This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Orthographic-Perspective Epipolar Geometry

Viktor Larsson[1], Marc Pollefeys[1,2], Magnus Oskarsson[3]

[1] Department of Computer Science, ETH Zurich
[2] Microsoft Mixed Reality and AI Lab Zurich
[3] Centre for Mathematical Sciences, Lund University *

Abstract

In this paper we consider the epipolar geometry between orthographic and perspective cameras. We generalize many of the classical results for the perspective essential matrix to this setting and derive novel minimal solvers, not only for the calibrated case, but also for partially calibrated and non-central camera setups. While orthographic cameras might seem exotic, they occur naturally in many applications. They can e.g. model 2D maps (such as floor plans), aerial/satellite photography and even approximate narrow field-of-view cameras (e.g. from telephoto lenses). In our experiments we highlight various applications of the developed theory and solvers, including Radar-Camera calibration and aligning Structure-from-Motion models to aerial or satellite images.

1. Introduction

In this paper we direct our attention at something that, at first glance, might seem like an exotic creature in the land of epipolar geometry, namely the essential matrix for mixed orthographic and perspective cameras. By this we mean the geometry of a two-view scene, where one camera is a fully calibrated perspective camera and the other is an orthographic camera. A schematic of the geometry is given in Figure 1. This case was first considered by Zhang et al. in [27] where the ortho-perspective essential matrix was derived. In this work we extend their analysis and derive a parallel theory to the classical results for the perspective essential matrix. Additionally we consider the case where the perspective camera is only partially calibrated (unknown focal length) or when it is a non-central (generalized) camera. For each of the cases we derive novel minimal solvers which allow for robust estimation in RANSAC [3] frameworks.

While optical systems that yield true orthographic projections are not commonplace, the orthographic camera is applicable in many other settings. For example, 2D-maps such as floor plans can be seen as an overhead orthographic image of the scene. Thus registering a camera to a 2D map is equivalent to relative pose estimation between an orthographic image and a perspective one. Orthographic projection can also approximate perspective projection in the case of narrow field-of-view or when the distance to the scene is large. In the experiments we will show that using this approximation can even be preferable to the full perspective model when the focal length needs to be estimated as well. Furthermore, while the camera models used for satellite or aerial photography are often complicated, they are also well-approximated by the simpler orthographic model.

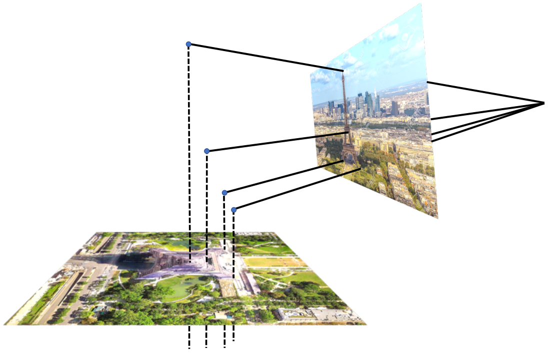

Figure 1. **Orthographic-Perspective Epipolar Geometry**

Related Work on Epipolar Geometry. Epipolar geometry deals with the geometry of two cameras viewing a scene. Often it is characterized in terms of the bifocal matching tensor (see Triggs [26]) which relates corresponding image points. For projective cameras this is the *fundamental matrix*. The matrix has only a single internal constraint (being rank-2) and can be minimally estimated from seven point correspondences [6]. For calibrated cameras the corresponding bifocal tensor is the *essential matrix*. While the essential matrix was first introduced by Longuet-Higgins [12], calibrated epipolar geometry was studied as early as 1883 by Hauck [7]. The essential matrix can be minimally estimated from five correspondences and there have been multiple solvers proposed in the literature (see

*Funded by the strategic research projects ELLIIT and eSSENCE, and the H2020 ECSEL JU project ADACORSA under GA 876019

e.g. [14, 22, 4]). Bifocal tensors have also been studied for non-perspective cameras. For example, Shapiro et al. [20] derived the fundamental matrix for affine cameras, which was later specialized for weak-perspective/orthographic in [21, 16]. Dai et al. [2] derived essential matrices for modelling rolling shutter effects. There are also works which consider heterogeneous camera setups; e.g. [24, 17] consider para-catadioptric/perspective and [8] studies the multifocal tensors for cameras with mixed dimensionality (2D camera vs. 1D line camera).

The work that is most related to ours is from Zhang et al. [27] which also considers the epipolar geometry of orthographic and perspective cameras. In [27] the authors derive the ortho-perspective essential matrix and identify its internal structure. They also propose a method for projecting a given matrix onto the set of ortho-perspective essential matrices. While the paper discusses the possibility of solving for the essential matrix from five point correspondences, no minimal solver is derived. In this paper we build on the analysis from [27] and derive new properties and results for the ortho-perspective essential matrix. We develop new internal constraints that are analogous so the classical trace-constraints and show how these can be used to derive a minimal solver.

In the supplementary material we provide additional discussion on related works for the applications we consider.

2. Mixed Perspective and Orthographic

We now direct our attention to the special case when one camera is perspective and the other is orthographic. Let

$$\mathbf{x}_p = (x,\ y,\ 1)^T \quad \text{and} \quad \mathbf{x}_o = (m_x,\ m_y,\ 1) \qquad (1)$$

be the image points in the perspective and orthographic cameras respectively. We assume that the perspective camera is calibrated and the image point is given in the normalized image plane. In the coordinate frame of the perspective camera, the 3D point corresponding to $\mathbf{x}_p$ is $\mathbf{X} = \lambda \mathbf{x}_p$ for some $\lambda > 0$. Let the orthographic camera be

$$P_o = \begin{bmatrix} \mathbf{r}_1^T & t_1 \\ \mathbf{r}_2^T & t_2 \end{bmatrix}, \quad \mathbf{r}_1^T \mathbf{r}_2 = 0,\ \|\mathbf{r}_1\| = \|\mathbf{r}_2\| = 1 \qquad (2)$$

From the orthographic projection we get the following

$$m_x = \lambda \mathbf{r}_1^T \mathbf{x}_p + t_1, \qquad (3)$$

$$m_y = \lambda \mathbf{r}_2^T \mathbf{x}_p + t_2. \qquad (4)$$

Eliminating λ we get $\mathbf{r}_1^T \mathbf{x}_p (m_y - t_2) = \mathbf{r}_2^T \mathbf{x}_p (m_x - t_1)$, which can be rewritten as

$$\mathbf{x}_o^T E \mathbf{x}_p = 0, \quad \text{where} \quad E = \begin{bmatrix} -\mathbf{r}_2^T \\ \mathbf{r}_1^T \\ t_1 \mathbf{r}_2^T - t_2 \mathbf{r}_1^T \end{bmatrix}. \qquad (5)$$

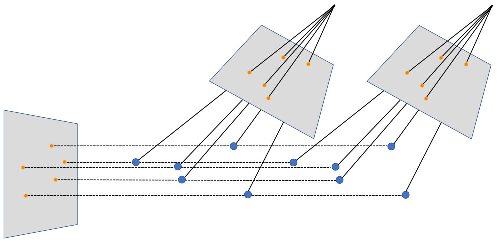

Figure 2. **Translational ambiguity.** The orthographic camera determines the scale of the reconstruction. However it is not possible to determine the relative translation of the cameras along the viewing direction of the orthographic camera.

This is the *ortho-perspective essential matrix* which was first identified in [27]. In the following sections we discuss various properties of the epipolar geometry. In the supplementary we also consider planar induced homographies between the orthographic and perspective image.

2.1. Scale/Translation Ambiguity

In classical epipolar geometry the scale of the scene is unobservable. In the ortho-perspective setting the global scale is fixed to the scale of the orthographic camera. Instead an ambiguity appears in the translation along the viewing direction of the orthographic camera (see Figure 2).

The ortho-perspective essential matrix and epipolar constraints are homogeneous in the scale of $\mathbf{r}_1$ and $\mathbf{r}_2$. It is therefore not possible to estimate the scaling factor in the case of scaled orthographic projection (weak-perspective) as this only rescales E. This ambiguity is directly coupled to the scale of the scene, as can be seen from (3)-(4) where any rescaling of $\mathbf{r}_1, \mathbf{r}_2$ can be compensated by the depth λ. Therefore we will in the remainder of the paper consider $\mathbf{r}_1$ and $\mathbf{r}_2$ as orthogonal vectors of the same length.

2.2. Internal Constraints on the Essential Matrix

From (5) it is clear that the essential matrix satisfies

$$\mathbf{e}_1^T \mathbf{e}_2 = 0, \quad \mathbf{e}_1^T \mathbf{e}_1 - \mathbf{e}_2^T \mathbf{e}_2 = 0, \quad det(E) = 0 \qquad (6)$$

where $\mathbf{e}_k^T$ is the kth row of E. These constraints were also derived in [27]. However, similar to perspective case there also exist internal constraints that are analogous to the classical trace constraints. In Section 4 we will show how these extra equations are very useful for deriving a minimal solver.

Theorem 1 (Ortho-perspective Trace Constraints.). *For a real non-zero matrix E the following are equivalent*

i) E is an ortho-perspective essential matrix (as in (5))

ii) E satisfies the constraints

$$2EE^TDE = tr\left(EE^TD\right)E, \qquad (7)$$

where $D = diag(1, 1, 0)$.

Remark. The structure of the constraint is perhaps not too surprising considering orthographic cameras are the limit as focal length tends to infinity and $D = \lim_{f\to\infty} K^{-1}$.

Proof. Let $\mathbf{e}_k^T$ denote kth row of E. By considering each row of the trace constraints (7) we get

$$2(\mathbf{e}_1^T\mathbf{e}_1)\mathbf{e}_1 + 2(\mathbf{e}_1^T\mathbf{e}_2)\mathbf{e}_2 = (\mathbf{e}_1^T\mathbf{e}_1 + \mathbf{e}_2^T\mathbf{e}_2)\mathbf{e}_1 \quad (8)$$

$$2(\mathbf{e}_2^T\mathbf{e}_1)\mathbf{e}_1 + 2(\mathbf{e}_2^T\mathbf{e}_2)\mathbf{e}_2 = (\mathbf{e}_1^T\mathbf{e}_1 + \mathbf{e}_2^T\mathbf{e}_2)\mathbf{e}_2 \quad (9)$$

$$2(\mathbf{e}_3^T\mathbf{e}_1)\mathbf{e}_1 + 2(\mathbf{e}_3^T\mathbf{e}_2)\mathbf{e}_2 = (\mathbf{e}_1^T\mathbf{e}_1 + \mathbf{e}_2^T\mathbf{e}_2)\mathbf{e}_3 \quad (10)$$

The implication $i \Rightarrow ii$ follows directly by inserting (5) into (8)-(10). To prove the converse assume that E satisfies the constraints. From (8) we get

$$2(\mathbf{e}_1^T\mathbf{e}_2)\mathbf{e}_2 = (\mathbf{e}_2^T\mathbf{e}_2 - \mathbf{e}_1^T\mathbf{e}_1)\mathbf{e}_1 \quad (11)$$

Thus either $\mathbf{e}_1 \parallel \mathbf{e}_2$ or both coefficients are zero, i.e.

$$\mathbf{e}_1^T\mathbf{e}_2 = 0, \quad \|\mathbf{e}_1\| = \|\mathbf{e}_2\|. \quad (12)$$

Now assume first that they are parallel, i.e. $\mathbf{e}_2 = \alpha\mathbf{e}_1$ for some $\alpha \in \mathbb{R}$. Inserting into (11) yields

$$2\alpha^2\|\mathbf{e}_1\|^2\mathbf{e}_1 = (\alpha^2 - 1)\|\mathbf{e}_1\|^2\mathbf{e}_1, \quad (13)$$

which implies $\alpha^2 = -1 \implies \alpha = \pm i$. However, by assumption E is real and thus $\mathbf{e}_1$ cannot be parallel to $\mathbf{e}_2$. Then (12) must hold and we have that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal vectors of the same length. Let $\mathbf{r}_1 = \mathbf{e}_2, \mathbf{r}_2 = -\mathbf{e}_1$. It remains to show that $\mathbf{e}_3$ is of the correct form. From (10),

$$2(\mathbf{e}_3^T\mathbf{r}_2)\mathbf{r}_2 + 2(\mathbf{e}_3^T\mathbf{r}_1)\mathbf{r}_1 = (\|\mathbf{r}_1\|^2 + \|\mathbf{r}_2\|^2)\mathbf{e}_3 \quad (14)$$

showing that $\mathbf{e}_3$ is indeed a linear comb. of $\mathbf{r}_1$ and $\mathbf{r}_2$. □

Remark. From the proof above we can see that trace constraints allow for complex solutions on the form

$$E = [\ \mathbf{a}^T;\ \pm i\mathbf{a}^T;\ \mathbf{b}^T\ ] \quad (15)$$

2.3. Twisted Pairs

For each regular essential matrix there are four consistent camera pairs. For the ortho-perspective essential matrix there are only two solutions corresponding to a sign change in $\mathbf{r}_1$ and $\mathbf{r}_2$. From equation (3)-(4) we can see that this corresponds to changing sign of the scalar λ. Thus, as is the case for the regular essential matrix, this ambiguity can be resolved by considering the cheirality constraints of the perspective camera. This is illustrated in Figure 3.

The next result shows that it possible to determine the correct sign *without* triangulating any points.

Theorem 2. *Given corresponding points $\mathbf{x}_o$ and $\mathbf{x}_p$, the sign of the ortho-perspective essential matrix E is consistent with positive perspective depth if and only if*

$$\mathbf{x}_o^T E\mathbf{e}_1\mathbf{e}_2^T\mathbf{x}_p > 0 \quad (16)$$

where $\mathbf{e}_k^T \in \mathbb{R}^3$ denotes the kth row of E.

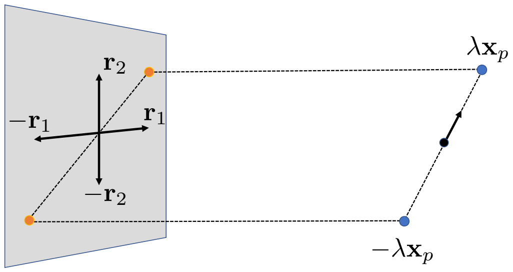

Figure 3. **Twisted pair for ortho-perspective essential matrices.** For each essential matrix there are two possible factorization which can be disambiguated using cheirality.

Proof. Multiplying (3) with $\mathbf{r}_1^T\mathbf{x}_p$ we get

$$(\mathbf{r}_1^T\mathbf{x}_p)(m_x - t_1) = \lambda(\mathbf{r}_1^T\mathbf{x}_p)^2 \quad (17)$$

thus $\lambda > 0$ if the left-hand side is positive. Using (5) we get

$$\mathbf{x}_o^T E\mathbf{e}_1\mathbf{e}_2^T\mathbf{x}_p = \mathbf{x}_o^T \begin{pmatrix} 1 \\ 0 \\ -t_1 \end{pmatrix} \mathbf{r}_1^T\mathbf{x}_p = (m_x - t_1)\mathbf{r}_1^T\mathbf{x}_p \quad (18)$$

□

Remark. From the proof we can see that there is an analogous constraint using the second equation (4), which has the opposite sign, i.e. $\mathbf{x}_o^T E\mathbf{e}_2\mathbf{e}_1^T\mathbf{x}_p < 0$. The constraint degenerates if $\mathbf{e}_2^T\mathbf{x}_p = 0$ (or $\mathbf{e}_1^T\mathbf{x}_p = 0$ respectively) in which the other constraint can be used. Note that if both are zero then $\mathbf{x}_p = \gamma\mathbf{r}_3$ which is the right epipole (see Section 2.4).

2.4. Other Properties of the Essential Matrix

The following properties of the ortho-perspective essential matrix E can easily be verified by explicit calculation.

Factorization. The matrix E can be factorized as

$$E = [(t_1, t_2, 1)]_\times DR = \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ -t_2 & t_1 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \end{bmatrix} \quad (19)$$

Epipoles. The epipoles of E are given by

$$E\mathbf{e}_p = 0 \implies \mathbf{e}_p = \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad (20)$$

$$E^T\mathbf{e}_o = 0 \implies \mathbf{e}_o = (t_1, t_2, 1)^T \quad (21)$$

Singular Values. Assuming E is scaled such that first row is a unit vector the singular values of E are

$$\sigma_1 = \sqrt{1 + t_1^2 + t_2^2}, \quad \sigma_2 = 1, \quad \sigma_3 = 0 \quad (22)$$

This was also shown in [27].

Remark. For $t_1 = t_2 = 0$ the singular values become $\sigma_1 = \sigma_2 = 1$, $\sigma_3 = 0$. In this case that the ortho-perspective essential matrix is also a perspective essential matrix, corresponding to the same relative rotation but with forward translation, i.e. $\mathbf{t} = (0,0,1)^T$. The left epipole is then in the image center, yielding radial epipolar lines. To understand why the matrices coincide in this configuration, note that the differences between the orthographic and the perspective projection is only in the radial scaling, which in this case is also along the epipolar lines.

2.5. Projection to the Essential Manifold

In some applications it is desirable to find the closest essential matrix to a given 3×3 matrix, i.e. to solve

$$\min_{E} \|E - \hat{E}\|_F^2 \quad \text{s.t. } \textit{E is an essential matrix.} \tag{23}$$

for a given matrix $\hat{E} \in \mathbb{R}^{3\times 3}$. This is for example useful as a post-processing step when estimating E from non-minimal number of points using DLT [5]. For regular essential matrices this projection has a nice closed form expression in terms of the singular value decomposition (by simply setting the singular values to 1,1,0). Unfortunately things are not as easy in the ortho-perspective case. Instead we now present a simple two-step approach which approximately solves (23) for ortho-perspective essential matrices.

1. Estimating the right epipole. We start by estimating the right epipole from $\hat{E}$ by minimizing

$$\min_{\mathbf{e}_p} \|\hat{E}\mathbf{e}_p\|_F^2 \quad \text{s.t.} \quad \|\mathbf{e}_p\| = 1 \tag{24}$$

This problem has a closed form solution given by the right singular vector corresponding to the smallest singular value.

2. Projection with fixed epipole. Next we solve the projection assuming the right epipole is known, i.e.

$$\min_{E} \|E - \hat{E}\|_F^2 \quad \text{s.t. } \textit{E o-p. essential and } E\mathbf{e}_p = 0 \tag{25}$$

Let $B \in \mathbb{R}^{3\times 2}$ be an orthonormal basis for the vectors orthogonal to $\mathbf{e}_p$, i.e. $B^T\mathbf{e}_p = 0$ and $B^TB = I_2$. The ortho-perspective essential matrices that have $\mathbf{e}_p$ as a right epipole can then be parameterized as

$$E = \begin{bmatrix} \multicolumn{2}{c}{Q} \\ x & y \end{bmatrix} B^T, \quad Q \in \mathbb{R}^{2\times 2},\ Q^TQ = \gamma I \tag{26}$$

The cost in (25) can then be simplified as

$$\left\| \begin{bmatrix} \multicolumn{2}{c}{Q} \\ x & y \end{bmatrix} B^T - \hat{E} \right\|_F^2 = \left\| \begin{bmatrix} \multicolumn{2}{c}{Q} \\ x & y \end{bmatrix} - \hat{E}B \right\|_F^2 \tag{27}$$

The optimization problem now separates in the unknowns (Q, x and y) and we can solve for them independently. While x, y are unconstrained and given directly as the third row of $\hat{E}B$, we have constraints that Q should be a scaled orthogonal matrix. Fortunately this has a closed form solution given by computing the SVD of the top 2×2 block of $\hat{E}B$ and replacing the singular values with their average.

Once we have recovered Q, x and y we can compute the scale γ and convert back into the parameters from (5) with

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = \frac{1}{\gamma} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} QB^T, \quad \begin{bmatrix} t_1 & t_2 \end{bmatrix} = -\frac{1}{\gamma^2} \begin{bmatrix} x & y \end{bmatrix} Q^T.$$

While this projection step does not solve (23) optimally, we found that it gives very good solutions in practice. In Section 5.1.1 we show experimental results verifying this.

In [27] the authors proposed a similar approach for projecting onto an ortho-perspective essential matrix which instead first solve for the epipole in the orthographic camera (or equivalently the translation). To recover the rotation the method then computes another 3×3 singular value decomposition (compared to 2×2 in the proposed method). Furthermore, as we will show in the experimental evaluation (Section 5.1.1), this approach yields less accurate estimates.

2.6. Unknown Focal Length

So far we have assumed that the perspective camera was calibrated. We now turn our focus to the scenario where the perspective focal length f_p is unknown and needs to be estimated. From the fundamental matrix F we get the essential matrix by multiplying with K from the right, $E = FK$, with $K = \operatorname{diag}(f_p, f_p, 1)$. Inserting this expression for E in (7) and multiplying with K^{-1} from the right gives the following constraint

$$2FKK^TF^TDF = \operatorname{tr}\left(FKK^TF^TD\right)F. \tag{28}$$

In this expression the focal length only appears as $\beta = f_p^2$, and we can rewrite the expression in the following way;

$$FKK^TFD = \beta FDF^T + F(I-D)F^T = \beta A + B. \tag{29}$$

We can now rewrite (28) as

$$2\beta AF + 2BF = \beta\operatorname{tr}(A)F + \operatorname{tr}(B)F, \tag{30}$$

which can be written

$$\underbrace{\begin{bmatrix} 2A\mathbf{f}_i - \operatorname{tr}(A)\mathbf{f}_i & 2B\mathbf{f}_i - \operatorname{tr}(B)\mathbf{f}_i \end{bmatrix}}_{M_i} \begin{bmatrix} \beta \\ 1 \end{bmatrix} = \mathbf{0}, i = 1,2,3. \tag{31}$$

where $\mathbf{f}_i$ is the ith column of F. Since (31) should have a solution, we get that necessary conditions for a fundamental matrix F with an unknown focal length are that all 2×2 sub-determinants of M_i should vanish. These constraints are nine polynomials of total degree six in the entries of F, and define together with the rank-2 constraint, the manifold of possible ortho-perspective fundamental matrices with an unknown focal length. It is also easy to verify (*e.g.* using a computer algebra system such as Macaulay2) that the ideal associated with this manifold is generated by the subset of (three) equations for $i = 3$ together with the rank-2 constraint. All three equations include the factor $\mathbf{f}_3^TD\mathbf{f}_3$ which can be divided away. So in essence the manifold is generated by three polynomials of total degree four and one of degree three (the rank constraint). This is similar to the case for the ordinary essential matrix with one-sided unknown focal length found in [9].

3. Non-Central Cameras

In the previous sections we considered the case of a central perspective camera, i.e. the viewing rays of the perspective camera intersect in a single point (the camera center). Now we extend our analysis to the setting of non-central cameras (sometimes called generalized cameras) where the viewing rays do not intersect. This can for example model a calibrated multi-camera system, or as we will show later, it can be used to register a full Structure-from-Motion model to the orthographic image. Each viewing ray can then originate from different cameras in the Structure-from-Motion reconstruction and thus have different camera centers.

Now consider a correspondence between an orthographic camera and a non-central camera. Again let $\mathbf{x}_o$ denote the image point in the ortho-image and parameterize the viewing ray in the non-central cameras as $\lambda \mathbf{x}_p + \mathbf{c}_p$. In contrast to a central camera, the non-central camera fixes the scale of the reconstruction. Thus to register to the ortho-image (which also fixes the scale) we need parameterize the relative scale change s. The projection equations are

$$m_x = \mathbf{r}_1^T(\lambda \mathbf{x}_p - s\mathbf{c}_p) + t_1, \quad m_y = \mathbf{r}_2^T(\lambda \mathbf{x}_p - s\mathbf{c}_p) + t_2 \quad (32)$$

Eliminating the depth λ we get

$$\mathbf{r}_1^T \mathbf{x}_p(m_y - t_2 + s\mathbf{r}_2^T\mathbf{c}_p) - \mathbf{r}_2^T \mathbf{x}_p(m_x - t_1 + s\mathbf{r}_1^T\mathbf{c}_p) = 0 \quad (33)$$

This constraint can be rewritten as $\mathbf{x}_o^T E \mathbf{x}_p =$

$$-s(\mathbf{r}_1^T\mathbf{x}_p)(\mathbf{r}_2^T\mathbf{c}_p) + s(\mathbf{r}_2^T\mathbf{x}_p)(\mathbf{r}_1^T\mathbf{c}_p) = \quad (34)$$

$$s\mathbf{c}_p^T(\mathbf{r}_1\mathbf{r}_2^T - \mathbf{r}_2\mathbf{r}_1^T)\mathbf{x}_p = -s\mathbf{c}_p^T[\mathbf{r}_3]_\times \mathbf{x}_p \quad (35)$$

where $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$. The generalized epipolar constraint is

$$\mathbf{x}_o^T E \mathbf{x}_p + s\mathbf{c}_p^T [\mathbf{r}_3]_\times \mathbf{x}_p = 0 \quad (36)$$

This constraint is analogous to the classical generalized epipolar constraint derived in [23] where it was derived using Plücker lines. In the supplementary material we show how (36) can also be derived in a similar fashion.

4. Minimal Solvers

In this section we derive minimal solvers for the ortho-perspective essential matrix. Due to the very similar structure as the regular essential matrix, similar solution strategies are applicable. The ortho-perspective essential matrix has five degrees of freedom and can thus be minimally estimated from five point correspondences.

$$\begin{array}{r|cccccccccccc|cccccccc|}
 & x^3 & x^2y & xy^2 & y^3 & x^2z & xyz & y^2z & xz^2 & yz^2 & z^3 & x^2 & xy & y^2 & xz & yz & z^2 & x & y & z & 1 \\ \hline
f_1 & 1 & & & & & & & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_1\ \text{———}} \\
f_2 & & 1 & & & & & & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_2\ \text{———}} \\
f_3 & & & 1 & & & & & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_3\ \text{———}} \\
f_4 & & & & 1 & & & & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_4\ \text{———}} \\
f_5 & & & & & 1 & & & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_5\ \text{———}} \\
f_6 & & & & & & 1 & & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_6\ \text{———}} \\
f_7 & & & & & & & 1 & & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_7\ \text{———}} \\
f_8 & & & & & & & & 1 & & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_8\ \text{———}} \\
f_9 & & & & & & & & & 1 & & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_9\ \text{———}} \\
det(E) & & & & & & & & & & 1 & & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_{10}\ \text{———}} \\
\mathbf{e}_1^T\mathbf{e}_2 & & & & & & & & & & & 1 & & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_{11}\ \text{———}} \\
|\mathbf{e}_1|^2 - |\mathbf{e}_2|^2 & & & & & & & & & & & & 1 & \multicolumn{8}{c|}{\text{———}\ \mathbf{m}_{12}\ \text{———}} \\ \hline
\end{array}$$

Figure 4. **Elimination template for o.p. essential matrix solver**. The first nine equations f_i are the trace-constraints (7) and the other equations are the original constraints on E from (6).

Each point correspondence yields one linear constraint, $\mathbf{x}_o^T E \mathbf{x}_p = 0$. From five correspondences we can extract a 4-dimensional linear space of possible 3×3 matrices,

$$E = E_1 x + E_2 y + E_3 z + E_4 w \quad (37)$$

Since the constraints are homogeneous we can fix the scale by setting $w = 1$. Inserting (37) into the trace constraints (7) and $\det(E) = 0$ yields 10 cubic equations in x, y, z. This equation system has nearly identical structure to classical five-point algorithm and indeed we find that this system also has 10 solutions. It would here be possible to directly apply the same solution strategy as in the classical five point solvers (see e.g. [14, 22, 4]). However, in the ortho-perspective setting we know that the trace constraints allow for additional complex solutions of the form (15). These solutions can be removed by adding the additional constraints from (6). Except for the determinant, these equations are quadratic in terms in the nullspace parameters x, y, z. Similar to [22] we derive an action matrix-based solver. We stack all equations into a matrix where each column contains the coefficients for one monomial. Performing linear elimination on this allow us to directly recover the action matrix for multiplication with x as $M_x =$

$$[-\mathbf{m}_3; -\mathbf{m}_5; -\mathbf{m}_6; -\mathbf{m}_8; -\mathbf{m}_{11}; -\mathbf{m}_{12}; \mathbf{u}_2; \mathbf{u}_5] \quad (38)$$

where $\mathbf{m}_i$ follows the notation from Figure 4 and $\mathbf{u}_i$ is the ith canonical basis vector, i.e. $\mathbf{u}_1 = (1, 0, \ldots, 0)$ etc. From the eigenvectors of M_x we can then recover the solutions. Note that this approach is identical to [22], except that the two additional constraints allow us to eliminate two extra monomials (using the last 8, instead of the last 10).

Remark. We also experimented with only using the original constraints (6) derived in [27]. Applying the method from Larsson et al. [10] yields a solver with template size 42×54 returning 12 solutions. The four additional solutions are complex and have $\mathbf{e}_1^T\mathbf{e}_1 = \mathbf{e}_2^T\mathbf{e}_2 = 0$. More comparisons can be found in the supplementary material.

4.1. Unknown focal length

For the case with an unknown focal length for the perspective camera we have an additional parameter to estimate, and hence we need minimally six point correspondences to solve for the relative pose. Each point correspondence yields one linear constraint, $\mathbf{x}_o^T F \mathbf{x}_p = 0$, where $E = FK$, with $K = diag(f_p, f_p, 1)$. From six correspondences we can extract a three-dimensional linear space of possible 3×3 matrices,

$$F = F_1 x + F_2 y + F_3 z. \tag{39}$$

We again fix the scale by setting $z = 1$. Inserting the expression of $F(x, y)$ into the unknown focal length constraints leads to a system of three fourth degree polynomials in (x, y) and one third degree polynomial $(p_4(x, y))$. These polynomials contain the parameters (x, y) in 15 monomials. The problem can easily be verified to have at most nine solutions. We can construct a solver based on an action matrix in the following way. We take y as action variable and

$$b = \begin{bmatrix} x^2y & xy^2 & y^3 & x^2 & xy & y^2 & x & y & 1 \end{bmatrix}, \tag{40}$$

as a linear basis for the quotient space. After multiplication of the action variable y, we have the monomial vector yb, which will still be contained in the 15 monomials from the initial equations. We can add the two polynomials $yp_4(x, y)$ and $xp_4(x, y)$ to our equations. Since $p_4(x, y)$ is of degree three, these two new polynomials will be of degree four, and not add any new monomials. These six polynomials will directly give a compact template from which our action matrix can be extracted by Gaussian elimination, in the same manner as in Section 4. From the eigenvectors of the action matrix we find x and y and from these we recover F from (39). The focal length is then found from (31).

4.2. Non-Central

We now derive a solver for the generalized case discussed in Section 3. The problem (including estimating the relative scale) is minimal with six correspondences. Each correspondence gives a linear constraint (36) on E and $s\mathbf{r}_3$. From these we can extract a $12 - 6 = 6$ dimensional nullspace basis, such that the solutions can be written

$$[E \;\; s\mathbf{r}_3] = \sum_{k=1}^{6} \alpha_k B_k, \quad B_k \in \mathbb{R}^{3\times4}. \tag{41}$$

Since the constraints are homogeneous we can fix the scale by setting $\alpha_6 = 1$. From the trace constraints (7) and (6) we get polynomial equations in $\alpha_1, ..., \alpha_5$. However, $s\mathbf{r}_3$ is not independent from E and we must also enforce $E\mathbf{r}_3 = 0$, yielding three additional quadratic equations in α_k. Applying the solver generator from Larsson et al. [10] we find that the problem has 16 solutions. The resulting solver performs linear elimination on a 170×186 linear system followed by solving a 16×16 eigenvalue problem.

Remark. Note that the proposed 6-point solver does not correspond to the regular generalized 6-point relative pose solver since it also estimates a scale between the reconstructions. As far as we are aware, there is currently no known minimal solution for the regular generalized relative pose problem with unknown scale (which would use 7 points).

5. Experiments

5.1. Synthetic Experiments

To evaluate the numerical stability of the proposed solvers we generate synthetic scenes. We uniformly sample image points in a 1000×1000 image for the perspective camera. The field-of-view for the perspective camera is uniformly sampled from $[45°, 90°]$. The points are then back-projected to a random depth and reprojected into a randomly oriented orthographic camera. The orthographic camera's translation and the scene scale is then set such that orthographic image also becomes 1000×1000. For the generalized solver we randomly sample different camera centers for each correspondence. For the experiment we generate 10,000 random noise-free instances where we apply the solvers from Sections 4–4.2. Figure 5 show the distribution of the errors in the estimated epipolar geometries. All of the proposed solvers yield stable estimates. The runtime of our MATLAB implementations (on a 2.5 GHz i7 Macbook Pro) are 0.5ms, 0.3ms and 0.85ms for the three full solvers.

5.1.1 Evaluation of Essential Matrix Projection

In Section 2.5 we presented a two-step projection method which approximately finds the closest ortho-perspective essential matrix to a given 3×3 matrix. We now evaluate how close this approximation comes to the optimal projection. Similarly to the previous section we generate random synthetic instances. For each image we add zero-mean gaussian noise to the image correspondences and estimate the essential matrix using DLT [5] from 25 point correspondences. Due to the noise, these matrices will not satisfy the internal constraints. We compare projecting with the method from Section 2.5 with the method from Zhang et al. [27]. We also include the result of local refinement initialized both from the ground truth essential matrix (used to generate the scene) and from the result the two methods, and the best result is kept (denoted *GT*). Figure 6 shows the average relative errors ($\|E - \hat{E}\|_F / \|\hat{E}\|_F$). The proposed method yields results very close to the ground-truth projection.

5.2. Application: Approximating Cameras with Large Focal Lengths as Orthographic

For cameras with very large focal lengths (narrow field-of-view) the viewing rays are approximately parallel. In this section we experiment with approximating such cam-

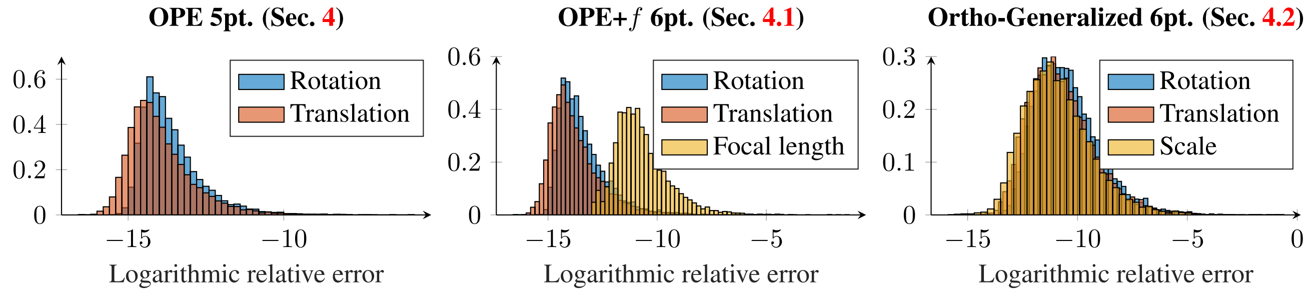

Figure 5. **Numerical Stability Evaluation.** The figures show the distribution of the $\log_{10}$ relative errors for 10,000 synthetic instances (see Section 5.1 for more details). Each of the three proposed solvers yield stable estimates.

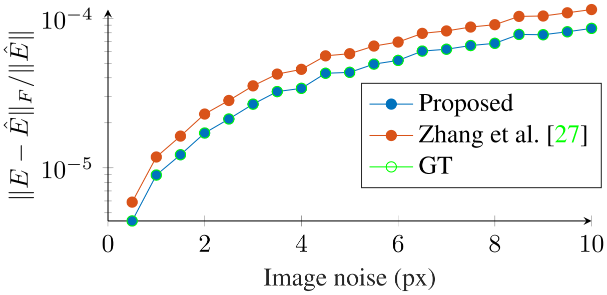

Figure 6. **Projection onto Essential Manifold.** The average relative distance $\|E - \hat{E}\|_F / \|\hat{E}\|_F$ after projection for varying noise. The proposed projection method gives better estimates compared to the method from Zhang et al.[27] and is very close to optimal.

eras with the orthographic projection model. We consider a dataset containing 1557 DSLR images with varying focal lengths (24mm, 50mm, 105mm, 150mm, 300mm and 600mm). We create a psuedo-ground truth by reconstructing the scene using COLMAP [18]. For the experimental setup we then select pairs of images; one with a shorter focal length (24mm, 50mm, 105mm) and one with a larger (150mm, 300mm, 600mm). The image with the larger focal length we approximate as an orthographic camera. We consider all pairs with at least 100 (SIFT [13]) matches which gave 10237 (150mm), 9391 (300mm) and 3313 (600mm) pairs respectively. Note that the extreme difference in field-of-view (105.8° for 24mm and 3.4° for 600mm) leads to some very challenging image pairs (see Figure 7).

For each image pair we estimate the epipolar geometry using RANSAC with the ortho-perspective essential matrix solvers; 5-point (OPE; Section 4) and the 6-point (OPE+f_p; Section 4.1) which estimates focal length for the perspective camera. The RANSAC is limited to 1000 iterations and the best model is selected with MSAC-scoring [25] on the symmetric epipolar error. We also compare with solvers that try to jointly estimate the focal length; the 6-point solver from [9] (f_o+E) which estimates one-sided focal length (used to estimate the larger focal length) and the 7-point fundamental matrix solver [6] ($F = f_o$+E+f_p) which estimates both focal lengths. For the solvers which require partial calibration (OPE and f_o+E) we use the focal lengths from the EXIF tags.

		OPE	f_o+E	OPE+f_p	F	E
f_o		5pt.	6pt. [9]	6pt.	7pt. [6]	5pt. [14]
150mm (N=10237)	AUC10	16.05	**44.52**	18.66	**44.13**	69.53
	F1	0.87	**0.92**	0.90	**0.93**	0.90
300mm (N=9391)	AUC10	27.13	**34.71**	25.36	**31.21**	67.05
	F1	0.91	**0.93**	0.93	**0.94**	0.92
600mm (N=3313)	AUC10	**37.78**	21.76	**29.20**	20.18	59.09
	F1	0.92	**0.93**	0.93	**0.94**	0.92

Table 1. **Ortho-approximation for large focal lengths.** The table shows the AUC for 10 degrees rotation error and the F1 score of the inliers w.r.t. the ground truth inliers. Orthographic approximation works better for large focal lengths whereas the performance degrades for the solver which tries to estimate the focal length (OPE vs. f_o+E). The same trend occurs when the focal length in the *perspective* camera is unknown (OPE+f_p vs. F).

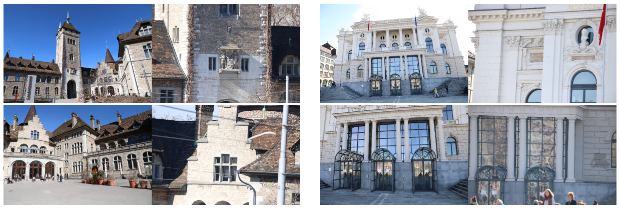

Figure 7. **Example image pairs** from the evaluation in Section 5.2. The focal lengths are between 24mm and 600mm.

Table 1 shows the area-under-curve (AUC) for the rotation error (i.e. the number of successful trials up to some threshold as a percentage of the complete square) and the F1 score for the inlier classification (compared to the structure-from-motion ground truth).

The experiment shows that as the focal length increases the orthographic approximation improves. In addition to requiring one less point-correspondence, the ortho-perspective solvers show improved accuracy for large focal lengths compared to the fully-perspective counterparts. We believe this is partially due to the difficulty of estimating the focal length for the extremely narrow field-of-view cameras. For comparison we also show the result of estimating the regular essential matrix [14] (using known intrinsics).

5.3. Application: 2D Radar Calibration

The last years have seen a renewed interest in radar applications, especially in combination with image data. The

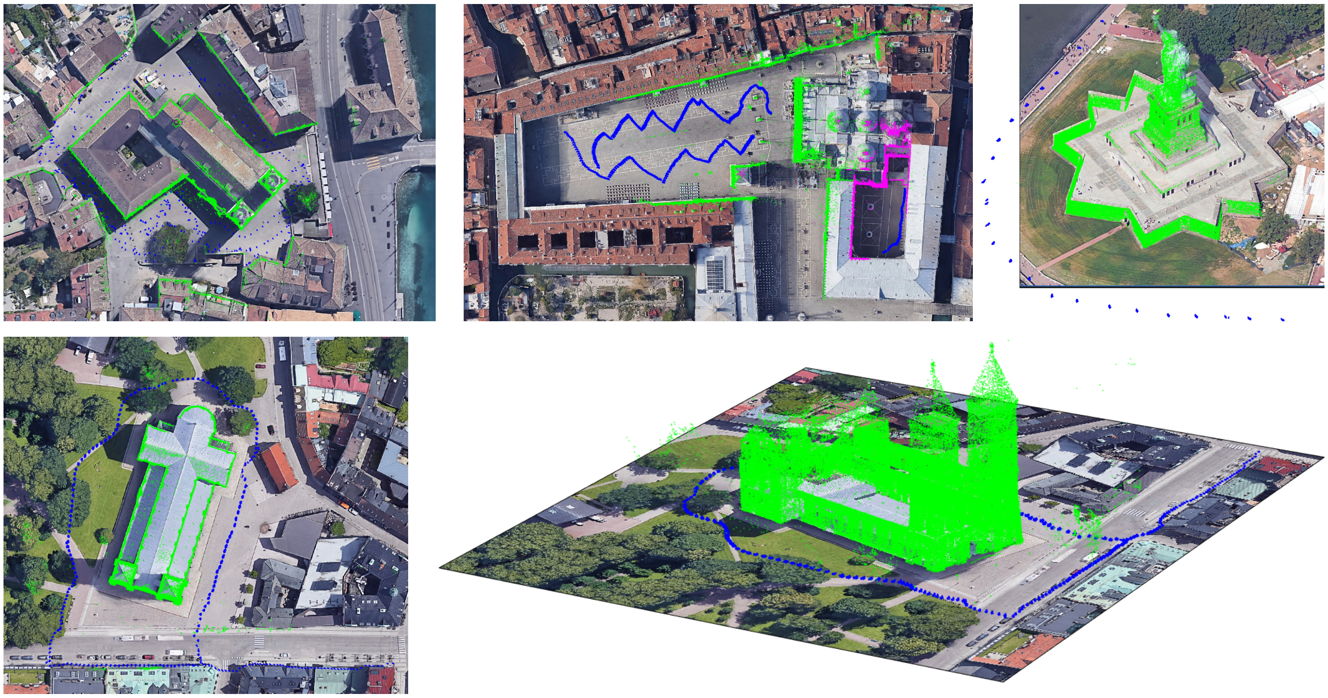

Figure 8. **Aligning Structure-from-Motion Reconstructions with Orthographic Images.** The images show SfM reconstructions (3D points in green, cameras in blue) overlayed on ortho-photo. *Top left:* Grossmünster Church [11]. *Top middle:*. Two disjoint reconstructions of the San Marco square (green) and the Doge Palace (magenta) registered to the same orthophoto. *Top right:* Statue of Liberty [15] registered to a 45° degree satellite image. *Bottom:* Lund Cathedral [15]. All ortho-images are taken from Google Maps [1]

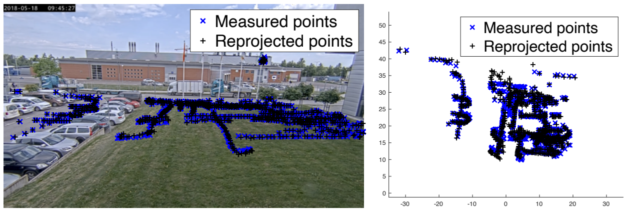

Figure 9. Reprojection errors for the radar experiment using the proposed method. *Left:* Camera image. *Right:* Radar image.

attraction of the combination of radar and image data is their complementary nature and failure modes.We show here how our theory can be applied to relative calibration of a 2D radar with a camera. For our calibration setup, persons walking were tracked both in the radar and a calibrated camera, and the goal is to automatically find the relative pose between the camera and the radar. We model the radar using an orthographic camera, and use our minimal 5-point solver in RANSAC. Reprojection results can be seen in Figure 9. See the supplementary for more results and details.

5.4. Application: Aligning Structure-from-Motion Reconstruction to Overhead Images

In this section we consider the problem of aligning Structure-from-Motion reconstructions with orthographic images using the generalized solver from Section 4.2. We assume that we have a sparse set of 2D-2D correspondences between different images in the SfM reconstruction and the orthographic image. For the experiment we manually created around 50-100 correspondences in total. However these correspondences might be automatically found using aerial to ground matching methods (see e.g. [19]) depending on the viewing angle of the orthographic image. We consider five datasets from [15, 11]. Using COLMAP [18] we reconstruct the scenes. The cameras from each reconstruction is now considered as a single generalized camera which we want to register to the orthographic image. Using the solver from Section 4.2 in RANSAC we estimate the similarity transform aligning the reconstruction with the ortho-image. Figure 8 shows the aligned reconstructions. In Figure 8 (Top middle) we show two different reconstructions co-registered to the same ortho-image.

6. Conclusions

We have in this paper presented a unifying theory for epipolar geometry in the mixed case of orthographic and perspective projections. We have derived a number of basic properties of the ortho-perspective essential that can be used to construct efficient minimal solvers for the calibrated case, the case with unknown focal length and for generalized cameras. These solvers can be used for bootstrapping estimation in a number of different applications, including aligning Structure-from-Motion reconstructions with orthographic cameras, approximating cameras with large focal lengths, and in radar and camera extrinsic calibration.

References

[1] Google maps. maps.google.com. 8

[2] Y. Dai, H. Li, and L. Kneip. Rolling shutter camera relative pose: Generalized epipolar geometry. In *Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[3] M. A. Fischler and R. 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. 1

[4] R. Hartley and H. Li. An efficient hidden variable approach to minimal-case camera motion estimation. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 34(12):2303–2314, 2012. 2, 5

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, USA, 2003. 4, 6

[6] R. I. Hartley. Projective reconstruction and invariants from multiple images. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 1994. 1, 7

[7] G. Hauck. Neue constructionen der perspective und photogrammetrie.(theorie der trilinearen verwandtschaft ebener systeme, i. artikel.). *Journal für die reine und angewandte Mathematik*, 1883(95):1–35, 1883. 1

[8] K. Kozuka and J. Sato. Multiple view geometry for mixed dimensional cameras. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2008. 2

[9] Z. Kukelova, J. Kileel, B. Sturmfels, and T. Pajdla. A clever elimination strategy for efficient minimal solvers. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 5, 7

[10] V. Larsson, K. Astrom, and M. Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 5, 6

[11] V. Larsson, N. Zobernig, K. Taskin, and M. Pollefeys. Calibration-free structure-from-motion with calibrated radial trifocal tensors. In *European Conference on Computer Vision (ECCV)*, 2020. 8

[12] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 1981. 1

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 2004. 7

[14] D. Nistér. An efficient solution to the five-point relative pose problem. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 26(6):756–770, 2004. 2, 5, 7

[15] C. Olsson and O. Enqvist. Stable structure from motion for unordered image collections. In *Scandinavian Conference on Image Analysis (SCIA)*, 2011. 8

[16] M. Oskarsson. Two-view orthographic epipolar geometry: Minimal and optimal solvers. *Journal of Mathematical Imaging and Vision (JMIV)*, 2018. 2

[17] L. Puig, P. Sturm, and J. J. Guerrero. Hybrid homographies and fundamental matrices mixing uncalibrated omnidirectional and conventional cameras. *Machine Vision and Applications (MVA)*, 2013. 2

[18] J. L. Schonberger and J.-M. Frahm. Structure-from-motion revisited. In *Computer Vision and Pattern Recognition (CVPR)*, 2016. 7, 8

[19] Q. Shan, C. Wu, B. Curless, Y. Furukawa, C. Hernandez, and S. M. Seitz. Accurate geo-registration by ground-to-aerial image matching. In *International Conference on 3D Vision (3DV)*, 2014. 8

[20] L. S. Shapiro, A. Zisserman, and M. Brady. 3d motion recovery via affine epipolar geometry. *International Journal of Computer Vision (IJCV)*, 1995. 2

[21] I. Shimshoni, R. Basri, and E. Rivlin. A geometric interpretation of weak-perspective motion. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 1999. 2

[22] H. Stewenius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006. 2, 5

[23] H. Stewénius, M. Oskarsson, K. Aström, and D. Nistér. Solutions to minimal generalized relative pose problems. In *Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS)*, 2005. 5

[24] P. Sturm. Mixing catadioptric and perspective cameras. In *Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS)*, 2002. 2

[25] P. H. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding (CVIU)*, 2000. 7

[26] B. Triggs. Matching constraints and the joint image. In *International Conference on Computer Vision (ICCV)*, 1995. 1

[27] Z. Zhang, P. Anandan, and H.-Y. Shum. What can be determined from a full and a weak perspective image? In *International Conference on Computer Vision (ICCV)*, 1999. 1, 2, 3, 4, 5, 6, 7