improved the convergence of **NLχ**. This result suggests that **NLχ** is not stable due to the inequality constraints.

6.2. Real data evaluation

We evaluated the performance of the methods on a real image dataset, `HPatches`[2] [1], in this section. This dataset consists of 59 sequences of a planar image where each sequence has 6 images captured from different viewpoints. The ground-truth homography H_{gt} is provided from the first image to other five images in each sequence. Hence, there are $(6 - 1) \times 59 = 295$ image pairs in total.

We integrated **DLT** and **ConvexDLT** with the locally optimized RANSAC (LO-RANSAC) [6] and an M-estimator using graduated non-convexity [26]. LO-RANSAC was set to have a threshold by five pixels, a confidence by $p = 0.995$, and the maximum number of iterations by 2500. This configuration assumes that the inlier ratio of an image pair is higher than 21.45%. The inner iterations of the LO step was five, and the size of the LO sample was chosen by the larger of 12 or 50% of the number of tentative inliers. The signed area constraint [23] was optionally applied for fast sample rejection. We configured the M-estimator based on [26] but employed Tukey's biweight function, which showed a better performance than the truncated L2 function in the preliminary experiments.

For each image pair, we evaluated the methods as follows. We first obtained initial keypoint matches by SIFT implemented in OpenCV. The ground-truth inliers were determined as point matches of which Euclidean distance computed by H_{gt} is less than five pixels. Then, we executed LO-RANSAC and M-estimator on the initial point matches. We measured the following three criteria: the NSPT error, computational time, and F1-score of predicted inliers.

Figure 7 summarizes the quantitative results over 100 independent trials for each image pair. The result of the NSPT error indicates that the proposed **ConvexDLT** methods can improve the robustness of homography estimations for both LO-RANSAC and M-estimator. LO-RANSAC with **Sign** rejects a sample that is not geometrically consistent, how-
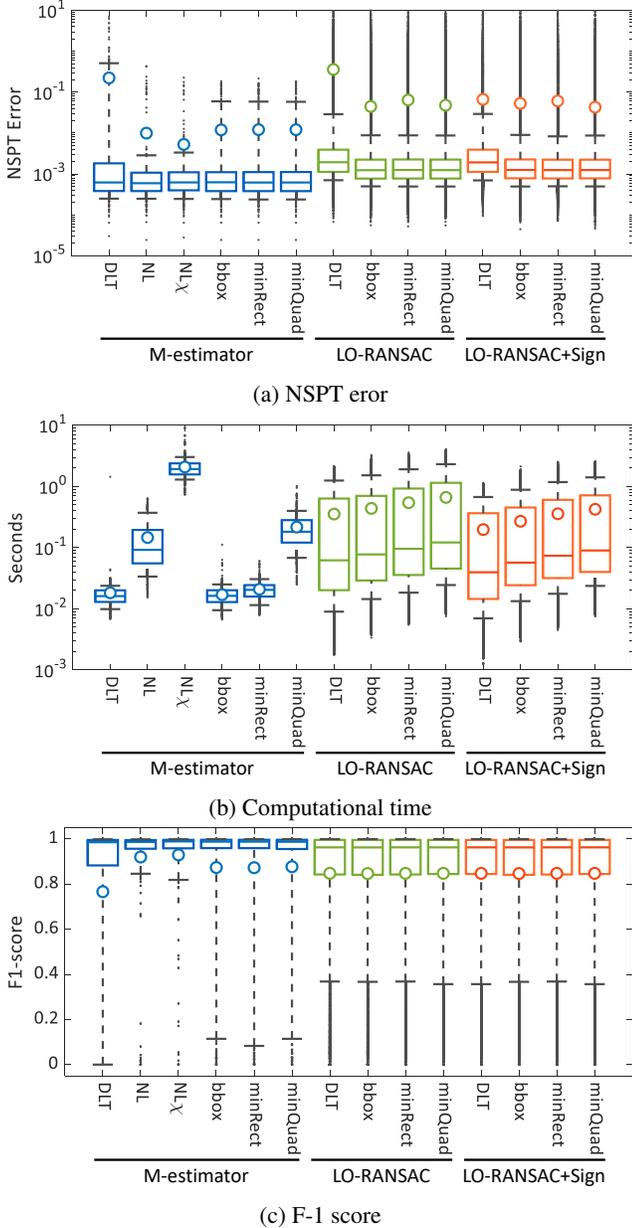
(a) NSPT eror

(b) Computational time

(c) F-1 score

Figure 7: Quantitative results on `HPatches` dataset. Box plots in blue, green, and orange correspond to the result of M-estimator, LO-RANSAC, and LO-RANSAC with the signed area constraint, respectively. A circle stands for the mean value for each method.

ever, there is no geometrical verification at the LO step. The proposed methods can give more accurate homographies than **DLT** if tentative inliers are contaminated by outliers at the LO step. Among the three **ConvexDLT**, there is no significant differences on estimation accuracy.

Comparing the LO-RANSACs with the M-estimators, we can observe an interesting result that the M-estimator

with the proposed methods outperforms the LO-RANSACs with **DLT** in all aspects of the three criteria. Particularly, the M-estimator with **ConvexDLT+bbox** or **ConvexDLT+minRect** achieve the fastest computation time with a small variance while giving the lowest NSPT error.

6.3. Discussion

Considering the trade-off between runtime and accuracy, we can conclude that **ConvexDLT+minRect** is the best balanced among the three approaches. **ConvexDLT+minQuad** is theoretically optimal; however, **ConvexDLT+minRect** has another advantage that many algorithms are publicly available for finding the minimum rotated rectangle, which have been studied for decades in computational geometry.

7. Limitations and Future Work

One thing to note is that the proposed constraint is not equivalent to the cheiral inequalities [15]. The proposed constraint ensures that the four corner points of a quadrilateral have the same projective orientation, *i.e.* $\text{sign}(H_{31}x_i + H_{32}y_i + H_{33}) = \text{sign}(H_{31}x_j + H_{32}y_j + H_{33})$. Therefore, the proposed methods cannot handle transformations that preserve a convexity, such as reflections. Another thing is that the proposed solvers are not invariant to exchanging two images. Since the constraint matrix S_1 is derived from an ellipse in one of the two images, the forward and (the inverse of) the backward homographies by the proposed solvers are generally not coincident. A quick workaround is to select a better one that gives the smaller reprojection error; however, more sophisticated approaches need to be investigated for a future research. It is also an interesting topic to incorporating the new constraint with deep-learning methods, *e.g.* [8, 27], to stabilize the training process.

8. Conclusion

In this paper, we revealed a new algebraic constraint on the classical homography estimation problem to enforce a homography matrix to be a transformation between two convex quadrilaterals. We proposed a generalized eigenvalue solver to satisfy the new constraint and demonstrated by experiments that the proposed method can improve estimation accuracy, stability, and computational efficiency in the presence of outliers. Moreover, we reported that an M-estimator with the proposed method showed superior performance to RANSAC in the real data experiment. M-estimator has been considered inferior to RANSAC in the computer vision community, while M-estimator has the advantage that it always returns an identical result for an identical input. We expect that this paper will provide an opportunity to revisit not only the homography estimation problem but also M-estimator.

References

[1] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5173–5182, 2017. 4327

[2] Daniel Barath and Zuzana Kukelova. Homography from two orientation-and scale-covariant features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1091–1099, 2019. 4321

[3] Daniel Barath, Michal Polic, Wolfgang Förstner, Torsten Sattler, Tomas Pajdla, and Zuzana Kukelova. Making affine correspondences work in camera geometry computation. In *European Conference on Computer Vision*, pages 723–740. Springer, 2020. 4326

[4] Selim Benhimane and Ezio Malis. Homography-based 2d visual tracking and servoing. *The International Journal of Robotics Research*, 26(7):661–676, 2007. 4321

[5] Eric Brachmann and Carsten Rother. Neural- Guided RANSAC: Learning where to sample model hypotheses. In *ICCV*, 2019. 4321, 4322

[6] Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In *Joint Pattern Recognition Symposium*, pages 236–243. Springer, 2003. 4321, 4322, 4327

[7] John D'Errico. A suite of minimal bounding objects, 2014. 4325, 4326, 4327

[8] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016. 4321, 4328

[9] Elan Dubrofsky and Robert J Woodham. Combining line and point correspondences for homography estimation. In *International Symposium on Visual Computing*, pages 202–213. Springer, 2008. 4321

[10] David Eppstein, Mark Overmars, Günter Rote, and Gerhard Woeginger. Finding minimum areak-gons. *Discrete & Computational Geometry*, 7(1):45–58, 1992. 4325

[11] 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. 4321, 4322

[12] Andrew Fitzgibbon, Maurizio Pilu, and Robert B Fisher. Direct least square fitting of ellipses. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):476–480, 1999. 4325

[13] 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. 4321

[14] Radim Halır and Jan Flusser. Numerically stable direct least squares fitting of ellipses. In *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, volume 98, pages 125–132. Citeseer, 1998. 4325

[15] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 4321, 4322, 4326, 4328

[16] Kenichi Kanatani, Hirotaka Niitsuma, and Prasanna Rangrajan. High accuracy homography computation without iterations. *Memoirs of the Faculty of Engineering, Okayama University*, 44:50–59, 2010. 4321

[17] Yasushi Kanazawa and Kenichi Kanatani. Image mosaicing by stratified matching. *Image and Vision computing*, 22(2):93–103, 2004. 4321

[18] Juho Kannala, Mikko Salo, and Janne Heikkilä. Algorithms for computing a planar homography from conics in correspondence. In *BMVC*, pages 77–86, 2006. 4321

[19] Zuzana Kukelova, Jan Heller, Martin Bujnak, and Tomas Pajdla. Radial distortion homography. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 639–647, 2015. 4321

[20] Stéphane Laveau and Olivier Faugeras. Oriented projective geometry for computer vision. In *European Conference on Computer Vision*, pages 147–156. Springer, 1996. 4323

[21] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 4321

[22] Ezio Malis and Manuel Vargas. Deeper understanding of the homography decomposition for vision-based control. (RR-6303), 2007. 4321

[23] Pablo Márquez-Neila, Javier López-Alberca, José M Buenaposada, and Luis Baumela. Speeding-up homography estimation in mobile devices. *Journal of Real-Time Image Processing*, 11(1):141–154, 2016. 4321, 4323, 4327

[24] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *European conference on computer vision*, pages 128–142. Springer, 2002. 4321

[25] Lionel Moisan, Pierre Moulon, and Pascal Monasse. Automatic Homographic Registration of a Pair of Images, with A Contrario Elimination of Outliers. *Image Processing On Line*, 2:56–73, 2012. https://doi.org/10.5201/ipol.2012.mmm-oh. 4323

[26] Gaku Nakano and Takashi Shibata. Fast and robust homography estimation by adaptive graduated non-convexity. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2556–2560. IEEE, 2019. 4321, 4322, 4327

[27] Ty Nguyen, Steven W Chen, Shreyas S Shivakumar, Camillo Jose Taylor, and Vijay Kumar. Unsupervised deep homography: A fast and robust homography estimation model. *IEEE Robotics and Automation Letters*, 3(3):2346–2353, 2018. 4321, 4328

[28] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: a universal framework for random sample consensus. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):2022–2038, 2012. 4323

[29] Prasanna Rangarajan and Panos Papamichalis. Estimating homographies without normalization. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 3517–3520. IEEE, 2009. 4321

[30] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011. 4321

[31] Shinya Sumikura, Mikiya Shibuya, and Ken Sakurada. Openvslam: a versatile visual slam framework. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2292–2295, 2019. 4321

[32] Richard Szeliski et al. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2007. 4321

[33] Godfried T Toussaint. Solving geometric problems with the rotating calipers. In *Proc. IEEE Melecon*, volume 83, page A10, 1983. 4325

[34] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. Pose tracking from natural features on mobile phones. In *Proc. 7th IEEE International Symposium on Mixed and Augmented Reality*, pages 125–134, Cambridge, UK, September 2008. Best Paper Award 2008, lasting impact award 2018. 4321

[35] Gehua Yang, Charles V Stewart, Michal Sofka, and Chia-Ling Tsai. Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE transactions on pattern analysis and machine intelligence*, 29(11):1973–1989, 2007. 4321

[36] Zhengyou Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and vision Computing*, 15(1):59–76, 1997. 4322

[37] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195, 1998. 4326

[38] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. 4321

[39] Dingfu Zhou, Yuchao Dai, and Hongdong Li. Ground-plane-based absolute scale estimation for monocular visual odometry. *IEEE Transactions on Intelligent Transportation Systems*, 21(2):791–802, 2020. 4321