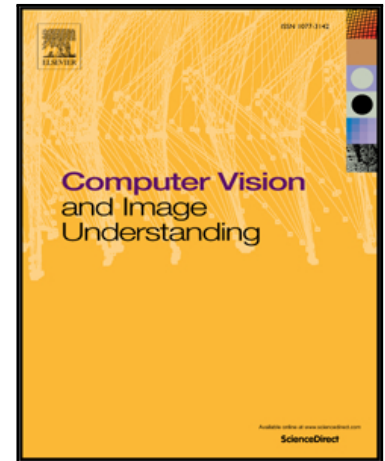


## Accepted Manuscript

A differential geometry approach to camera-independent image correspondence

József Molnár, Iván Eichhardt

PII: S1077-3142(18)30017-1  
DOI: [10.1016/j.cviu.2018.02.005](https://doi.org/10.1016/j.cviu.2018.02.005)  
Reference: YCVIU 2669



To appear in: *Computer Vision and Image Understanding*

Received date: 22 March 2017  
Revised date: 11 October 2017  
Accepted date: 8 February 2018

Please cite this article as: József Molnár, Iván Eichhardt, A differential geometry approach to camera-independent image correspondence, *Computer Vision and Image Understanding* (2018), doi: [10.1016/j.cviu.2018.02.005](https://doi.org/10.1016/j.cviu.2018.02.005)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Highlights**

- A unified, camera-independent theory for 3D reconstruction problems is proposed.
- A camera-independent, generalized epipolar geometry is presented.
- Epipolar constraints are derived from compatibility equation.
- The theory is applied to perspective, axial and general spherical cameras.



Computer Vision and Image Understanding  
journal homepage: [www.elsevier.com](http://www.elsevier.com)

## A differential geometry approach to camera-independent image correspondence

József Molnár<sup>a,\*\*</sup>, Iván Eichhardt<sup>b,c</sup>

<sup>a</sup>MTA BRC, Temesvári krt. 62, H-6726 Szeged, Hungary

<sup>b</sup>ELTE IK, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

<sup>c</sup>MTA SZTAKI, Kende u. 13-17, Budapest H-1111, Hungary

### ABSTRACT

Projective geometry is a standard mathematical tool for image-based 3D reconstruction. Most reconstruction methods establish pointwise image correspondences using projective geometry. We present an alternative approach based on differential geometry using oriented patches rather than points. Our approach assumes that the scene to be reconstructed is observed by any camera, existing or potential, that satisfies very general conditions, namely, the differentiability of the surface and the bijective projection functions. We show how the notions of the differential geometry such as diffeomorphism, pushforward and pullback are related to the reconstruction problem. A unified theory applicable to various 3D reconstruction problems is presented. Considering two views of the surface, we derive reconstruction equations for oriented patches and pose equations to determine the relative pose of the two cameras. Then we discuss the generalized epipolar geometry and derive the generalized epipolar constraint (compatibility equation) along the epipolar curves. Applying the proposed theory to the projective camera and assuming that affine mapping between small corresponding regions has been estimated, we obtain the minimal pose equation for the case when a fully calibrated camera is moved with its internal parameters unchanged. Equations for the projective epipolar constraints and the fundamental matrix are also derived. Finally, two important nonlinear camera types, the axial and the spherical, are examined.

© 2018 Elsevier Ltd. All rights reserved.

### 1. Introduction

Most approaches to multi-view stereo reconstruction (Furukawa and Ponce, 2010; Habbecke and Kobbelt, 2007; Seitz et al., 2006) perspective, affine or weak perspective camera models (Hartley and Zisserman, 2005). Solutions for central and non-central catadioptric cameras (Svoboda and Pajdla, 2002; Micusik and Pajdla, 2004) are also available. Despite the great variety of the approaches, almost all of them rely on *projective geometry* as a basic tool to describe relations between scene points and image points, or establish correspondence between points in different views.

In this section, we first discuss stereo reconstruction approaches based on projective geometry as the mainstream of the related research. Special attention is paid to the way correspondence is established for homography estimation. Then

we discuss possible alternatives to the mainstream which use *differential geometry*.

Most methods search for pointwise or region correspondences. The essential difference between region-based affine correspondence and point correspondence is discussed in (Bentolila and Francos, 2014a). Attempts to avoid correspondence, e.g. (Kutulakos and Seitz, 1999; Domokos et al., 2012), have also been made. Brightness and texture gradients reveal surface geometry; they can be used in shape from shading and shape from texture, respectively (Sonka et al., 2008). These methods operate on single images and do not require correspondences.

Affine-covariant regions and features (Mikolajczyk et al., 2005; Tuytelaars and Mikolajczyk, 2008) can be used to find image correspondences and estimate affine distortion of a surface patch between views (see also Oxford University, KU Leuven, INRIA, CMP (2007)). Alternatively, one can apply the correspondence-free approach (Domokos et al., 2012) to register shapes and estimate local homography.

In the framework of projective geometry, studies (Köser,

<sup>\*\*</sup>Corresponding author: Tel.: +36-30-231-0952;  
e-mail: [jmolnar64@digikabel.hu](mailto:jmolnar64@digikabel.hu) (József Molnár)

2009; Köser et al., 2008; Köser and Koch, 2008) investigate the following aspects of the affine approximation of local inter-image warp: a) general homography; b) infinite homography resulting from conjugate rotation in the perspective camera model; c) surface normal estimation and d) pose estimation for model-independent calibrated camera. The general homography is derived from two affine correspondences. For the homography of the conjugate rotation, Köser et al. present a minimal parameterization having seven DOFs. Their only constraint, a linear equation, is derived from the orthogonality of the rotation connecting the two components of the homography that cannot be determined from a single affine correspondence. The authors also derive the following constraints on the additional parameter of the general homography that has eight DOFs: (i) linear constraints using an extra point/line correspondence; or, alternatively, (ii) a quadratic constraint that restricts the internal calibration, i.e., the known aspect ratio and zero skew. In this paper, we also study the problems of surface normal and pose estimation and compare our approach with (Köser, 2009).

In the study (Rothganger et al., 2007), the authors consider affine-covariant patches and derive locally affine projection constraints by linearizing the perspective projection function in a vicinity of patch center. The constraints are used to find rigid components in a dynamic scene and build the 3D models of the components. Other authors (Perd'och et al., 2006; Riggi et al., 2006) apply local affine approximation to obtain additional corresponding points for a more robust solution.

Most of the current approaches for calculating the affine fundamental matrix use pointwise correspondences; some methods (Arandjelovic and Zisserman, 2010; Bentolila and Francos, 2014a,b) use affine region correspondences. The method (Arandjelovic and Zisserman, 2010) represents an affine covariant region by an ellipse posing the problem of affine region correspondence between two images as matching of two ellipses.

The limitations of the approach (Arandjelovic and Zisserman, 2010) are discussed in the study (Bentolila and Francos, 2014a) that formulates explicit constraints on the epipolar geometry resulting from affine correspondences treated as derivatives of the corresponding homographies. A requirement for a fundamental matrix to be compatible with a homography is formulated. Employing this compatibility requirement, a pair of affine correspondences is shown to constrain the location of the epipole to a conic. Given three correspondences, one can obtain the epipole as the intersection of two conics, then calculate the fundamental matrix.

In (Bentolila and Francos, 2014b), the same authors introduce a metric for measuring the distance between affine transformations and apply it to the estimation of homography and fundamental matrix based on affine region correspondences. In section 4, we discuss the relation of our approach to the results of (Bentolila and Francos, 2014a,b).

The mainstream research has led to the development of solutions providing impressive results in both sparse and dense reconstruction of scenes and objects with varying geometry and surface properties. Applications to vision-based

SLAM (Lemaire et al., 2007; Davison et al., 2007) have also resulted in significant improvement in localization and mapping by mobile devices, autonomous robots and vehicles.

Differential properties of surfaces expressed by image gradients and affine distortions of local regions have been used in various areas related to 3D reconstruction. In particular, affine propagation of patch correspondences in wide-baseline stereo was proposed in (Megyesi et al., 2006). The importance of the oriented patches for multiview stereo was recognized and utilized in (Furukawa and Ponce, 2010). The study (Habbecke and Kobbelt, 2007) uses surface growing in multi-view reconstruction by image warping estimating the surface normal vector as a linear function of the camera matrix and the homography.

In this paper, we consider a surface viewed by two cameras assuming that the Jacobian of the local mapping between the two views is known. We propose a comprehensive differential geometry framework for a wide class of camera models including the perspective one. In particular, we derive relationships between local distortions of small corresponding regions, the parameters of the cameras and the local geometry of the surface in the regions. This work can be viewed as a unifying and generalizing theoretical foundation for the partial theoretical and experimental results published by us and other authors in (Megyesi et al., 2006; Köser, 2009; Molnár et al., 2014b; Tanacs et al., 2014; Molnár et al., 2014a; Barath et al., 2015).

We address neither low-level data acquisition or establishing correspondence nor the problems related to the complexity of real world scenes. Coping with phenomena such as self-shading of non-convex objects or non-Lambertian reflectance are important problems in themselves. For interested readers we recommend the following studies (Magda et al., 2001; Belhumeur and Kriegman, 1998; Adato et al., 2010; Gkioulekas et al., 2015).

In spirit, our theory is related to the work (Devernay and Faugeras, 1994) that also relies on differential geometry. However, the study (Devernay and Faugeras, 1994) considers only the perspective camera model and uses a parameterization dependent, non-invariant representation, while we use a *very general camera model* and an invariant representation. Our camera model is a mapping restricted only by the differentiability of the surface and the bijective projection functions. Perspective, affine, weak-perspective and central and non-central catadioptric camera models are all special cases of our model.

The main contributions of this paper are as follows. For our general camera model, we obtain a) correspondence equations applicable to scene reconstruction; b) a pose equation that can be used to calculate the relative pose of the cameras; c) a generalized epipolar constraint along the epipolar curves and d) compatibility equations for local correspondences and general epipolar geometry. The proposed theory results in the minimal pose equation for the special case of the widely applied perspective camera model. This allows one to determine the new pose of a fully calibrated camera moved to another position with its internal parameters unchanged. In particular, we derive (i) the projective fundamental relation involving the fundamental matrix, as a specific solution of the general epipolar differential equation; (ii) the differential constraint for the fun-



damental matrix and (iii) the algebraic form of the epipolar constraint introduced in (Bentolila and Francos, 2014a). This form enables robust calculation of the epipoles using overdetermined system of equations. Finally, we examine the cases of the axial and the spherical cameras and derive the fundamental quantities and the coordinate gradients for both cases.

The structure of the paper is the following. Section 2 introduces notations and theoretical background. Then derivations for reconstruction, pose estimation, epipolar geometry and compatibility equations for a surface observed by a general camera are presented. In sections 3 and 4, we apply the general theory to the perspective camera. Section 5 studies two important nonlinear camera models. In section 6, we show and analyze test results for the following problems: (i) epipole calculation in order to determine the center of distortion; (ii) pose estimation and (iii) surface reconstruction. Section 7 concludes the paper by discussion and outlook.

## 2. Theory for surface viewed by general camera

### 2.1. Notations

The notations we use are widely used in classical differential geometry. For vectors and tensors, we use bold letters and italics for the coordinates. For spatial coordinates, we use italic capital letters with superscripts:  $X^1, X^2, X^3$ ; for 3D vectors, we use bold capital letters, while lowercase bold letters are used for 2D vectors. *Homogeneous representations* are marked with *tilde* to be distinguished from their inhomogeneous counterparts. Italic letters  $u^1, u^2$  are used for Gaussian point coordinates constrained to the embedded manifolds. Partial derivatives are denoted by subscripts. The world coordinate system given by standard basis in space is defined by three orthonormal basis vectors  $\mathbf{e}_1, \mathbf{e}_2$  and  $\mathbf{e}_3$ . 3D points  $\mathbf{X} \in \mathbb{R}^3$  are identified by their coordinates in the standard basis:  $\mathbf{X} = X^1\mathbf{e}_1 + X^2\mathbf{e}_2 + X^3\mathbf{e}_3$ . An embedded surface  $\mathbf{S} \subset \mathbb{R}^3$  is defined by a two-parameter vector-valued function:

$$\mathbf{S}(u^1, u^2) = X^1(u^1, u^2)\mathbf{e}_1 + X^2(u^1, u^2)\mathbf{e}_2 + X^3(u^1, u^2)\mathbf{e}_3. \quad (1)$$

The tangent space for a surface  $\mathbf{S}$  at a surface point  $(u^1, u^2)$  is spanned by the local (covariant) basis vectors  $\mathbf{S}_k = \frac{\partial \mathbf{S}}{\partial u^k}$ ,  $\mathbf{S}_k = \mathbf{S}_k(u^1, u^2)$ ,  $k = 1, 2$ . The corresponding contravariant basis vectors  $\mathbf{S}^l$ ,  $l = 1, 2$ , are defined to satisfy the identity equations  $\mathbf{S}^l \cdot \mathbf{S}_k = \delta_k^l$ , where  $\delta_k^l$  is the Kronecker delta, and the scalar product is denoted by dot.

The normal vector of the surface is given by  $\mathbf{N} = \mathbf{S}_1 \times \mathbf{S}_2$ , where the cross product is denoted by  $\times$ . Signed surface area element is defined by the triple scalar product  $|\mathbf{nS}_1\mathbf{S}_2| \doteq \mathbf{n} \cdot (\mathbf{S}_1 \times \mathbf{S}_2)$ , where  $\mathbf{n} = \frac{\mathbf{N}}{|\mathbf{N}|}$  is the unit normal vector of the surface. The cross-tensor of the normal vector  $\mathbf{N}_\times = \mathbf{S}_2\mathbf{S}_1 - \mathbf{S}_1\mathbf{S}_2$  is a difference of two dyadic (direct) products of the local basis vectors. A dyadic product is denoted by a simple sequence of the constituent vectors.

The dot product between dyads and vectors is defined so that  $\mathbf{uv} \cdot \mathbf{w} = (\mathbf{v} \cdot \mathbf{w})\mathbf{u}$ . Therefore,  $\mathbf{N}_\times \cdot \mathbf{v} = \mathbf{N} \times \mathbf{v}$  for any vector  $\mathbf{v}$ . For the representation of vectors and second order tensors purely with their coordinates, we use column vectors and two-dimensional matrices.

### 2.2. Camera-independent correspondence equations

Consider an observed scene in the 3D space  $\mathbb{R}^3$ . The visible parts of the scene objects are treated as 2D surfaces embedded in  $\mathbb{R}^3$  given by Eq. (1). Different images of a surface are distinguished with lower indices  $i, j$ ; only these two letters are used to identify the projection functions, any other letter in subscript means either partial derivative or coordinate.

We assume that images of spatial points are projections given by two functions assigning two image coordinates  $(x^1, x^2)$  to spatial points. Spatial points  $X^m$  lying on the surface  $X^m(u^1, u^2)$  are mapped onto the  $i$ -th image by composite functions of coordinates  $k = 1, 2$  as follows:

$$x_i^k = x_i^k(X^1(u^1, u^2), X^2(u^1, u^2), X^3(u^1, u^2)) = \hat{x}_i^k(u^1, u^2). \quad (2)$$

To simplify notation, the hat in the right-hand side will be omitted. We suppose that the mappings in Eq. (2) are bijections in a small open disk around the point  $(u^1, u^2)$ . Assuming that both the projection functions and the surface are smooth, this is the condition for differentiability. The inverse functions of the bijective mappings,  $u^1(x_i^1, x_i^2)$  and  $u^2(x_i^1, x_i^2)$ , also exist.

Consider a surface observed by two cameras that provide images  $i$  and  $j$ . A small shift on the surface results in small shifts  $\mathbf{dx}_i$  and  $\mathbf{dx}_j$  in the two images. As shown in (Molnár and Chetverikov, 2014), they are related as follows:

$$\mathbf{dx}_j = \mathbf{J}_j \cdot \mathbf{J}_i^{-1} \cdot \mathbf{dx}_i \doteq \mathbf{J}_{ij} \cdot \mathbf{dx}_i, \quad (3)$$

where the Jacobian of the image mapping  $i \rightarrow j$

$$\mathbf{J}_{ij} = \begin{bmatrix} \frac{\partial x_j^1}{\partial x_i^1} & \frac{\partial x_j^1}{\partial x_i^2} \\ \frac{\partial x_j^2}{\partial x_i^1} & \frac{\partial x_j^2}{\partial x_i^2} \end{bmatrix} = \begin{bmatrix} \frac{\partial x_j^1}{\partial u^1} & \frac{\partial x_j^1}{\partial u^2} \\ \frac{\partial x_j^2}{\partial u^1} & \frac{\partial x_j^2}{\partial u^2} \end{bmatrix} \begin{bmatrix} \frac{\partial u^1}{\partial x_i^1} & \frac{\partial u^1}{\partial x_i^2} \\ \frac{\partial u^2}{\partial x_i^1} & \frac{\partial u^2}{\partial x_i^2} \end{bmatrix}^{-1}. \quad (4)$$

The images are two-dimensional Euclidean manifolds (planes). Relations between regions of two images can be considered as a set of local **diffeomorphisms** whose differential is the Jacobian (4). These diffeomorphisms, however, have physical origin: they are induced by the scene objects with the help of light rays. We seek representation that reflects this physical origin.

The equation (4) is parameterized by  $(u^1, u^2)$ . The partial derivatives of any function  $f \in \{x_i^1, x_j^1, x_i^2, x_j^2\}$  can be written as

$$\frac{\partial f}{\partial u^k} = \frac{\partial X^1}{\partial u^k} \frac{\partial f}{\partial X^1} + \frac{\partial X^2}{\partial u^k} \frac{\partial f}{\partial X^2} + \frac{\partial X^3}{\partial u^k} \frac{\partial f}{\partial X^3} = \mathbf{S}_k \cdot \nabla f, \quad (5)$$

where  $k = 1, 2$  and  $\mathbf{S}_k$  are the partial derivatives of the surface (1),  $\nabla f$  is the spatial gradient of  $f$ . After applying this result to the projection functions, the components of the Jacobians  $\mathbf{J}_i, \mathbf{J}_j$  take the following form:

$$\mathbf{J}_m = \begin{bmatrix} \mathbf{S}_1 \cdot \nabla x_m^1 & \mathbf{S}_2 \cdot \nabla x_m^1 \\ \mathbf{S}_1 \cdot \nabla x_m^2 & \mathbf{S}_2 \cdot \nabla x_m^2 \end{bmatrix}, \quad m = i, j. \quad (6)$$

Substitute Eq. (6) into Eq. (3). Then the products of the components of Eq. (6) enter  $\mathbf{J}_{ij}$ . For example, the determinant of

$\mathbf{J}_i$  expressed by the dyadic products is equivalent to the surface normal cross-tensor:

$$\begin{aligned}\det \mathbf{J}_i &= \nabla x_i^1 \cdot (\mathbf{S}_1 \mathbf{S}_2 - \mathbf{S}_2 \mathbf{S}_1) \cdot \nabla x_i^2 \\ &= -\nabla x_i^1 \cdot \mathbf{N}_\times \cdot \nabla x_i^2 \\ &= -|\mathbf{N}| |\nabla x_i^1 \nabla x_i^2|. \end{aligned} \quad (7)$$

The Jacobian becomes

$$\mathbf{J}_{ij} = \frac{1}{|\nabla x_i^1 \nabla x_i^2|} \begin{bmatrix} |\nabla x_j^1 \nabla x_i^2| & |\nabla x_j^1 \nabla x_i^1| \\ |\nabla x_j^2 \nabla x_i^2| & |\nabla x_j^2 \nabla x_i^1| \end{bmatrix}, \quad (8)$$

where  $|\nabla x_i^1 \nabla x_i^2|$  is the triple scalar product of the gradients and the normal unit vector  $\mathbf{n}$  of the surface. In the equation, the gradients represent the paths of light, while the normal vector represents the surface. These quantities are invariant first-order differentials. Eq. (8) is a general formula that can be applied to any camera type and any reasonably smooth surface, since neither specific projection function nor specific surface is assumed.

### 2.3. Alternative interpretation

Using the Helmholtz reciprocity principle (Zickler et al., 2002), we can think of reversing the directions of the light paths: from images to surfaces. This view leads to an alternative interpretation of image correspondence. Suppose the observed surface is parameterized by its local image coordinates **pushed forward** to the surface creating its local map. For example, image  $i$  induces the following parameterization:

$$\mathbf{S}(x_i^1, x_i^2) = X^1(x_i^1, x_i^2)\mathbf{e}_1 + X^2(x_i^1, x_i^2)\mathbf{e}_2 + X^3(x_i^1, x_i^2)\mathbf{e}_3. \quad (9)$$

We need the local basis  $\mathbf{S}_{1i} = \frac{\partial \mathbf{S}}{\partial x_i^1}$ ,  $\mathbf{S}_{2i} = \frac{\partial \mathbf{S}}{\partial x_i^2}$  expressed with invariants. Applying Eq. (5) to the coordinate functions  $x_i^1$  and  $x_i^2$  with  $u^1 = x_i^1$  and  $u^2 = x_i^2$ , we obtain

$$\mathbf{S}_p \cdot \nabla q = \delta_{pq}, \quad p, q \in \{x_i^1, x_i^2\}, \quad (10)$$

where  $\delta_{pq}$  is the Kronecker delta. This fulfills the definition of the inverse basis for  $\nabla x_i^1, \nabla x_i^2$ . The inverse (contravariant) basis vectors will be denoted by  $\mathbf{S}_i^1, \mathbf{S}_i^2$ . Since they lie on the tangent plane of the surface, the following must hold:

$$\begin{aligned}\mathbf{S}_i^1 &= \nabla x_i^1|_T, \quad \mathbf{S}_i^2 = \nabla x_i^2|_T \\ \nabla z|_T &= \nabla z \cdot (\mathbf{I} - \mathbf{nn}), \quad z \in \{x_i^1, x_i^2\}. \end{aligned} \quad (11)$$

Here  $\nabla z|_T$  is the projection of  $\nabla z$  to the tangent plane with the projector  $\mathbf{I} - \mathbf{nn}$ ,  $\mathbf{I}$  the identity tensor,  $\mathbf{nn}$  the direct (dyadic) product. The cross-product of these contravariant vectors is perpendicular to the tangent plane, hence it is a surface normal with the length  $l_i = \mathbf{n} \cdot (\mathbf{S}_i^1 \times \mathbf{S}_i^2)$ . Using Eq. (11), we have

$$\begin{aligned}l_i &= \mathbf{n} \cdot [(\nabla x_i^1 - (\nabla x_i^1 \cdot \mathbf{n})\mathbf{n}) \times (\nabla x_i^2 - (\nabla x_i^2 \cdot \mathbf{n})\mathbf{n})] \\ &= |\nabla x_i^1 \nabla x_i^2|. \end{aligned} \quad (12)$$

We observe that  $l_i$  equals the denominator in the Jacobian (8). Since the contravariant and covariant basis vectors are related

as  $\mathbf{S}_{1i} = \frac{1}{l_i}(\mathbf{S}_i^2 \times \mathbf{n})$ ,  $\mathbf{S}_{2i} = \frac{1}{l_i}(\mathbf{n} \times \mathbf{S}_i^1)$ , we have

$$\begin{aligned}\mathbf{S}_{1i} &= \frac{1}{|\nabla x_i^2 \nabla x_i^1|} [\nabla x_i^2 - (\nabla x_i^2 \cdot \mathbf{n})\mathbf{n}] \times \mathbf{n} = \frac{\mathbf{n} \times \nabla x_i^2}{|\nabla x_i^1 \nabla x_i^2|}, \\ \mathbf{S}_{2i} &= \frac{1}{|\nabla x_i^1 \nabla x_i^2|} \mathbf{n} \times [\nabla x_i^1 - (\nabla x_i^1 \cdot \mathbf{n})\mathbf{n}] = \frac{\nabla x_i^1 \times \mathbf{n}}{|\nabla x_i^1 \nabla x_i^2|}. \end{aligned} \quad (13)$$

Any vector  $\mathbf{v}$  in the tangential plane can be decomposed in two ways:

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{S}_1)\mathbf{S}_1 + (\mathbf{v} \cdot \mathbf{S}_2)\mathbf{S}_2 = (\mathbf{v} \cdot \mathbf{S}_1)\mathbf{S}_1^1 + (\mathbf{v} \cdot \mathbf{S}_2)\mathbf{S}_2^2, \quad (14)$$

where  $v^1 = \mathbf{v} \cdot \mathbf{S}_1$ ,  $v^2 = \mathbf{v} \cdot \mathbf{S}_2$  are the contravariant,  $v_1 = \mathbf{v} \cdot \mathbf{S}_1$ ,  $v_2 = \mathbf{v} \cdot \mathbf{S}_2$  the covariant vector coordinates. Applying such decomposition to Eq. (3), the components of  $d\mathbf{x}_i = \mathbf{S}_{1i}dx_i^1 + \mathbf{S}_{2i}dx_i^2$  in projection  $j$  can be expressed as

$$dx_j^k = \mathbf{S}_j^k \cdot (\mathbf{S}_{1i}dx_i^1 + \mathbf{S}_{2i}dx_i^2), \quad k = 1, 2. \quad (15)$$

Using (11) and (13), the Jacobian (4) can be written as

$$\mathbf{J}_{ij} = \begin{bmatrix} \nabla x_j^1|_T \cdot \frac{(\mathbf{n} \times \nabla x_i^2)}{|\nabla x_i^1 \nabla x_i^2|} & \nabla x_j^1|_T \cdot \frac{(\nabla x_i^1 \times \mathbf{n})}{|\nabla x_i^1 \nabla x_i^2|} \\ \nabla x_j^2|_T \cdot \frac{(\mathbf{n} \times \nabla x_i^2)}{|\nabla x_i^1 \nabla x_i^2|} & \nabla x_j^2|_T \cdot \frac{(\nabla x_i^1 \times \mathbf{n})}{|\nabla x_i^1 \nabla x_i^2|} \end{bmatrix} \doteq \begin{bmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{bmatrix}. \quad (16)$$

This form, which is equivalent to Eq. (8), expresses the image mapping  $i \rightarrow j$  by invariant first-order differential quantities, the projection gradients and the unit normal vector. The symbols  $a_1^1, a_2^1, \dots$  are introduced to simplify notation. The components  $a_j^k$  of  $\mathbf{J}_{ij}$  can be estimated from image correspondences. Once this has been done, their equivalence with the invariant expressions (8) or (16) can be used for different purposes.

Applying the decomposition Eq. (14) to the tangential vectors  $\nabla x_j^1|_T, \nabla x_j^2|_T$ , for components  $k = 1, 2$  we obtain

$$\nabla x_j^k|_T = (\nabla x_j^k \cdot \mathbf{S}_{1i})\nabla x_i^1|_T + (\nabla x_j^k \cdot \mathbf{S}_{2i})\nabla x_i^2|_T, \quad (17)$$

The expressions in parentheses are the components of  $\mathbf{J}_{ij}$ , hence Eq. (17) can be rewritten as

$$\begin{bmatrix} \nabla x_j^1|_T \\ \nabla x_j^2|_T \end{bmatrix} = \mathbf{J}_{ij} \cdot \begin{bmatrix} \nabla x_i^1|_T \\ \nabla x_i^2|_T \end{bmatrix}, \quad (18)$$

which means that the contravariant basis vectors transform as coordinate differentials. We call this important relation the **pose equation** for the reason that will be explained later. The pose equation states that the same relationship exists between two images of a surface as between projection gradients constrained to the tangent plane.

Using Eq. (11), Eq. (18) can be re-written as

$$\nabla x_j^k \cdot (\mathbf{I} - \mathbf{nn}) = a_1^k \nabla x_i^1 \cdot (\mathbf{I} - \mathbf{nn}) + a_2^k \nabla x_i^2 \cdot (\mathbf{I} - \mathbf{nn}). \quad (19)$$

Taking the dot product of both sides with  $(\nabla x_i^1 \times \nabla x_i^2)$  and using the identities  $\mathbf{a} \cdot \mathbf{I} = \mathbf{a}$  and  $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0$ ,  $\mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0$  for arbitrary vectors  $\mathbf{a}, \mathbf{b}$ , we derive the following scalar equation system:

$$\frac{1}{l_i} \begin{bmatrix} |\nabla x_i^1 \nabla x_i^1 \nabla x_i^2| \\ |\nabla x_i^2 \nabla x_i^1 \nabla x_i^2| \end{bmatrix} = \begin{bmatrix} \nabla x_j^1|_T \\ \nabla x_j^2|_T \end{bmatrix} \cdot \begin{bmatrix} \nabla x_i^1|_T \\ \nabla x_i^2|_T \end{bmatrix}. \quad (20)$$

Here, the right-hand side is the counterpart of Eq. (18) in the normal direction. Recall that  $l_i$  was introduced in Eq. (12), while  $\nabla z|_{\mathbf{n}} = (\nabla z \cdot \mathbf{n})\mathbf{n}$  is the projection of  $\nabla z$ ,  $z \in \{x_i^1, x_i^2, x_j^1, x_j^2\}$ , to the normal direction. The left-hand side is the basic expression for the **epipolar geometry** discussed in Section 2.4.

It is worth mentioning that in some special cases, e.g., for a calibrated depth camera, it is possible to **pull back** the metric of the observed surface patch using a single image and the depth information. (The latter is necessary to calculate the normal vectors.) The simplest way to do this is to retrieve the inverse metric components

$$g^{kl} = \mathbf{S}^k \cdot \mathbf{S}^l = (\nabla x^k \times \mathbf{n}) \cdot (\nabla x^l \times \mathbf{n}), \quad (21)$$

where  $k, l = 1, 2$ , then invert the matrix:

$$\begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} = \begin{bmatrix} g^{11} & g^{12} \\ g^{12} & g^{22} \end{bmatrix}^{-1}. \quad (22)$$

Having the metric components  $g_{kl} = g_{kl}(x^1, x^2)$  as functions of the image coordinates, one can measure on the surface lengths, angles, areas and other properties working in the image domain alone.

#### 2.4. Epipolar geometry

Now we impose further restrictions on the projection functions (2). We assume that each image point has a ray associated with the point. The rays may not intersect, that is, points in space may not have same image coordinates, except for the case when they have a common projection center. We emphasize that this does not necessarily mean central projection, since each image point may have its own origin denoted by  $\mathbf{C} = \mathbf{C}(x^1, x^2)$ . We only assume that origins and rays vary smoothly keeping all differentiability criteria valid.

The ray  $\mathbf{X}(t)$ ,  $t \in (0, \infty]$ ,  $\mathbf{X}(0) = \mathbf{C}$  is specified by constant coordinates  $x^1(\mathbf{X}(t)) = (x^1)_0$ ,  $x^2(\mathbf{X}(t)) = (x^2)_0$  for any ray parameter  $t$ . The derivative w.r.t.  $t$  is  $\nabla x^k \cdot \dot{\mathbf{X}} = 0$ , where  $\dot{\mathbf{X}}(t) = \frac{d\mathbf{X}}{dt}$  is the direction of the ray. That is,  $\dot{\mathbf{X}}(t)$  is perpendicular to both gradients, hence

$$\dot{\mathbf{X}}(t) = c(\nabla x^1 \times \nabla x^2) \quad (23)$$

for any real constant  $c$  which can be selected freely. Since the ray direction  $\frac{\dot{\mathbf{X}}(t)}{|\dot{\mathbf{X}}(t)|}$  is independent of  $t$ , the unit vector  $\frac{\nabla x^1 \times \nabla x^2}{|\nabla x^1 \times \nabla x^2|}$  depends only on the image coordinates  $(x^1)_0, (x^2)_0$ . Integrating this normalized version of Eq. (23), we obtain the **equation for back-projected ray**:

$$\mathbf{X}(t) = \mathbf{C} + \frac{\nabla x^1 \times \nabla x^2}{|\nabla x^1 \times \nabla x^2|} t = \mathbf{C} + \frac{\nabla x^1 \times \nabla x^2}{r} t, \quad (24)$$

$$r \doteq |\nabla x^1 \times \nabla x^2|,$$

where the constant vector  $\mathbf{C} = \mathbf{X}(0)$  is the origin of the ray, the ‘projection center’ associated with the coordinates  $(x^1)_0, (x^2)_0$ .

Observing by camera  $j$  a back-projected ray of camera  $i$ , we have the following correspondence equation:

$$x_j^k(t) = x_j^k(\mathbf{C}_i) + \frac{1}{r_i} (\nabla x_i^1 \times \nabla x_i^2) t, \quad k = 1, 2. \quad (25)$$

Since the normalized cross product  $\frac{1}{r_i} (\nabla x_i^1 \times \nabla x_i^2)$  is independent of  $t$ ,

$$\frac{dx_j^k}{dt} = \nabla x_j^k \cdot \frac{\nabla x_i^1 \times \nabla x_i^2}{r_i}. \quad (26)$$

From this, we obtain the first-order ordinary differential equation

$$\frac{dx_j^2}{dx_j^1} = \frac{|\nabla x_j^2 \nabla x_i^1 \nabla x_i^2|}{|\nabla x_j^1 \nabla x_i^1 \nabla x_i^2|} \quad (27)$$

expressed as a ratio of triple scalar products that contains neither  $t$  nor  $r_i$ . The initial condition is given by the ‘epipoles’  $x_j^2(x_j^1(\mathbf{C}_i)) = x_j^2(\mathbf{C}_i)$ , and solution associating possible image coordinate pairs  $(x_j^1, x_j^2(x_j^1))$  to the image point  $(x_i^1, x_i^2)$  is uniquely defined.

According to Eq. (20), the differential equation compatible with Eq. (8) can be expressed via image gradients and the entries of  $\mathbf{J}_{ij}$ :

$$\frac{dx_j^2}{dx_j^1} = \frac{\mathbf{n} \cdot (\nabla x_j^2 - a_1^2 \nabla x_i^1 - a_2^2 \nabla x_i^2)}{\mathbf{n} \cdot (\nabla x_j^1 - a_1^1 \nabla x_i^1 - a_2^1 \nabla x_i^2)}. \quad (28)$$

Eq. (28) can be considered as the **compatibility equation**, that is, the correspondence equation compatible with the epipolar geometry. It provides equations for the components of  $\mathbf{J}_{ij}$ , i.e., the components of  $\mathbf{J}_{ij}$  are not independent along the epipolar curves. Examples will be given in section 3.

In the case of central projection with constant  $\mathbf{C}_i$  and  $\mathbf{C}_j$ , the vector  $(\mathbf{C}_i - \mathbf{C}_j)$  and the two rays  $(\nabla x_i^1 \times \nabla x_i^2), (\nabla x_j^1 \times \nabla x_j^2)$  define the epipolar plane. Its images are the above mentioned epipolar curves. With an epipolar plane given, the two associated epipolar curves are defined by

$$\frac{dx_i^2}{dx_i^1} = \frac{|\nabla x_i^2 \nabla x_j^1 \nabla x_j^2|}{|\nabla x_i^1 \nabla x_j^1 \nabla x_j^2|}, \quad x_i^2(x_i^1(\mathbf{C}_j)) = x_i^2(\mathbf{C}_j), \quad (29)$$

and similarly for  $j$ , with  $i$  and  $j$  swapped. Any observed object point on an epipolar plane has two projected points on its associated epipolar curves. Searching a point along the corresponding epipolar curves means searching an object point on the epipolar plane.

### 3. Application to the projective camera

As long as the differentiability criteria are valid, the presented theory does not assume any particular camera model. Below, we apply the theory to the finite projective CCD camera because of its practical importance. *Main results of this section are:* i) the normal vector and triangulation equations (38) and (40) for reconstruction; ii) the minimal pose equations (44) and iii) the derivation of the fundamental matrix from the most general differential equation (27) of the epipolar geometry.

In the case of perspective views, the projection functions are given by the projection matrix in the form of  $\mathbf{P} = \mathbf{K} \cdot [\mathbf{R}, \mathbf{t}]$ , where  $\mathbf{K}$  is an upper-triangular matrix,  $\mathbf{R}$  the rotation matrix,

$\mathbf{t}$  the translation vector. In homogeneous coordinates, a spatial point  $\mathbf{X}$  is projected onto image point  $\mathbf{x}$  as

$$\tilde{\mathbf{x}} = \mathbf{P} \cdot \tilde{\mathbf{X}}, \quad (30)$$

where  $\tilde{\mathbf{X}} = [X^1 \ X^2 \ X^3 \ 1]^T$  and  $\tilde{\mathbf{x}} = s[x^1 \ x^2 \ 1]^T$  with unknown scale factor  $s$ . In practice, the skew-free (CCD) camera model is widely used. In this case  $\mathbf{K}$  and  $\mathbf{K}^{-1}$  take the simple form

$$\mathbf{K} = \begin{bmatrix} \alpha & 0 & u^1 \\ 0 & \beta & u^2 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}^{-1} = \begin{bmatrix} \frac{1}{\alpha} & 0 & -\frac{u^1}{\alpha} \\ 0 & \frac{1}{\beta} & -\frac{u^2}{\beta} \\ 0 & 0 & 1 \end{bmatrix}. \quad (31)$$

Introduce  $\boldsymbol{\rho}^k = [r_1^k \ r_2^k \ r_3^k]$  for the  $k$ -th row of the rotation matrix. Then the projection function becomes

$$\begin{aligned} x^1 &= \frac{1}{s} [(\alpha \rho^1 + u^1 \rho^3) \cdot \mathbf{X} + p_4^1], \\ x^2 &= \frac{1}{s} [(\beta \rho^2 + u^2 \rho^3) \cdot \mathbf{X} + p_4^2], \\ s &= \rho^3 \cdot \mathbf{X} + p_4^3 \end{aligned} \quad (32)$$

with  $\mathbf{X} = [X^1 \ X^2 \ X^3]^T$  and  $\mathbf{K} \cdot \mathbf{t} = [p_4^1 \ p_4^2 \ p_4^3]^T$ , the fourth column of  $\mathbf{P}$ . The gradient components  $\nabla x^k = [\frac{\partial x^k}{\partial X^1} \ \frac{\partial x^k}{\partial X^2} \ \frac{\partial x^k}{\partial X^3}]^T$  are

$$\begin{aligned} \frac{\partial x^1}{\partial X^l} &= \frac{1}{s} [\alpha r_l^1 - (x^1 - u^1) r_l^3], \\ \frac{\partial x^2}{\partial X^l} &= \frac{1}{s} [\beta r_l^2 - (x^2 - u^2) r_l^3], \quad l = 1, 2, 3. \end{aligned} \quad (33)$$

The following problems can be addressed using the proposed theory: 1. **Reprojection.** For a calibrated camera system and an approximately reconstructed surface, transformation between images can be estimated to evaluate similarity and refine the surface. This problem is considered in (Molnár and Chetverikov, 2014). 2. **Reconstruction.** For a calibrated camera system and estimated Jacobian (16), the surface normal vector and the relative distance to the tangent plane can be computed enabling reconstruction from sparse correspondences. The Jacobian is the local affine transformation with the two origins aligned that can be estimated by different means (Mikolajczyk et al., 2005; Tuytelaars and Mikolajczyk, 2008; Domokos et al., 2012). 3. **Pose estimation.** For a fully calibrated camera and a second camera with only internal parameters known, the pose of the second camera can be calculated given the Jacobian. Below, we address problems 2 and 3, which are inverse problems, assuming that the Jacobian components  $a_1^1, a_2^1, \dots$  have been estimated from images.

Later on, specifically for perspective camera, we present further applications of the theory: we derive the fundamental matrix and epipolar compatibility constraints for the components of the Jacobian (8).

### 3.1. Reconstruction

The process of reconstruction involves normal vector calculation followed by triangulation. For **normal vector calculation**, consider a calibrated camera pair. One can estimate the

components of the Jacobian  $a_j^k$  from region correspondences. Then Eq. (8) can be used to calculate the unknown unit normal vector. To eliminate the common denominator, one can use row, column, or cross ratios. Without loss of generality, we deduce the equation for the 3D surface normal using the cross ratios  $\frac{a_1^1}{a_2^1}$  and  $\frac{a_2^1}{a_1^1}$  as

$$\begin{aligned} \frac{\mathbf{n} \cdot (\nabla x_i^2 \times \nabla x_j^1)}{\mathbf{n} \cdot (\nabla x_j^2 \times \nabla x_i^1)} &= \frac{a_1^1}{a_2^1} \\ \frac{\mathbf{n} \cdot (\nabla x_j^1 \times \nabla x_i^1)}{\mathbf{n} \cdot (\nabla x_i^2 \times \nabla x_j^2)} &= \frac{a_2^1}{a_1^1} \end{aligned} \quad (34)$$

Rearranging, we obtain

$$\begin{aligned} \mathbf{n} \cdot [a_2^2(\nabla x_i^2 \times \nabla x_j^1) - a_1^1(\nabla x_j^2 \times \nabla x_i^1)] &= 0, \\ \mathbf{n} \cdot [a_1^2(\nabla x_j^1 \times \nabla x_i^1) - a_2^1(\nabla x_i^2 \times \nabla x_j^2)] &= 0, \end{aligned} \quad (35)$$

where we have two known vectors, both perpendicular to the normal:

$$\begin{aligned} \mathbf{v} &= a_2^2(\nabla x_i^2 \times \nabla x_j^1) - a_1^1(\nabla x_j^2 \times \nabla x_i^1), \\ \mathbf{w} &= a_1^2(\nabla x_j^1 \times \nabla x_i^1) - a_2^1(\nabla x_i^2 \times \nabla x_j^2). \end{aligned} \quad (36)$$

The surface unit normal can be readily computed as

$$\mathbf{n} = \frac{\mathbf{v} \times \mathbf{w}}{|\mathbf{v} \times \mathbf{w}|}. \quad (37)$$

Applying this to the projective camera with the projection function gradients (33) and the scaled gradients  $s_i \nabla x_i^k, s_j \nabla x_j^k$ ,  $k = 1, 2$ , the scaled vectors  $\mathbf{V} = s_i s_j \mathbf{v}$  and  $\mathbf{W} = s_i s_j \mathbf{w}$  yield the following result:

$$\mathbf{n} = \frac{\mathbf{V} \times \mathbf{W}}{|\mathbf{V} \times \mathbf{W}|}. \quad (38)$$

In contrast to this rather geometric approach, in his PhD study (Köser (2009), pp. 107–111) the author presented a purely linear algebraic approach to determine the surface normal from two views of a calibrated camera pair. His approach uses the Jacobian of the homography equation induced by the observed locally planar surface patch as  $\tilde{\mathbf{H}}_{norm} = \mathbf{H}_\pi \cdot \tilde{\mathbf{H}}_{inorm}$ . The normalized image coordinates are calculated as  $\tilde{\mathbf{x}}_{norm} = \mathbf{K}^{-1} \tilde{\mathbf{x}}$ , hence only the external camera parameters enter the equations, and  $\mathbf{H}_\pi$  contains the relative pose of the camera pair and the observed surface normal. (See Molton et al. (2004).) The resulting system of linear equations is overdetermined, and it can be solved for the two independent components of the surface normal using a least-squares method.

Now, we can apply **triangulation** to complete reconstruction. The ratio of the scale factors  $\frac{s_j}{s_i}$ , which is equal to ratio of the depths, is given by any component of (8). This can be used to calculate the spatial position of the observed patch by determining the minimal distance between the back-projected rays pointing to the patch. Using the notations of Fig. 1 and Eq. (24) and introducing

$$\mathbf{w}_m = \frac{\nabla x_m^1 \times \nabla x_m^2}{r_m}, \quad m = i, j, \quad (39)$$

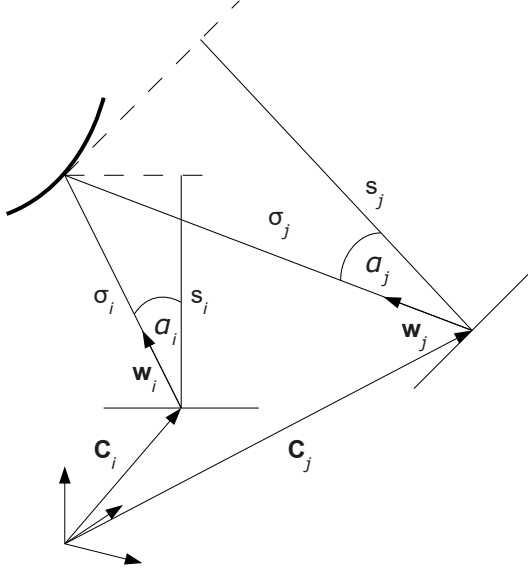


Fig. 1. The notations of triangulation.  $s_i, s_j$  are the depths and  $\sigma_i, \sigma_j$  are the Euclidean distances of the observed point relative to the camera centers  $C_i, C_j$ .  $w_i, w_j$  are used to denote the unit directional vectors of the observed point from camera  $i$  and  $j$  respectively.

the difference of the back-projected rays can be expressed as

$$\Delta \mathbf{X} \doteq \mathbf{X}_j - \mathbf{X}_i = \mathbf{C}_j - \mathbf{C}_i + \left( \frac{\sigma_j}{\sigma_i} \mathbf{w}_j - \mathbf{w}_i \right) t \quad (40)$$

with the gradients calculated at the image coordinates of the observed patch. Minimizing  $(\Delta \mathbf{X})^2$  is a one-parameter minimization problem that provides a closed-form solution for  $t = t_{obj}$  and the spatial position  $\mathbf{X}_{obj} = \mathbf{C}_i + \mathbf{w}_i t_{obj}$ . According to Fig. 1, the ratio of the distances

$$\frac{\sigma_j}{\sigma_i} = \frac{s_j \cos \alpha_i}{s_i \cos \alpha_j} \quad (41)$$

The cosine functions are calculated as the scalar products of two unit vectors: the directions of the observed patch  $\mathbf{w}_m$  and the directions of the cameras. The latter ones can be calculated in different ways. If the camera poses, hence their rotations, are known due to pose estimation, then the camera directions are  $\mathbf{R}_m \cdot \mathbf{e}_3$ . For a calibrated setup, the camera directions can be calculated using the same formula, Eq. (39), with the gradients evaluated at the principal points.

The normal equations (35) are not valid when the constituent image gradients are parallel vectors. For cameras with identical calibrations, this occurs when the camera centers and the observed reference point – the point in which the Jacobian is given – are collinear. This is the case of parallax-free motion well-known from projective geometry (Hartley and Zisserman, 2005). Another degenerate case occurs when the gradient vectors of camera  $i$  and the normal vector are coplanar, hence the common denominator in Eq. (8) is zero. This happens when camera  $i$  observes an apparent contour point (outline point). Observing an outline point by camera  $j$  does not lead to a degenerate case.

### 3.2. Pose estimation

Assume a camera had been calibrated, then moved with the internal parameters unchanged. Without loss of generality, we can suppose that camera  $i$  has been calibrated to the origin of the tangent plane  $\mathbf{n} = \mathbf{e}_3$  ( $Z = 0$ ). Then the pose equation (18) becomes

$$\nabla x_j^k|_T = a_1^k \nabla x_i^1|_T + a_2^k \nabla x_i^2|_T, \quad k = 1, 2. \quad (42)$$

Eq. (42) has an intrinsic ambiguity since flipped views of the surface lead to the same equation. If this ambiguity has been resolved in the context of a scene containing several objects, no additional ambiguities may appear.

The right-hand side of Eq. (42) has known entries: the parameters of the completely calibrated camera and the estimated Jacobian components. The left-hand side has 7 unknowns: 6 components of the rotation matrix and  $s$ . The number of equations available is also 7: 4 independent equations (42) written for the tangential ( $k = 1, 2$ ) components of (33) and the constraints on the rotation matrices, i.e., the norms of the columns are 1 and their dot product is zero. Equations (42) can therefore be considered as **minimal pose equations**.

Since all unknowns are in camera  $j$ , in the equations below we omit this index. Introduce  $\mathbf{r}_k = [r_k^1 \ r_k^2 \ r_k^3]^T$ ,  $k = 1, 2, 3$ , for the  $k$ -th column of  $\mathbf{R}$  in the decomposition  $\mathbf{P} = \mathbf{K} \cdot [\mathbf{R}, \mathbf{t}]$  of camera  $j$ . The right-hand side of Eq. (42) can be given in the standard basis. Denote these known components by  $A_l^k$ ,  $k, l = 1, 2$ :

$$a_1^k \nabla x_i^1|_T + a_2^k \nabla x_i^2|_T \doteq A_1^k \mathbf{e}_1 + A_2^k \mathbf{e}_2. \quad (43)$$

Using properties of  $\mathbf{R}$  and Eqs. (31) and (43), one can derive

$$\begin{aligned} (B_1^1 s + C_1^1 r_1^3)^2 + (B_2^1 s + C_2^1 r_1^3)^2 + (r_1^3)^2 &= 1, \\ (B_2^1 s + C_2^1 r_1^3)^2 + (B_2^2 s + C_2^2 r_2^3)^2 + (r_2^3)^2 &= 1, \\ (B_1^1 s + C_1^1 r_1^3)(B_2^1 s + C_2^1 r_2^3) &+ (B_2^1 s + C_2^1 r_1^3)(B_2^2 s + C_2^2 r_2^3) + r_1^3 r_2^3 = 0. \end{aligned} \quad (44)$$

Here we introduced the notations  $B_k^1 \doteq \frac{1}{\alpha} A_k^1$ ,  $B_k^2 \doteq \frac{1}{\beta} A_k^2$ ,  $k = 1, 2$ ,  $C^1 \doteq \frac{1}{\alpha}(x^1 - u^1)$ ,  $C^2 \doteq \frac{1}{\alpha}(x^2 - u^2)$ . Recall that  $r_k^i$  is the element of  $\mathbf{R}$  in  $i$ -th row and  $k$ -th column.

The first two equations in (44) can be parametrically solved for  $r_1^3$  and  $r_2^3$  as functions of  $s$ , then the absolute value of the left-hand side in the third equation can be used as error function for  $s$ . Iteration with fixed number of steps can be applied. The maximum value for  $s$  is estimated as the lower bound of the discriminants of the first two equations (44). Finally, 4 solutions are available for positive  $s$ , from which the unique solution can be chosen by reprojection. Note that if the pose is completely or partially known, Eq. (42) can be solved for the internal calibration parameters (Eichhardt and Hajder, 2016).

A different approach to pose estimation was presented in Köser and Koch (2008). (See also Köser (2009), pp. 113–119, for further details.) This approach based on the known geometry of a reference scene object, e.g., using orthophoto, rather than its image acquired by a calibrated reference camera. The further assumptions in (Köser and Koch, 2008; Köser,

2009) are similar to ours: the reference scene object defines the world coordinate system where the pose equations are given. The equations are derived using the Jacobian of the perspective between the scene object and its image geometry. Similarly to our pose estimation equations, the Jacobian and the rotation matrix properties contribute four and two equations, respectively, for the six independent pose parameters. However, the solutions of the equation system in (Köser and Koch, 2008; Köser, 2009) and our approach are different. In the approach (Köser and Koch, 2008; Köser, 2009), the equations can be transformed into quadratic form geometrically interpreted as the intersection of conics and lines, hence an analytical solution is also possible.

### 3.3. Epipolar lines and fundamental matrix

The first-order ordinary differential equation (27) is valid for a wide class of camera models. In this section, we solve this equation for the case of the projective camera model. We show that the solution leads to the fundamental equation containing the fundamental matrix.

For the perspective camera, the gradients are

$$\begin{aligned} s \nabla x^l &= \mathbf{p}^l - x^l \mathbf{p}^3, \quad l = 1, 2, \\ s &= \mathbf{p}^3 \cdot \mathbf{X} + p_4^3, \end{aligned} \quad (45)$$

where  $(\mathbf{p}^T)^k = [p_1^k \ p_2^k \ p_3^k]$ ,  $k = 1, 2, 3$ , is the  $k$ -th row of the left  $3 \times 3$  submatrix of  $\mathbf{P}$ . In Eq. (27), the scale factors  $s_i, s_j$  are eliminated:

$$\begin{aligned} \frac{dx_j^2}{dx_j^1} &= \frac{|\nabla x_j^2 \nabla x_i^1 \nabla x_i^2|}{|\nabla x_j^1 \nabla x_i^1 \nabla x_i^2|} = \\ &= \frac{(\mathbf{p}_j^2 - x_j^2 \mathbf{p}_j^3) \cdot [(\mathbf{p}_i^1 - x_i^1 \mathbf{p}_i^3) \times (\mathbf{p}_i^2 - x_i^2 \mathbf{p}_i^3)]}{(\mathbf{p}_j^1 - x_j^1 \mathbf{p}_j^3) \cdot [(\mathbf{p}_i^1 - x_i^1 \mathbf{p}_i^3) \times (\mathbf{p}_i^2 - x_i^2 \mathbf{p}_i^3)]}. \end{aligned} \quad (46)$$

This can be re-arranged as

$$\frac{x_j^2 - \frac{(x_i^1 D_{23}^2 - x_i^2 D_{13}^2 + D_{12}^2)}{(x_i^1 D_{23}^3 - x_i^2 D_{13}^3 + D_{12}^3)}}{x_j^1 - \frac{(x_i^1 D_{23}^1 - x_i^2 D_{13}^1 + D_{12}^1)}{(x_i^1 D_{23}^3 - x_i^2 D_{13}^3 + D_{12}^3)}} = \frac{x_j^2 - \xi^2}{x_j^1 - \xi^1}, \quad (47)$$

where

$$\xi^k \doteq \frac{(x_i^1 D_{23}^k - x_i^2 D_{13}^k + D_{12}^k)}{(x_i^1 D_{23}^3 - x_i^2 D_{13}^3 + D_{12}^3)}, \quad k = 1, 2.$$

Here the notation  $D_{mn}^l \doteq |\mathbf{p}_j^l \mathbf{p}_i^m \mathbf{p}_i^n|$ ,  $l, m, n \in \{1, 2, 3\}$ , was introduced for the triple scalar products with the first vector from camera  $j$  and two vectors from camera  $i$ . For a fixed image point  $(x_i^1, x_i^2)$  whose corresponding epipolar line is sought in image  $j$ , the expression (47) is a function of  $(x_j^1, x_j^2)$  and

$$\frac{dx_j^2}{dx_j^1} = \frac{x_j^2 - \xi^2}{x_j^1 - \xi^1}, \quad (48)$$

with the point  $(\xi^1, \xi^2)$  lying on the epipolar line.

O.d.e. (48) is separable in its variables, and its general solution

$$x_j^2 = \kappa_j x_j^1 + (\xi^2 - \kappa_j \xi^1) \quad (49)$$

is a one-parameter family of straight lines with the slope  $\kappa_j$ . For a particular solution we need an initial value condition to be satisfied. Denote the epipole coordinates by  $\epsilon_j^1, \epsilon_j^2$ . Then the initial condition is

$$\epsilon_j^2 = \kappa_j \epsilon_j^1 + (\xi^2 - \kappa_j \xi^1), \quad \kappa_j = \frac{\epsilon_j^2 - \xi^2}{\epsilon_j^1 - \xi^1},$$

and Eq. (49) transforms to

$$(\epsilon_j^1 - \xi^1)x_j^2 - (\epsilon_j^2 - \xi^2)x_j^1 + (\epsilon_j^2 \xi^1 - \epsilon_j^1 \xi^2) = 0. \quad (50)$$

Each of the following ratios expresses the same property, the slope  $\kappa_j$  of the epipolar line:

$$\frac{\epsilon_j^2 - \xi^2}{\epsilon_j^1 - \xi^1} = \frac{x_j^2 - \xi^2}{x_j^1 - \xi^1} = \frac{\epsilon_j^2 - x_j^2}{\epsilon_j^1 - x_j^1}. \quad (51)$$

All of them lead to the same solution (50).

Eq. (50) is related to the fundamental matrix. It can be written in the form expressing that three points are on the same line:

$$\det \begin{bmatrix} x_j^1 & x_j^2 & 1 \\ \epsilon_j^1 & \epsilon_j^2 & 1 \\ \xi^1 & \xi^2 & 1 \end{bmatrix} = 0, \quad (52)$$

or, equivalently, using the notation of Eq. (47),

$$\begin{aligned} \tilde{\mathbf{x}}_j \cdot [\tilde{\boldsymbol{\epsilon}}_j]_{\times} \cdot \begin{bmatrix} D_{23}^1 & -D_{13}^1 & D_{12}^1 \\ D_{23}^2 & -D_{13}^2 & D_{12}^2 \\ D_{23}^3 & -D_{13}^3 & D_{12}^3 \end{bmatrix} \cdot \tilde{\mathbf{x}}_i = 0, \quad \text{that is,} \\ \tilde{\mathbf{x}}_j \cdot \mathbf{F} \cdot \tilde{\mathbf{x}}_i = 0. \end{aligned} \quad (53)$$

Here the fundamental matrix appears in the factorized form  $\mathbf{F} = [\boldsymbol{\epsilon}]_{\times} \cdot \mathbf{H}$  with the homography  $\mathbf{H}$ . The properties  $\text{rank}(\mathbf{F}) = 2$  and  $\boldsymbol{\epsilon}_j \cdot \mathbf{F} = 0$  are obvious.

Introduce the notations  $f_{kl}$ ,  $k, l = 1, 2, 3$  for the components of the fundamental matrix and  $f_{\alpha\beta}$ ,  $\alpha, \beta = 1, 2$  for the components of its upper-left submatrix  $\mathbf{F}_{2 \times 2}$ . Then Eq. (53) can be re-written as

$$\sum_{\alpha, \beta} f_{\alpha\beta} x_i^\beta x_j^\alpha + \sum_{\beta} f_{3\beta} x_i^\beta + \sum_{\alpha, \beta} f_{\alpha 3} x_j^\alpha + f_{33} = 0. \quad (54)$$

Differentiating w.r.t. the spatial coordinates, we obtain the differential form of Eq. (53)

$$\begin{aligned} \nabla x_j^1 \left( \sum_{\beta} f_{1\beta} x_i^\beta + f_{13} \right) + \nabla x_j^2 \left( \sum_{\beta} f_{2\beta} x_i^\beta + f_{23} \right) + \\ \nabla x_i^1 \left( \sum_{\alpha} f_{\alpha 1} x_j^\alpha + f_{31} \right) + \nabla x_i^2 \left( \sum_{\alpha} f_{\alpha 2} x_j^\alpha + f_{32} \right) = 0 \end{aligned} \quad (55)$$

containing the components of the projection function gradients.

#### 4. Applications of the compatibility equation

The components of the Jacobian (8) in the correspondence equation cannot vary independently if the compatibility with the epipolar geometry is stipulated. This section is devoted to the applications of the compatibility equation (28). *Main results are:* i) the constraint on the fundamental matrix Eq. (62); ii) the epipolar equations (70), (78) and iii) the specific form of the Jacobian for rectified image pair Eq. (83), Eq. (87).

Taking the dot product of the differential fundamental form (55) with  $\nabla x_i^1 \times \nabla x_i^2$ , we have

$$\begin{aligned} \nabla x_j^1 \cdot (\nabla x_i^1 \times \nabla x_i^2) & \left( \sum_{\beta} f_{1\beta} x_i^{\beta} + f_{13} \right) + \\ \nabla x_j^2 \cdot (\nabla x_i^1 \times \nabla x_i^2) & \left( \sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23} \right) = 0, \end{aligned} \quad (56)$$

hence the slope  $\kappa_j = \frac{\nabla x_j^2 \cdot (\nabla x_i^1 \times \nabla x_i^2)}{\nabla x_j^1 \cdot (\nabla x_i^1 \times \nabla x_i^2)}$  expressed in terms of the fundamental matrix is

$$\kappa_j = - \frac{\sum_{\beta} f_{1\beta} x_i^{\beta} + f_{13}}{\sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23}}. \quad (57)$$

Using (57), the differential fundamental form (55) can be rearranged as

$$-\kappa_j \nabla x_j^1 + \nabla x_j^2 + \nabla x_i^1 \frac{\sum_{\alpha} f_{\alpha 1} x_j^{\alpha} + f_{31}}{\sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23}} + \nabla x_i^2 \frac{\sum_{\alpha} f_{\alpha 2} x_j^{\alpha} + f_{32}}{\sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23}} = 0. \quad (58)$$

The rearranged compatibility equation (28)

$$\begin{aligned} \kappa_j \mathbf{n} \cdot (\nabla x_j^1 - a_1^1 \nabla x_i^1 - a_2^1 \nabla x_i^2) \\ = \mathbf{n} \cdot (\nabla x_j^2 - a_1^2 \nabla x_i^1 - a_2^2 \nabla x_i^2) \end{aligned} \quad (59)$$

holds for any normal vector and is in fact a vector equation:

$$\kappa_j \nabla x_j^1 - \kappa_j a_1^1 \nabla x_i^1 - \kappa_j a_2^1 \nabla x_i^2 = \nabla x_j^2 - a_1^2 \nabla x_i^1 - a_2^2 \nabla x_i^2. \quad (60)$$

Comparing Eq. (58) and Eq. (60), we obtain

$$\begin{aligned} \kappa_j a_1^1 - a_1^2 &= \frac{\sum_{\alpha} f_{\alpha 1} x_j^{\alpha} + f_{31}}{\sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23}}, \\ \kappa_j a_2^1 - a_2^2 &= \frac{\sum_{\alpha} f_{\alpha 2} x_j^{\alpha} + f_{32}}{\sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23}}. \end{aligned} \quad (61)$$

Using Eq. (57), the above equation can be expressed via the transpose of the Jacobian  $\mathbf{J}_{ij}$ :

$$-\begin{bmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{bmatrix} \begin{bmatrix} \sum_{\beta} f_{1\beta} x_i^{\beta} + f_{13} \\ \sum_{\beta} f_{2\beta} x_i^{\beta} + f_{23} \end{bmatrix} = \begin{bmatrix} \sum_{\alpha} f_{\alpha 1} x_j^{\alpha} + f_{31} \\ \sum_{\alpha} f_{\alpha 2} x_j^{\alpha} + f_{32} \end{bmatrix} \quad (62)$$

This provides two constraints on the components of the fundamental matrix. Together with the pointwise constraint (53), three constraints are available.

For the coordinates shifted to the origin, Eq. (62) simplifies to

$$-\mathbf{J}_{ij}^T \cdot \mathbf{f}_3 = \mathbf{f}_3,$$

where the notations  $\mathbf{f}_3 = [f_{13} \ f_{23}]^T$  and  $\mathbf{f}_3 = [f_{31} \ f_{32}]^T$  are used. In the studies (Bentolila and Francos, 2014a,b), this formula was obtained using the derivatives of a homography shifted to the origin.

Since the degree of freedom of the fundamental matrix is seven, two independent local region correspondences (point correspondences plus known Jacobians) and a single point pair correspondence are sufficient to determine the fundamental matrix. In the study (Bentolila and Francos, 2014a), the authors proved that two independent local region correspondences constrain the epipole to a second-order rational curve (conic), hence the epipole position can be determined as the intersection of two conics. We will now discuss further consequences of the fundamental constraint (62).

##### 4.1. Epipole constraint

With the notations introduced above, the fundamental matrix can be partitioned as

$$\begin{bmatrix} \mathbf{F}_{2 \times 2} & \mathbf{f}_3 \\ \mathbf{f}_3^T & f_{33} \end{bmatrix}. \quad (63)$$

Then the compatibility constraint (62) associated with a local region can be expressed with inhomogeneous vectors as

$$-\mathbf{J}_{ij}^T \cdot (\mathbf{F}_{2 \times 2} \cdot \mathbf{x}_i + \mathbf{f}_3) = \mathbf{x}_j \cdot \mathbf{F}_{2 \times 2} + \mathbf{f}_3. \quad (64)$$

The pointwise fundamental constraint (54) can be re-grouped in different forms:

$$\begin{aligned} \mathbf{x}_j \cdot \mathbf{F}_{2 \times 2} \cdot \mathbf{x}_i + \mathbf{x}_j \cdot \mathbf{f}_3 + \mathbf{x}_i \cdot \mathbf{f}_3 + f_{33} &= \\ \mathbf{x}_j \cdot (\mathbf{F}_{2 \times 2} \cdot \mathbf{x}_i + \mathbf{f}_3) + (\mathbf{x}_i \cdot \mathbf{f}_3 + f_{33}) &= \\ \mathbf{x}_i \cdot (\mathbf{F}_{2 \times 2}^T \cdot \mathbf{x}_j + \mathbf{f}_3) + (\mathbf{x}_j \cdot \mathbf{f}_3 + f_{33}) &= 0. \end{aligned} \quad (65)$$

Eq. (65) is satisfied by the epipolar lines. In image  $i$  this can be written with the epipole coordinates  $\epsilon_i^1$ ,  $\epsilon_i^2$  and a scale factor  $s_i$  as

$$s_i \begin{bmatrix} x_i^1 & x_i^2 \end{bmatrix} \begin{bmatrix} x_i^2 - \epsilon_i^2 \\ \epsilon_i^1 - x_i^1 \end{bmatrix} + s_i (x_i^1 \epsilon_i^2 - x_i^2 \epsilon_i^1) = 0. \quad (66)$$

Comparing this equation to the last form in Eq. (65), we have

$$s_i \begin{bmatrix} x_i^2 - \epsilon_i^2 & \epsilon_i^1 - x_i^1 \end{bmatrix}^T = \mathbf{F}_{2 \times 2}^T \cdot \mathbf{x}_j + \mathbf{f}_3, \quad (67)$$

for image  $i$ . Similarly, for image  $j$

$$s_j \begin{bmatrix} x_j^2 - \epsilon_j^2 & \epsilon_j^1 - x_j^1 \end{bmatrix}^T = \mathbf{F}_{2 \times 2} \cdot \mathbf{x}_i + \mathbf{f}_3. \quad (68)$$

Substituting these equations into (64), we obtain the compatibility constraint equations (62) translated to the epipoles in the form

$$s_i \begin{bmatrix} x_i^2 - \epsilon_i^2 \\ \epsilon_i^1 - x_i^1 \end{bmatrix} = -s_j \mathbf{J}_{ij}^T \cdot \begin{bmatrix} x_j^2 - \epsilon_j^2 \\ \epsilon_j^1 - x_j^1 \end{bmatrix} \quad (69)$$

with the pointwise fundamental equation (53) already satisfied. This formula has important consequences discussed below.

Applying rotation by  $90^\circ$  to Eq. (69), we have

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_m^2 - \epsilon_m^2 \\ \epsilon_m^1 - x_m^1 \end{bmatrix} = \begin{bmatrix} x_m^1 - \epsilon_m^1 \\ x_m^2 - \epsilon_m^2 \end{bmatrix}, \quad m = i, j. \quad (70)$$

Substituting Eq. (70) into Eq. (69), the latter can be expressed by *differences of two-dimensional inhomogeneous vectors* (ideal points or directions in projective geometry) as

$$\begin{aligned} \mathbf{x}_i - \epsilon_i &= q \mathbf{J}_{ij}^{-1} \cdot (\mathbf{x}_j - \epsilon_j), \\ \mathbf{x}_j - \epsilon_j &= p \mathbf{J}_{ij} \cdot (\mathbf{x}_i - \epsilon_i), \end{aligned} \quad (71)$$

where

$$q \doteq -\det \mathbf{J}_{ij} \frac{s_j}{s_i} \quad \text{and} \quad p \doteq \frac{1}{q}.$$

These equations show that in the epipolar directions all finite vectors transform as coordinate differentials (infinitesimal vectors) up to an unknown scale factor. As a consequence of this property, if the epipole and two corresponding points with given Jacobians are collinear, the compatibility equations (71) become

$$\begin{aligned} \mathbf{x}_i^{(1)} - \epsilon_i &= q^{(1)} \mathbf{J}_{ij}^{-1(1)} \cdot (\mathbf{x}_j^{(1)} - \epsilon_j), \\ \mathbf{x}_i^{(2)} - \epsilon_i &= q^{(2)} \mathbf{J}_{ij}^{-1(2)} \cdot (\mathbf{x}_j^{(2)} - \epsilon_j). \end{aligned} \quad (72)$$

The collinearity conditions expressed with the scale factors  $\lambda_m$  as

$$\mathbf{x}_m^{(2)} - \epsilon_m = \lambda_m (\mathbf{x}_m^{(1)} - \epsilon_m), \quad m = i, j$$

provide the eigenvalue equation

$$\mathbf{J}_{ij}^{(2)} \cdot \mathbf{J}_{ij}^{-1(1)} \cdot (\mathbf{x}_j^{(1)} - \epsilon_j) = \lambda (\mathbf{x}_j^{(1)} - \epsilon_j), \quad (73)$$

where

$$\lambda \doteq \frac{\lambda_j q^{(2)}}{\lambda_i q^{(1)}}.$$

That is, the *epipolar direction vectors*  $\mathbf{x}_m - \epsilon_m$ ,  $m = i, j$  are the eigenvectors of the composition of Jacobians  $\mathbf{J}_{ij}^{(2)} \cdot \mathbf{J}_{ij}^{-1(1)}$  or its inverse. The components of the Jacobians along an epipolar line must obey this eigenvector constraint. Note that the eigenvector property (73) is valid for arbitrary normal vectors of the two patches. From projective geometry view,  $\mathbf{J}_{ij}^{(2)} \cdot \mathbf{J}_{ij}^{-1(1)} \cdot (\mathbf{x}_j^{(1)} - \epsilon_j)$  is the fix (ideal) point of the transformation.

Another formulation for the epipolar constraint can be also derived. Taking the dot product of both sides of Eq. (69) with the vector  $\mathbf{x}_i - \epsilon_i$  perpendicular to the vector in the left side, we obtain a homogeneous equation with the scale factors eliminated:

$$0 = (\mathbf{x}_i - \epsilon_i) \cdot \mathbf{J}_{ij}^T \cdot \begin{bmatrix} x_j^2 - \epsilon_j^2 \\ \epsilon_j^1 - x_j^1 \end{bmatrix} \quad (74)$$

or

$$0 = (\mathbf{x}_j - \epsilon_j) \times [\mathbf{J}_{ij} \cdot (\mathbf{x}_i - \epsilon_i)]. \quad (75)$$

Equation (75) expresses that the epipolar direction in image  $i$  transformed by the Jacobian  $\mathbf{J}_{ij}$  is parallel with the corresponding epipolar direction in image  $j$ . This geometric constraint reduces the necessary equations by one, hence three equations are enough to calculate the four coordinates of the two epipoles.

Applying rotation by  $90^\circ$  to  $\begin{bmatrix} x_j^2 - \epsilon_j^2 & \epsilon_j^1 - x_j^1 \end{bmatrix}^T$ , Eq. (69) can be alternatively written as

$$(\mathbf{x}_i - \epsilon_i) \cdot \hat{\mathbf{J}}_{ij} \cdot (\mathbf{x}_j - \epsilon_j) = 0, \quad (76)$$

where

$$\hat{\mathbf{J}}_{ij} = \begin{bmatrix} -a_1^2 & a_1^1 \\ -a_2^2 & a_2^1 \end{bmatrix}.$$

has a structure similar to the fundamental equation (53). This algebraic formulation of the epipole equation allows the use of overdetermined equation system for robust epipole calculation. Denoting the epipolar direction vectors by  $\mathbf{d}_m \doteq \mathbf{x}_m - \epsilon_m$ ,  $m = i, j$  and normalizing the equation with the first coordinate of the vectors, we have a correspondence equation for the slopes of the epipolar lines as

$$\begin{bmatrix} 1 & \kappa_i \end{bmatrix} \begin{bmatrix} -a_1^2 & a_1^1 \\ -a_2^2 & a_2^1 \end{bmatrix} \begin{bmatrix} 1 \\ \kappa_j \end{bmatrix} = 0, \quad (77)$$

where

$$\kappa_m = \frac{d_m^2}{d_m^1}.$$

From Eq. (77),  $\kappa_j$  can be expressed as

$$\kappa_j = \frac{a_1^2 + a_2^2 \kappa_i}{a_1^1 + a_2^1 \kappa_i}, \quad (78)$$

that is a one-dimensional homography between the slopes.

#### 4.2. Normal constraint

Applying Eq. (28) to Eq. (45), we obtain

$$\begin{aligned} \mathbf{n} \cdot (\kappa_j \nabla x_j^1 - \nabla x_j^2 + a_1^2 \nabla x_i^1 + \\ + a_2^2 \nabla x_i^2 - \kappa_j a_1^1 \nabla x_i^1 - \kappa_j a_2^1 \nabla x_i^2) = 0. \end{aligned} \quad (79)$$

Substituting (45) and (51), we have

$$\begin{aligned} s_j \mathbf{n} \cdot [(a_1^2 - \kappa_j a_1^1)(\mathbf{p}_i^1 - x_i^1 \mathbf{p}_i^3) + (a_2^2 - \kappa_j a_2^1)(\mathbf{p}_i^2 - x_i^2 \mathbf{p}_i^3)] \\ = s_i \mathbf{n} \cdot [\mathbf{p}_j^2 - \kappa_j \mathbf{p}_j^1 + (\kappa_j \epsilon_j^1 - \epsilon_j^2) \mathbf{p}_j^3]. \end{aligned} \quad (80)$$

$s_i, s_j$  are the homogeneous scale factors (i.e., projective depths) for cameras  $i$  and  $j$ . Since the equation must hold for any normal unit vector including  $\mathbf{n} = \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ , we have three equations from which two independent ratios can be used to eliminate the projective depths. These two equations represent the **epipolar constraint** on the components of  $\mathbf{J}_{ij}$  reducing its DOF to two. Note that for the normalized coordinates,  $s_i = d_i, s_j = d_j$  become ‘real’ Euclidean depths, and their ratio has a well-defined meaning. Below we use this epipolar constraint to parameterize



the Jacobian with the components of the normal of the observed patch in a frequently used special configuration.

**Rectified image pair** can be characterized by two special camera matrices and image coordinate system with origin in the optical center:

$$\mathbf{P}_i = \mathbf{K} [\mathbf{I}, \mathbf{0}] = \begin{bmatrix} \alpha & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{P}_j = \mathbf{K} [\mathbf{I}, -d\mathbf{e}_1] = \begin{bmatrix} \alpha & 0 & 0 & -\alpha d \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Using the finite CCD model (31), we have  $\mathbf{p}_i^1 = \mathbf{p}_i^1 = [\alpha \ 0 \ 0]$ ,  $\mathbf{p}_j^2 = \mathbf{p}_j^2 = [0 \ \beta \ 0]$ ,  $\mathbf{p}_i^3 = \mathbf{p}_i^3 = [0 \ 0 \ 1]$ .

Two trivial observations can be made for any imaged spatial point, namely,  $x_j^2 = x_i^2$  and  $s_j = s_i$ . The slope parameter  $\kappa_j$  given by Eq. (47) is zero:  $\kappa_j = 0$ . Since  $\mathbf{p}_j^2 = \mathbf{p}_i^2$ , Eq. (80) becomes

$$\mathbf{n} \cdot [a_1^2(\mathbf{p}_i^1 - x_i^1 \mathbf{p}_i^3) + (a_2^2 - 1)(\mathbf{p}_i^2 - x_i^2 \mathbf{p}_i^3)] = 0. \quad (81)$$

In the directions  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  this yields, respectively,

$$\begin{aligned} a_1^2 \alpha &= 0 \Rightarrow a_1^2 = 0, \\ (a_2^2 - 1)\beta &= 0 \Rightarrow a_2^2 = 1, \\ a_1^2(u^1 - x_i^1) + (a_2^2 - 1)(u^2 - x_i^2) &= 0. \end{aligned} \quad (82)$$

Note that the third condition is satisfied by the solutions of the first two expressing the fact that the depth parameters are identical:  $s_j = s_i$ . The epipolar constraint-compatible Jacobian is therefore written as

$$\mathbf{J}_{ij} = \begin{bmatrix} a_1^1 & a_2^1 \\ 0 & 1 \end{bmatrix}. \quad (83)$$

It has two degrees of freedom. The compatibility constraint (83) was used in paper (Tanacs et al., 2014) for image-based 3D reconstruction. The correspondence equation (8) can be used to translate the parameterization (83) into parameterization with the components of the unit normal vector. For a rectified image pair  $a_2^2 = 1$ , hence the ratio  $\frac{a_1^1}{a_2^1} = a_1^1$ . Applying Eq. (8), we have

$$\begin{aligned} a_1^1 &= \frac{\mathbf{n} \cdot (\nabla x_i^2 \times \nabla x_j^1)}{\mathbf{n} \cdot (\nabla x_j^2 \times \nabla x_i^1)} \\ &= \frac{\mathbf{n} \cdot [(\mathbf{p}_i^2 - x_i^2 \mathbf{p}_i^3) \times (\mathbf{p}_j^1 - x_j^1 \mathbf{p}_j^3)]}{\mathbf{n} \cdot [(\mathbf{p}_j^2 - x_j^2 \mathbf{p}_j^3) \times (\mathbf{p}_i^1 - x_i^1 \mathbf{p}_i^3)]} \\ &= \frac{-n_3 \alpha \beta - x_j^2 n_2 \alpha - x_i^1 n_1 \beta}{-n_3 \alpha \beta - x_i^2 n_2 \alpha - x_j^1 n_1 \beta}, \end{aligned} \quad (84)$$

where the components of the unit normal vector are denoted by  $n_1, n_2, n_3$ . Assuming  $n_3 \neq 0$ , this can be rearranged as

$$a_1^1 = \frac{\frac{n_1}{n_3} \beta (x_i^1 - x_j^1)}{-\frac{n_1}{n_3} \beta x_i^1 - \frac{n_2}{n_3} \alpha x_j^2 - \alpha \beta} + 1. \quad (85)$$

Similarly, for the ratio  $\frac{a_2^1}{a_2^2} = a_2^1$  we obtain

$$a_2^1 = \frac{\frac{n_2}{n_3} \alpha (x_i^1 - x_j^1)}{-\frac{n_1}{n_3} \beta x_i^1 - \frac{n_2}{n_3} \alpha x_j^2 - \alpha \beta}. \quad (86)$$

Consider the special case of square pixels:  $\beta = \alpha$ . Denoting the camera centers by  $u, v$  and introducing  $x^1 \doteq x - u$ ,  $x^2 \doteq y - v$ , the above equations can be transformed to

$$\begin{aligned} a_1^1 &= \frac{\frac{n_1}{n_3} [(x_j - u_j) - (x_i - u_i)]}{\frac{n_1}{n_3} (x_i - u_i) + \frac{n_2}{n_3} (y_j - v_j) + \alpha} + 1, \\ a_2^1 &= \frac{\frac{n_2}{n_3} [(x_j - u_j) - (x_i - u_i)]}{\frac{n_1}{n_3} (x_i - u_i) + \frac{n_2}{n_3} (y_j - v_j) + \alpha}. \end{aligned} \quad (87)$$

These equations were derived in study (Megyesi et al., 2006) by purely geometric considerations.

## 5. Other important camera types

In this section, we apply the proposed general theory to two particular types of nonlinear cameras that are of growing importance for the computer vision community. We derive the fundamental quantities – the coordinate gradients (8) – for the axial and the general spherical camera models focusing on the essential properties responsible for the nonlinear behavior of the models. For the spherical camera, sample applications to reconstruction are considered, including surface normal and relative distance calculation. The pose equations are also discussed.

For better readability of expressions, in this section we use the *summation convention*. Same index in subscript and superscript means summation over the complete range of the index, like

$$a_k b^k = \sum_k a_k b^k.$$

### 5.1. The axial camera

The axial camera can be interpreted as a generalization of the perspective camera such that image points with different distances from the principal point have rays intersecting the optical axis at different points. This focal distance dependency can be described by the radial function  $f(r)$  with  $r = \sqrt{(x^1)^2 + (x^2)^2}$ .

Consider a skew-free camera with rotationally symmetric projection around the optical axis and the image origin coinciding with the principal point. The projection functions are given by  $\mathbf{P} = \mathbf{K} \cdot [\mathbf{R}, \mathbf{t}]$ , where  $\mathbf{K} = \text{diag}[f(r) \ f(r) \ 1]$  encodes the nonlinearity,  $\mathbf{R} = [r_l^k]$ ,  $k, l = 1, 2, 3$ , and  $\mathbf{t} = [t^1 \ t^2 \ t^3]^T$ . In the component form,

$$\begin{aligned} s x^k &= f(r) (r_l^k X^l + t^k), \quad k = 1, 2, \\ s &= r_l^3 X^l + t^3. \end{aligned} \quad (88)$$

The perspective camera (31) with the skew removed ( $\alpha = \beta$ ) and the image origin at the principal point ( $u^1 = u^2 = 0$ ) is a

special case of axial cameras with the constant radial function  $f(r) = \alpha$ .

Taking the derivatives of (88) w.r.t. the spatial coordinates  $X^l$ , we have

$$\begin{aligned} r_l^3 x^1 + s \frac{\partial x^1}{\partial X^l} &= \frac{\partial f}{\partial X^l} \frac{s}{f} x^1 + f r_l^1, \\ r_l^3 x^2 + s \frac{\partial x^2}{\partial X^l} &= \frac{\partial f}{\partial X^l} \frac{s}{f} x^2 + f r_l^2, \end{aligned} \quad (89)$$

where

$$\frac{\partial f}{\partial X^l} = \frac{1}{r} \frac{df}{dr} \left( x^1 \frac{\partial x^1}{\partial X^l} + x^2 \frac{\partial x^2}{\partial X^l} \right). \quad (90)$$

This equation system is to be solved for the gradient components  $\nabla x^k$  of the projection functions for  $k = 1, 2$ . The solution is

$$\begin{aligned} \frac{\partial x^1}{\partial X^l} &= \frac{1}{S} \left[ \frac{1}{r} \frac{df}{dr} \left( r_l^2 x^1 x^2 - r_l^1 (x^2)^2 \right) + f r_l^1 - r_l^3 x^1 \right], \\ \frac{\partial x^2}{\partial X^l} &= \frac{1}{S} \left[ \frac{1}{r} \frac{df}{dr} \left( r_l^1 x^1 x^2 - r_l^2 (x^1)^2 \right) + f r_l^2 - r_l^3 x^2 \right], \\ S &\doteq s \left( 1 - \frac{r}{f} \frac{df}{dr} \right). \end{aligned} \quad (91)$$

Nonlinearity enters a gradient component via its dependence on the other image coordinate. In the camera coordinate system, the projection gradients can be obtained substituting  $s = X^3$  and  $r_l^k = \delta_l^k$ ,  $k, l = 1, 2, 3$ .

## 5.2. The general spherical camera

In this section, we derive the coordinate gradients and the components of the Jacobian for the general spherical camera model (Geyer and Daniilidis, 2000). This model is often used to describe central catadioptric systems as well as some cameras equipped with fish-eye lenses. We focus on the common part of these systems responsible for their nonlinear behavior.

The model proposed in (Geyer and Daniilidis, 2000) unifies all central catadioptric systems. The essential part of the theory (Geyer and Daniilidis, 2000) is the sequential projection of the spatial points. The points are first projected onto a unit ‘projection sphere’, then further projected by a perspective camera creating the catadioptric image. In our context, this process is reversed: the coordinates of the catadioptric image  $x^1, x^2$  parameterize the projection sphere, then the space is parameterized with the sphere points  $\mathbf{x}(x^1, x^2)$  and the distance  $x^3$  from the center of the sphere.

Various catadioptric systems use different conic mirrors, such as the parabolic, the hyperbolic, the elliptical and the planar (the standard perspective camera). They differ in the unit sphere parameterization. We do not restrict ourselves to any specific parameterization assuming only that it exists. An advantage of such ‘inverse’ approach is that a polynomial approximation of the transformation between the catadioptric image and the sphere points can be used directly without calculating the inverse polynomial.

A camera placed in the world coordinate system is specified by its center  $\mathbf{C}$  and rotation  $\mathbf{R}$ . The spatial points  $\mathbf{X}$  are then identified by the parameters  $x^1, x^2, x^3$  such that

$$\mathbf{X} = \mathbf{R}^T \cdot [x^3 \mathbf{x}(x^1, x^2)] + \mathbf{C}. \quad (92)$$

Note that we use the notation consistent with the projective camera. In Eq. (92), the coordinates are transformed from the camera to the world system by  $\mathbf{R}^T$ , while in the projective camera the coordinates are transformed from the world system to the camera by  $\mathbf{R}$ . The *gradients* of the catadioptric image coordinates can be easily calculated from

$$\delta_k^l = \frac{\partial \mathbf{X}}{\partial x^k} \cdot \frac{\partial x^l}{\partial \mathbf{X}} = \mathbf{g}_k \cdot \mathbf{g}^l, \quad k, l = 1, 2, 3. \quad (93)$$

Here  $\mathbf{g}_k = \frac{\partial \mathbf{X}}{\partial x^k}$  are the covariant,  $\mathbf{g}^l \doteq \nabla x^l$  the contravariant basis vectors, and

$$\nabla x^l = \mathbf{g}^l = g^{lk} \mathbf{g}_k, \quad (94)$$

where  $g^{lk}$  are the components of the inverse metric tensor.

From Eq. (92), the covariant basis vectors and the metric tensor components  $g_{lk} = \mathbf{g}_l \cdot \mathbf{g}_k$  are

$$\begin{aligned} \mathbf{g}_k &= x^3 \mathbf{R}^T \cdot \tilde{\mathbf{g}}_k, \\ g_{lk} &= (x^3)^2 \tilde{g}_{lk}, \quad k, l = 1, 2, \\ \mathbf{g}_3 &= \mathbf{R}^T \cdot \mathbf{x}, \\ g_{3k} &= g_{k3} = 0, \quad k = 1, 2 \\ g_{33} &= (\mathbf{R}^T \cdot \mathbf{x}) \cdot (\mathbf{R}^T \cdot \mathbf{x}) = \mathbf{x} \cdot \mathbf{x} = 1, \end{aligned} \quad (95)$$

where  $\tilde{\mathbf{g}}_k = \frac{\partial \mathbf{x}}{\partial x^k}$  and  $\tilde{g}_{lk} = \tilde{\mathbf{g}}_l \cdot \tilde{\mathbf{g}}_k$ ,  $k, l = 1, 2$ , are the covariant basis vectors and the metric components of the parameterization of the unit sphere, respectively. The third equation in (95) means that the point  $\mathbf{x}$  lies on the sphere:

$$x^3 (\mathbf{R}^T \cdot \mathbf{x}) \cdot \left( \mathbf{R}^T \cdot \frac{\partial \mathbf{x}}{\partial x^k} \right) = x^3 \mathbf{x} \cdot \frac{\partial \mathbf{x}}{\partial x^k} = \frac{x^3}{2} \frac{\partial \mathbf{x} \cdot \mathbf{x}}{\partial x^k} = 0. \quad (96)$$

Note that the rotation  $\mathbf{R}$  does not appear in the metric components, and the translation  $\mathbf{C}$  is eliminated by the derivation both from the metric components and the covariant basis vectors. The contravariant basis vectors are given by Eq. (94). Using matrix notation and the quantities of (95), the right-hand side of (94) can be rewritten as

$$\begin{bmatrix} (x^3)^2 \tilde{g}_{11} & (x^3)^2 \tilde{g}_{12} & 0 \\ (x^3)^2 \tilde{g}_{12} & (x^3)^2 \tilde{g}_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x^3 \mathbf{R}^T \cdot \tilde{\mathbf{g}}_1 \\ x^3 \mathbf{R}^T \cdot \tilde{\mathbf{g}}_2 \\ \mathbf{R}^T \cdot \mathbf{x} \end{bmatrix} = 0. \quad (97)$$

The first two contravariant basis vectors, i.e., the coordinate gradients we need, can be independently expressed from the third vector as

$$\begin{aligned} \begin{bmatrix} \nabla x^1 \\ \nabla x^2 \end{bmatrix} &= \frac{1}{x^3} \begin{bmatrix} \tilde{g}_{11} & \tilde{g}_{12} \\ \tilde{g}_{12} & \tilde{g}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}^T \cdot \tilde{\mathbf{g}}_1 \\ \mathbf{R}^T \cdot \tilde{\mathbf{g}}_2 \end{bmatrix} = \\ &= \frac{1}{x^3} \begin{bmatrix} \tilde{g}^{11} & \tilde{g}^{12} \\ \tilde{g}^{12} & \tilde{g}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{R}^T \cdot \tilde{\mathbf{g}}_1 \\ \mathbf{R}^T \cdot \tilde{\mathbf{g}}_2 \end{bmatrix}. \end{aligned} \quad (98)$$

Here the sums  $\frac{1}{x^3} \widetilde{g}^{lk} \widetilde{g}_k$ ,  $l = 1, 2$ , containing the components  $\widetilde{g}^{lk}$  of the inverse metric tensor are the image coordinate gradients  $\nabla x_{cam}^l$  in the coordinate system of the camera's unit projection sphere. The gradients in the camera coordinate system and in the world coordinate system are related by the rotation  $\mathbf{R}^T$ :

$$\nabla x^k = \mathbf{R}^T \cdot \nabla x_{cam}^k. \quad (99)$$

The coordinate gradients  $\nabla x_{cam}^l$  are purely expressed by the unit sphere's local basis vectors  $\widetilde{\mathbf{g}}_k = \frac{\partial \mathbf{x}}{\partial x^k}$  induced by the image coordinates and the distance  $x^3$  from the observed point to the center of the projection sphere.

The above method for calculating the gradients of the projection function requires consistent inverse calculation of the **Jacobian components**. Denote the Cartesian coordinates in the local coordinate systems attached to the projection spheres  $i$  and  $j$  by  $\mathbf{x}_i = [z_i^1 \ z_i^2 \ z_i^3]^T$ ,  $z_i^k = z_i^k(x_i^1, x_i^2)$ , and  $\mathbf{x}_j = [z_j^1 \ z_j^2 \ z_j^3]^T$ ,  $z_j^k = z_j^k(x_j^1, x_j^2)$ . Also, denote the partial derivatives as  $h_l^k \doteq \frac{\partial z_j^k}{\partial x_i^l}$ ,  $k, l = 1, 2, 3$ . Then the components of the Jacobian relating the coordinate differentials in the sphere points of cameras  $i$  and  $j$  as

$$\begin{bmatrix} dz_j^1 \\ dz_j^2 \\ dz_j^3 \end{bmatrix} = \begin{bmatrix} h_1^1 & h_1^2 & h_1^3 \\ h_2^1 & h_2^2 & h_2^3 \\ h_3^1 & h_3^2 & h_3^3 \end{bmatrix} \begin{bmatrix} dz_i^1 \\ dz_i^2 \\ dz_i^3 \end{bmatrix} \quad (100)$$

can be directly estimated. We wish to translate this Jacobian to the Jacobian that acts between image coordinates  $x_j^k$  and  $x_i^l$ ,  $k, l = 1, 2$ .

In the local coordinates, two close points lie on a sphere if

$$(z^1 + dz^1)^2 + (z^2 + dz^2)^2 + (z^3 + dz^3)^2 = (z^1)^2 + (z^2)^2 + (z^3)^2. \quad (101)$$

Omitting the higher order differentials, we have

$$z^1 dz^1 + z^2 dz^2 + z^3 dz^3 = 0, \quad (102)$$

and the third differential is

$$dz^3 = -\left(\frac{z^1}{z^3} dz^1 + \frac{z^2}{z^3} dz^2\right). \quad (103)$$

This differential constraint reduces the DOF of the Jacobian in (100) by one. Only two lines remain linearly independent. Choosing the first two lines and replacing  $dz_i^3$  by (103), we obtain the following relations between the coordinate differentials:

$$\begin{bmatrix} dz_j^1 \\ dz_j^2 \end{bmatrix} = \begin{bmatrix} h_1^1 - \frac{z_1^1}{z_1^3} h_1^3 & h_1^2 - \frac{z_1^2}{z_1^3} h_1^3 \\ h_2^1 - \frac{z_2^1}{z_2^3} h_2^3 & h_2^2 - \frac{z_2^2}{z_2^3} h_2^3 \end{bmatrix} \begin{bmatrix} dz_i^1 \\ dz_i^2 \end{bmatrix}. \quad (104)$$

Assuming bijections between the sphere points  $z^k$  and image points  $x^l$  in the whole domain of estimation, the differentials can be related as

$$\begin{bmatrix} dz^1 \\ dz^2 \end{bmatrix} = \begin{bmatrix} \frac{\partial z^1}{\partial x^1} & \frac{\partial z^1}{\partial x^2} \\ \frac{\partial z^2}{\partial x^1} & \frac{\partial z^2}{\partial x^2} \end{bmatrix} \begin{bmatrix} dx^1 \\ dx^2 \end{bmatrix}, \quad (105)$$

hence the Jacobian that maps the image differentials as  $d\mathbf{x}_j = \mathbf{J}_{ij} \cdot d\mathbf{x}_i$  is

$$\mathbf{J}_{ij} = \begin{bmatrix} \frac{\partial z_j^1}{\partial x_i^1} & \frac{\partial z_j^1}{\partial x_i^2} \\ \frac{\partial z_j^2}{\partial x_i^1} & \frac{\partial z_j^2}{\partial x_i^2} \end{bmatrix}^{-1} \begin{bmatrix} h_1^1 - \frac{z_1^1}{z_1^3} h_1^3 & h_1^2 - \frac{z_1^2}{z_1^3} h_1^3 \\ h_2^1 - \frac{z_2^1}{z_2^3} h_2^3 & h_2^2 - \frac{z_2^2}{z_2^3} h_2^3 \end{bmatrix} \begin{bmatrix} \frac{\partial z_i^1}{\partial x_i^1} & \frac{\partial z_i^1}{\partial x_i^2} \\ \frac{\partial z_i^2}{\partial x_i^1} & \frac{\partial z_i^2}{\partial x_i^2} \end{bmatrix}. \quad (106)$$

The Jacobian contains only the components of the local basis vectors of the projection sphere  $\frac{\partial \mathbf{x}_i}{\partial x_i^k}$ ,  $\frac{\partial \mathbf{x}_j}{\partial x_j^l}$ , and the estimated components  $h_l^k$ . The elements of the Jacobian (106) calculated based on the estimation (100) will be denoted by  $a_i^j$  as usual.

Having the coordinate gradients and the method to estimate the Jacobian components, we can apply them to solve the same problems as in the case of perspective camera. Sketches of the solutions for the spherical camera are presented below.

**Pose estimation.** We assume that camera  $i$  is fully calibrated to the origin of the tangent plane  $\mathbf{n} = \mathbf{e}_3$  ( $Z = 0$ ) of the world coordinate system. Camera  $j$  is internally calibrated, i.e., the points of the projection sphere are given as a function of the catadioptric image coordinates  $\mathbf{x}_j = \mathbf{x}_j(x^1, x^2)$ . The unknowns are all in camera  $j$ , so this index will not be used in the equations below.

For the spherical camera, the general pose equation (42) is specified by the components of the estimated Jacobian (106). Denote by  $A_i^k$  the world coordinates of the known entries on the right-hand side of (42):

$$a_1^k \nabla x_i^1|_T + a_2^k \nabla x_i^2|_T \doteq A_1^k \mathbf{e}_1 + A_2^k \mathbf{e}_2, \quad k = 1, 2. \quad (107)$$

Using Eq. (98), the pose equations for the general spherical camera can be written as

$$\frac{1}{x^3} \left[ (\mathbf{e}_1 \cdot \mathbf{R}^T \cdot \widetilde{g}^{lk} \widetilde{g}_k) \mathbf{e}_1 + (\mathbf{e}_2 \cdot \mathbf{R}^T \cdot \widetilde{g}^{lk} \widetilde{g}_k) \mathbf{e}_2 \right] = A_1^k \mathbf{e}_1 + A_2^k \mathbf{e}_2, \quad (108)$$

where  $l = 1, 2$ . These vector equations must be satisfied for each component providing four equations for the first two rows of  $\mathbf{R}^T$  and the distance  $x^3$  between the camera and the origin, which means the total of 7 unknowns. Additional three equations are provided by constraints on the rotation matrices. Similarly to the projective camera case, the equations can be solved for the 7 unknowns. The ambiguity caused by the quadratic constraints can be resolved by reprojection.

Equations for calculating the **normal vector** were given in section 3.1. Additional technical details, test results and discussion for omnidirectional cameras are provided in our conference paper (Molnár et al., 2014a). Once the unit normal vector has been obtained, the ratio  $\frac{x_j^3}{x_j^1}$  of the distances between the cameras and the object can be determined by any component of the Jacobian (8).

## 6. Tests

We proposed a novel theoretical framework providing an alternative to the mainstream approach. The purpose of the tests presented in this section is to demonstrate that our theory is

technically correct and operational. First, we show that the estimation of the center of distortion is possible using affine correspondences between a planar calibration object and its views. We present quantitative results for synthetic and semi-synthetic cases and also apply our approach to real-world image acquired by wide field of view cameras. Second, we use synthetic data and the perspective camera model to obtain quantitative evaluation results for the minimal pose equation (42) applying the solution (44). Then we show a collection of surface reconstruction results for real image data with radial and tangential lens distortion.

### 6.1. Estimating the center of distortion

Applying equation (64) to the epipolar geometry of radial distortion (Hartley and Kang, 2007) enables one to express the center of distortion (CoD) as the epipole in equation (76) not using elements of the fundamental matrix, but epipoles  $\mathbf{e}_i$  and  $\mathbf{e}_j$ , exclusively. The CoD in our approach is the second epipole,  $\mathbf{e}_j$ .

Here, we present an iterative algorithm to estimate  $\mathbf{e}_1$  and  $\mathbf{e}_2$  with the following steps:

1. Establish Affine Correspondences (AC) between a planar calibration pattern and the camera image. See Fig. 2 for an example. Using correspondences  $\Omega = \{1 \dots n\}$ , the problem can be formulated as

$$\arg \min_{\mathbf{e}_i, \mathbf{e}_j} \sum_{k \in \Omega} (\mathbf{x}_i^{(k)} - \mathbf{e}_i) \cdot \hat{\mathbf{J}}_{ij}^{(k)} \cdot (\mathbf{x}_j^{(k)} - \mathbf{e}_j), \quad (109)$$

where  $(\hat{\mathbf{J}}_{ij}^{(k)}, \mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)})$ ,  $k = 1 \dots n$ , are ACs and  $\hat{\mathbf{J}}_{ij}^{(k)}$ -s are defined in (76).

2. Initialize  $\mathbf{e}_j$  as the center of the image.
3. Find the solution of equation (109) in an iterative manner.
  - a) Assuming  $\mathbf{e}_j$  is known, the solution of the following over-determined system of linear equations gives  $\mathbf{e}_i$ :

$$\begin{bmatrix} (\hat{\mathbf{J}}_{ij}^{(1)} \cdot (\mathbf{x}_j^{(1)} - \mathbf{e}_j))^T \\ \vdots \\ (\hat{\mathbf{J}}_{ij}^{(n)} \cdot (\mathbf{x}_j^{(n)} - \mathbf{e}_j))^T \end{bmatrix} \mathbf{e}_i = \begin{bmatrix} \mathbf{x}_i^{(1)} \cdot \hat{\mathbf{J}}_{ij}^{(1)} \cdot (\mathbf{x}_j^{(1)} - \mathbf{e}_j) \\ \vdots \\ \mathbf{x}_i^{(n)} \cdot \hat{\mathbf{J}}_{ij}^{(n)} \cdot (\mathbf{x}_j^{(n)} - \mathbf{e}_j) \end{bmatrix}, \quad (110)$$

where  $\mathbf{e}_i = [e_i^1 \ e_i^2]^T$ . An equivalent form of this equation is  $\mathbf{P}\mathbf{e}_i = \mathbf{b}$  with  $\mathbf{P} \in \mathbb{R}^{n \times 2}$  and  $\mathbf{b} \in \mathbb{R}^n$ . Its solution is  $\mathbf{e}_i = \mathbf{P}^+ \mathbf{b}$  using the Moore-Penrose pseudoinverse.

- b) Assuming  $\mathbf{e}_i$  is known,  $\mathbf{e}_j$  can be estimated similarly to a).
- c) Repeat these steps until the rate of change in parameters becomes sufficiently small.

As for the ACs, we applied the so-called Tree-based Morse Regions (Xu et al., 2014) feature detector on both the calibration pattern and the camera image to extract corresponding Local Affine Frames (LAFs). A corresponding pair of LAFs  $(\mathbf{B}_i, \mathbf{x}_i)$  and  $(\mathbf{B}_j, \mathbf{x}_j)$  gives an AC  $(\mathbf{B}_j \mathbf{B}_i^{-1}, \mathbf{x}_i, \mathbf{x}_j)$  so that  $\mathbf{J}_{ij} \approx \mathbf{B}_j \mathbf{B}_i^{-1}$ .

#### 6.1.1. Synthetic tests

This section describes synthetic tests of our algorithm. LAFs were produced using the derivatives of the projection function at each point  $\mathbf{p}_i$  of the pattern. Here we used a camera model with slightly misaligned principal point and radial distortion center. We varied the camera pose randomly so that the viewing angle of the calibration pattern is at most 45 degrees.

We also conducted tests to study the influence of noise added to observations. The noise model used for 2D points was the zero-mean Gaussian noise  $\mathcal{N}(\mathbf{0}, \sigma \cdot \mathbf{I})$  with levels of increasing variance  $\sigma$ . However, noise in the affine transformations is not so simple to simulate in a purely synthetic test.

We degraded Jacobians in ACs using multiplicative noise as follows:

$$\mathbf{J}_{ij, (noisy)} = \mathbf{J}_{ij, (gt)} \cdot (\mathbf{I} + \alpha [\mathbf{a} \ \mathbf{b}]), \quad (111)$$

where  $\mathbf{a}, \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma \cdot \mathbf{I})$  and  $\alpha > 0$ . Observe the effect of the noise on the degrading estimation results on Fig. 3.

We conclude that the influence of noise on the affine parameters is a potential source of error, though our synthetic affine noise may be slightly different from real-world conditions.

#### 6.1.2. Semi-synthetic tests

Since true CoD is especially hard to acquire for real-world cameras, the most obvious source of ground truth data – with credible sources of errors in affine parameters – is semi-synthetic data. We sampled our calibration pattern (see Fig. 2) using a synthetic camera with misaligned principal point and radial distortion center. The resulting images are good input for feature detectors (Li et al., 2013) as feature matching is facilitated by their distinct texture.

Our results can be seen in Fig. 4 that shows that (i) the increasing level of radial distortion influences the performance of our algorithm; (ii) the performance improves as input resolution grows. We believe that the source of this negative effect is that the quality of the estimated affine parameters is degraded by the effects of radial distortion.

#### 6.1.3. Tests using real-world cameras

We printed the pattern we used in our semi-synthesized tests and acquired its images using an action camera and higher-end cameras with fish-eye lens. We used normal-quality printer and paper for tests with the GoPro Hero 4 action camera and high-quality for tests with the Canon Rebel T3i equipped with SuperFisheye<sup>11</sup> lens. It appears that in the latter cases the estimated centers of distortions are more consistent through different shots, both for our method and that of (Hartley and Kang, 2007). Note that, for the sake of completeness, we also included results from a number of calibration toolboxes (Ying et al., 2014; Scaramuzza et al., 2006; Bradski, 2000), however, we focus our comparative evaluation on the method (Hartley and Kang, 2007). The results are shown in Fig. 5.

In each test the method (Hartley and Kang, 2007) proved to produce more consistent results for shots from different views.

<sup>11</sup><http://www.superfisheye.com>

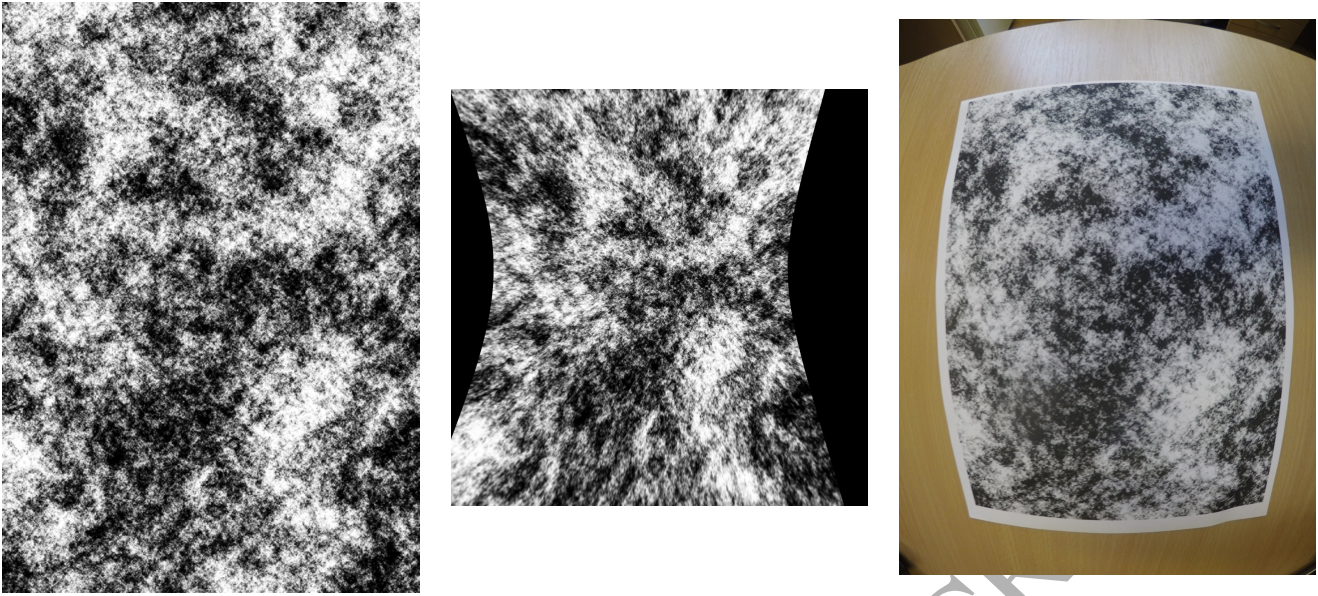


Fig. 2. Calibration pattern used for CoD tests and its shots by different cameras. From left to right: the pattern, a synthetic test camera view and a shot using a GoPro camera.

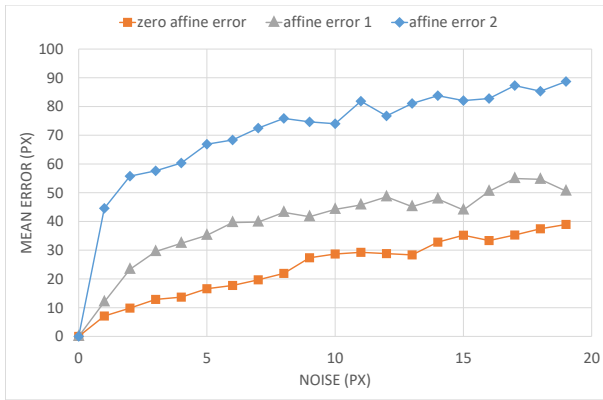


Fig. 3. CoD estimation: Synthetic tests. Diagram of estimation errors w.r.t. ground truth in pixels. “zero affine error”: no affine noise added, “affine error” 1 and 2: increasing levels of noise in affine parameters.

Our algorithm converged not far from these results, in a radius of about five pixels. We believe that, similarly to the semi-synthetic tests, imperfect affine parameters have a negative effect on our algorithm.

#### 6.1.4. Conclusion

We can summarize the tests for the estimation of the center of distortion as follows. In the synthetic tests under low noise, the proposed algorithm works well and yields accurate results comparable to those by the method (Hartley and Kang, 2007). The semi-synthetic and real-world tests demonstrate that the qualities of the image and the calibration pattern are crucial for our method to be able to extract precise affine transformations.

## 6.2. Further tests

### 6.2.1. Synthetic data

In this group of tests, a fully calibrated, virtual perspective camera views a virtual cube from a randomly generated position at a fixed distance from the cube. Then the camera is randomly moved to another position at the same distance preserving the visibility of the cube. A lower and an upper limit on the angle between the two views ( $15^\circ$  and  $45^\circ$ , respectively) were introduced to avoid too close and too far views.

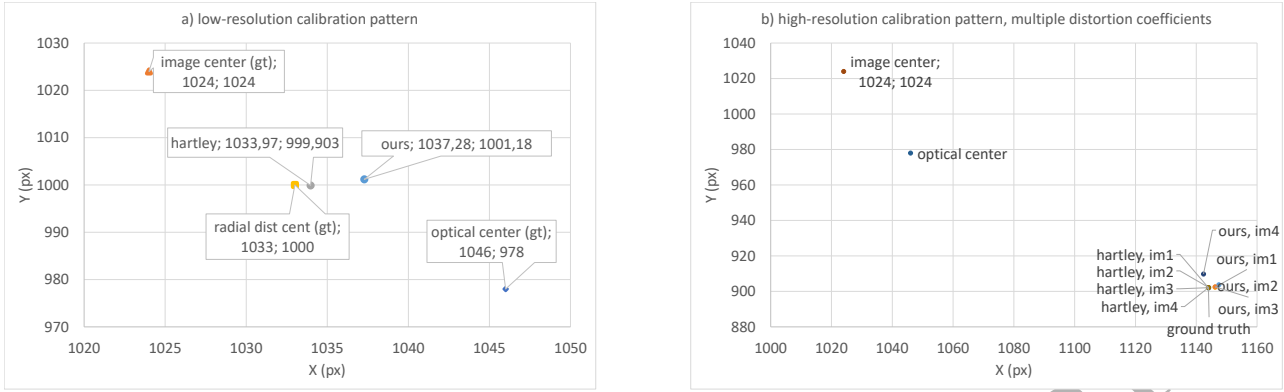
The precise Jacobian components  $a_1^1, a_2^1, \dots$  were calculated based on the known geometry of the stereo pair and the cube. The visible corner points of the cube were used to estimate the homography for each of the visible sides of the cube. To simulate the imprecision of the Jacobian estimation, random noise was added to the corner points in both projections. The strength of the additive Gaussian noise was defined by the squared variance of the distribution. For each noise level, 1000 pairs of views were generated.

Two different methods for homography estimation were tested, namely, the normalized DLT algorithm (Hartley and Zisserman, 2005) and a method proposed in (Barath and Hajder, 2016). The latter uses the epipolar geometry as a constraint, that is, it yields a homography compatible with the epipolar geometry. The obtained homography was linearized in the vicinities of the projections of the corners.

In half of the tests, the ground truth for the pose was used. To test the proposed method for pose estimation, in the other half we obtained the pose using our method. The normal vector was calculated by Eq. (38). For triangulation, we applied our method proposed in section 3.1. As an alternative, we also tested the homogeneous triangulation method (Hartley and Zisserman, 2005) based on the DLT.

For each noise level, the mean and the median errors of the trials were obtained. In the plots, the continuous line is the





**Fig. 4. CoD estimation: Semi-synthetic tests. On diagram a) the results conducted using a virtual camera with a resolution on 1024<sup>2</sup> pixels. Diagram b) shows results on images 1 to 4 shot using increasing radial distortion at a resolution of 2048<sup>2</sup>.**

mean, the dotted line the median value. A large difference between the two curves indicates that the method is less robust. The angular error is the angle, in degrees, between the calculated normal and the ground-truth normal. The spatial position error of the corner points was measured as the percentage of the size of the virtual scene where the range of each coordinate was set to  $[-1, 1]$ . For example, the position error of 0.001 means 0.2% of the scene size. Note that the  $Y$  (error) range varies from plot to plot, while the  $X$  (noise level) range is fixed.

Figure 6 illustrates the effect of homography estimation on the precision of position estimation. In this case, the proposed pose estimation method and the homogeneous triangulation method (Hartley and Zisserman, 2005) were used. The epipolar geometry-compatible homography yields more robust and precise results. However, the results provided by the normalized DLT algorithm (Hartley and Zisserman, 2005) are also acceptable.

Figure 7 demonstrates that the proposed method for pose estimation is reasonably robust and precise as its inclusion in the normal estimation pipeline does not lead to a drastic decrease in performance. Using very accurate pose estimation (simulated by the ground truth) results in excellent normals, but our method also delivers good normal values. In this case, the epipolar-compatible homography (Barath and Hajder, 2016) and the proposed triangulation method were used.

Finally, figure 8 illustrates the effect of triangulation on the precision of position estimation. In this case, the ground-truth pose and the epipolar-compatible homography were used. The proposed triangulation method yields acceptable results, but its performance is worse than that of the homogeneous triangulation (Hartley and Zisserman, 2005). We emphasize that further research is needed to develop *adequate error metric* and *numerical framework* for our approach in a way similar to the studies (Bentolila and Francos, 2014a,b).

Comparing the right-hand sides of figures 8 and 6, one can observe the effect of pose estimation on the position error which is, generally, consistent with the effect on the normal error. An interesting difference is that the dependence of the position er-

ror on the noise level is linear while for the angular error this dependence is non-linear for small noise.

Summarizing the results of the above tests, we conclude that a good selection of the components of the normal and position calculation pipeline is important for the success of the approach. An improper selection can result in a failure. The best results were achieved with the homography compatible with the epipolar geometry (Barath and Hajder, 2016) and the homogeneous triangulation (Hartley and Zisserman, 2005). The proposed triangulation method can also be used, but it needs a more robust implementation.

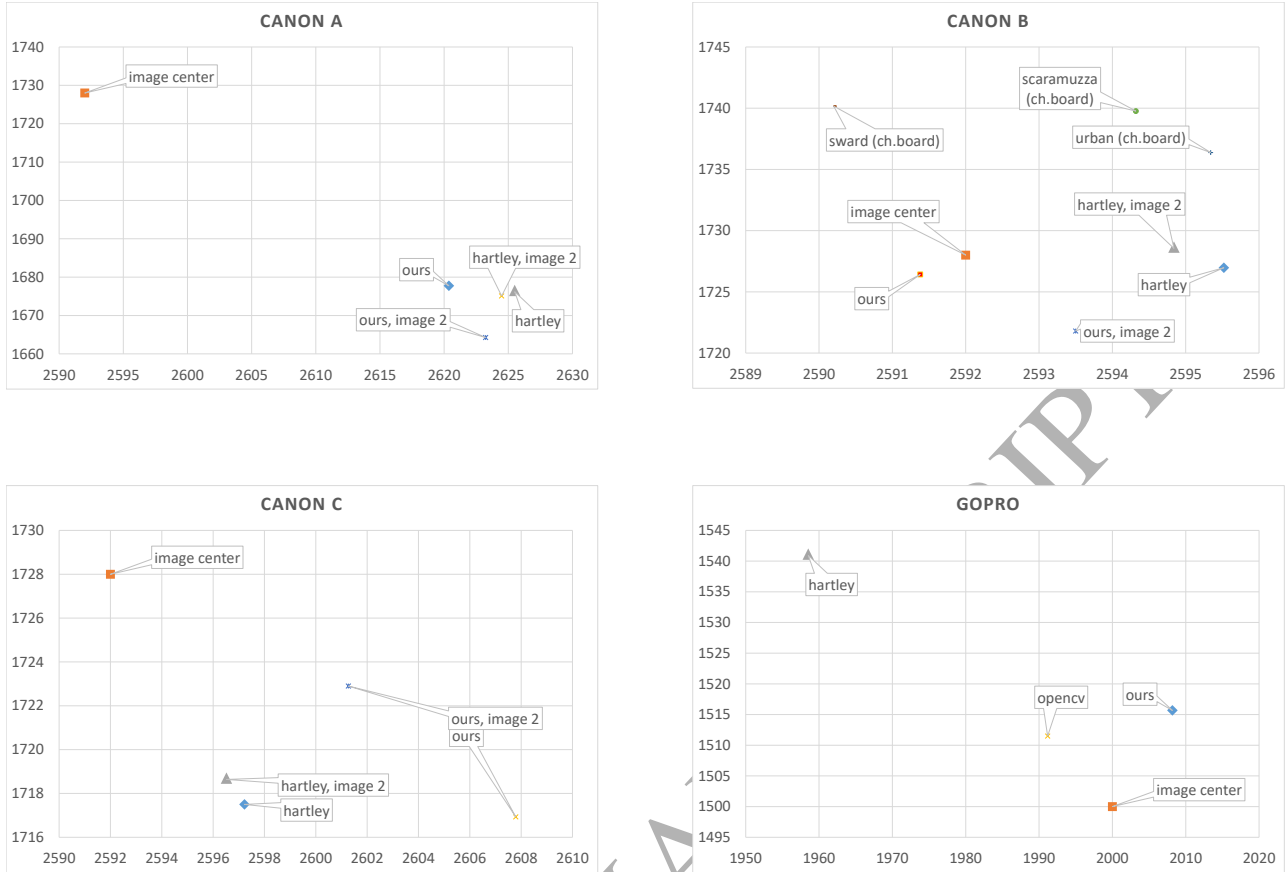
### 6.3. Real data

In this section we present sample surface reconstruction results obtained with the proposed approach. The images used for the reconstruction contain certain radial and tangential lens distortion typical for real data. The reconstructed surfaces range from planar to highly curved.

We start with planar surfaces such as building facades. In this case, the assumption of planarity can be used to robustify the approach. Figure 9 shows a collection of reconstruction results for buildings in Szeged, Hungary. These results were obtained for stereo pairs using homography estimation based on feature correspondence. More input data and reconstruction results for buildings are available at the project web site (University of Szeged, Institute for Computer Science and Control, 2008) where one can interactively view the reconstructed surfaces from desired viewpoints<sup>2</sup>.

Figure 10 illustrates reconstruction of the curved surface of a small object, the Bear (Barath and Eichhardt, 2016). The left side of the figure shows the input stereo images, while the right side displays two views of the surface reconstructed using ASIFT matching (Morel and Yu, 2009). Poisson surface reconstruction (Kazhdan et al., 2006) from oriented point clouds was applied with the orientations calculated by our method.

<sup>2</sup>These results were obtained by a multiview extension of the proposed approach.



**Fig. 5. CoD estimation, real-world tests.** For each Canon 1-2-3 cameras two shots were taken of the calibration object. Results on both images are displayed on the diagrams. Imperfection in the alignment of the Fisheye lens result in different CoD for each Canon cameras. Hartley's method produces consistent results among different shots. Our results are close but seem to be slightly less stable.

Finally, we present two multiview surface reconstruction results for two objects, 'Herz-Jesu-P8' and 'fountain-P11', from the dataset (EPFL, Computer Vision Laboratory, 2008). Both surfaces contain planar and curved regions. Figure 11 shows both non-textured and textured versions of the reconstructed object, while in figure 12 only the non-textured version is displayed that better demonstrates the details of the surface.

## 7. Discussion and conclusion

Traditional approaches to image correspondence are based on projective geometry that operates with points and lines to obtain the fundamental matrix or the trifocal tensor. The proposed alternative approach uses differential geometry and operates with two-dimensional entities, small surface patches.

The correspondence equation (8) is valid when a non-planar surface is close to the tangent plane, that is, the size of the patch is reasonably small. However, for projective camera viewing a planar patch, the Jacobian can be *exactly* determined from homography. This means that for flat surfaces the proposed theory provides exact solution to the surface normal and camera pose estimation problems.

Recently, we have applied the general theory to different kinds of camera models. Results for 3D reconstruction of planar patches viewed by omnidirectional cameras appeared in our study (Molnár et al., 2014a). A promising direction of research could be the development of a second-order theory of image correspondence along the lines proposed in (Molnár and Chetverikov, 2014).

The first-order theory allows for camera pose estimation. Additive second-order entries could possibly bring additional information allowing for autocalibration and reconstruction based on oriented surface patches with associated curvatures (Molnár and Chetverikov, 2014).

## Acknowledgment

This work was supported in part by the European Social Fund [EFOP-3.6.3-VEKOP-16-2017-00001]. It was also supported by the project "Industry 4.0 Research and Innovation Centre of Excellence" [GINOP-2.3.2-15-2016-00002].

Finally, we are grateful to Dmitry Chetverikov for his continuous help and encouragement during this work.

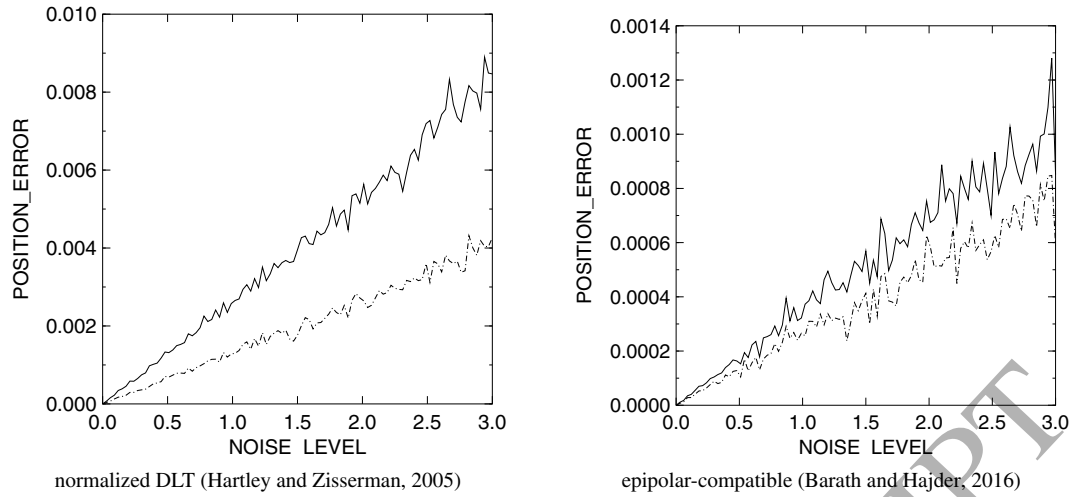


Fig. 6. Effect of homography estimation on position error. Continuous line: median values of the trials. Dotted line: mean values.

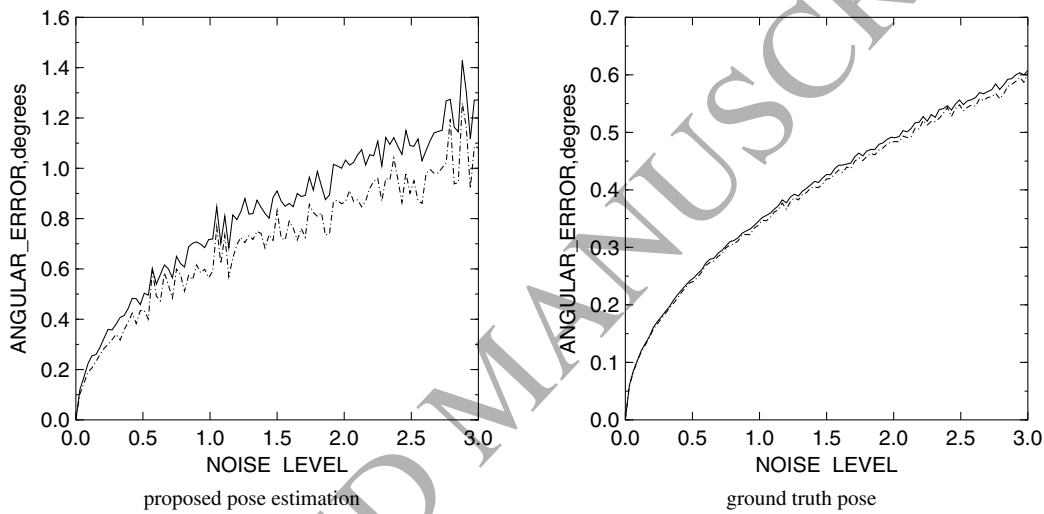


Fig. 7. Effect of pose estimation on normal error.

## References

- Adato, Y., Zickler, T.E., Ben-Shahar, O., 2010. Toward robust estimation of specular flow., in: Proc. British Machine Vision Conf., pp. 1–11.
- Arandjelovic, R., Zisserman, A., 2010. Efficient image retrieval for 3D structures, in: Proc. British Machine Vision Conf., pp. 1–11.
- Barath, D., Eichhardt, I., 2016. A Novel Technique for Point-wise Surface Normal Estimation, in: Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 686–693.
- Barath, D., Hajder, L., 2016. Novel Ways to Estimate Homography from Local Affine Transformations, in: Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 432–443.
- Barath, D., Molnar, J., Hajder, L., 2015. Optimal surface normal from affine transformation, in: International Conference on Computer Vision Theory and Applications, pp. 305–316.
- Belhumeur, P.N., Kriegman, D.J., 1998. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision* 28, 245–260.
- Bentolila, J., Francos, J., 2014a. Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding* 122, 105–114.
- Bentolila, J., Francos, J., 2014b. Homography and fundamental matrix estimation from region matches using an affine error metric. *Journal of Mathematical Imaging and Vision* 49, 481–491.
- Bradski, G., 2000. OpenCV library. *Dr. Dobbs's Journal of Software Tools*.
- Davison, A., Reid, I., Molton, N., Stasse, O., 2007. MonoSLAM: Real-time

single camera SLAM. *IEEE Trans. Pattern Analysis and Machine Intelligence* 29, 1052–1067.

- Devernay, F., Faugeras, O., 1994. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models, in: Conf. on Computer Vision and Pattern Recognition, pp. 208–213.
- Domokos, C., Nemeth, J., Kato, Z., 2012. Nonlinear shape registration without correspondences. *IEEE Trans. Pattern Analysis and Machine Intelligence* 34, 943–958.
- Eichhardt, I., Hajder, L., 2016. Improvement of camera calibration using surface normals, in: Proc. International Conf. on Pattern Recognition, IEEE, pp. 3734–3739.
- EPFL, Computer Vision Laboratory, 2008. Dense multi-view stereo dataset. [cvlabwww.epfl.ch/data/multiview/denseMVS.html](http://cvlabwww.epfl.ch/data/multiview/denseMVS.html).
- Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Analysis and Machine Intelligence* 32, 1362–1376.
- Geyer, C., Daniilidis, K., 2000. A Unifying Theory for Central Panoramic Systems and Practical Applications, in: Proc. European Conf. on Computer Vision, pp. 445–461.
- Gkioulekas, I., Walter, B., Adelson, E.H., Bala, K., Zickler, T., 2015. On the appearance of translucent edges, in: Conf. on Computer Vision and Pattern Recognition, pp. 5528–5536.
- Habbecke, M., Kobbelt, L., 2007. A surface-growing approach to multi-view stereo reconstruction, in: Conf. on Computer Vision and Pattern Recognition, pp. 1–8.
- Hartley, R., Kang, S.B., 2007. Parameter-free radial distortion correction with



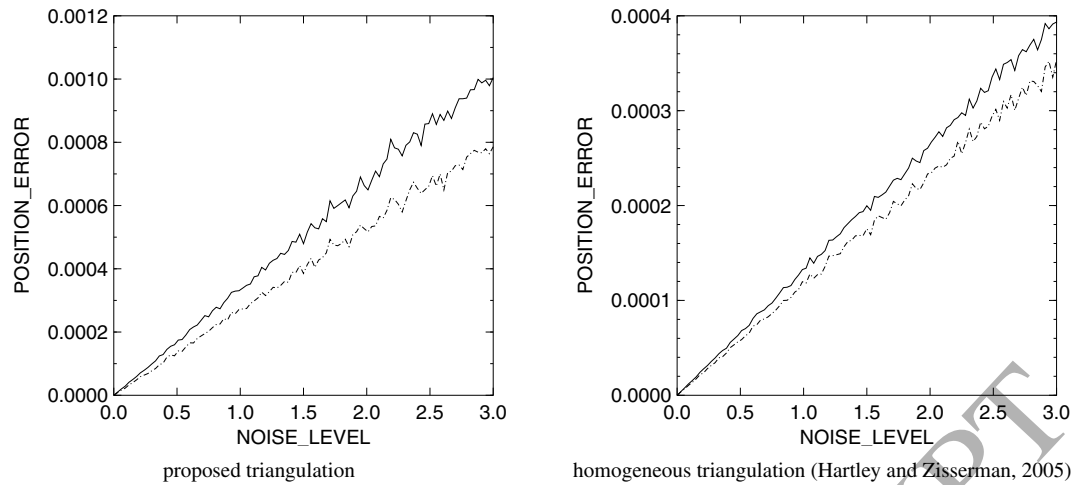


Fig. 8. Effect of triangulation on position error.

- center of distortion estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 29, 1309–1321.
- Hartley, R., Zisserman, A., 2005. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK.
- Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction, in: *Eurographics Symposium on Geometry Processing*, pp. 61–70.
- Köser, K., 2009. *Geometric Estimation with Local Affine Frames and Free-form Surfaces*. Ph.D. thesis. University of Kiel.
- Köser, K., Beder, C., Koch, R., 2008. Conjugate rotation: Parameterization and estimation from an affine feature correspondence, in: *Conf. on Computer Vision and Pattern Recognition*, pp. 1–8.
- Köser, K., Koch, R., 2008. Differential spatial resection-pose estimation using a single local image feature, in: *Proc. European Conf. on Computer Vision*, Springer, pp. 312–325.
- Kutulakos, K., Seitz, S., 1999. A theory of shape by space carving, in: *Proc. International Conf. on Computer Vision*, pp. 307–314.
- Lemaire, T., Berger, C., Jung, I.K., Lacroix, S., 2007. Vision-based SLAM: Stereo and monocular approaches. *International Journal of Computer Vision* 74, 343–364.
- Li, B., Heng, L., Koser, K., Pollefeys, M., 2013. A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern, in: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 1301–1307.
- Magda, S., Kriegman, D.J., Zickler, T., Belhumeur, P.N., 2001. Beyond lambert: Reconstructing surfaces with arbitrary brdfs, in: *Proc. International Conf. on Computer Vision*, IEEE, pp. 391–398.
- Megyesi, Z., Kós, G., Chetverikov, D., 2006. Dense 3D reconstruction from images by normal aided matching. *Machine Graphics & Vision* 15, 3–28.
- Micusik, B., Pajdla, T., 2004. Autocalibration and 3D reconstruction with non-central catadioptric cameras, in: *Conf. on Computer Vision and Pattern Recognition*, pp. 1–58.
- Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaf-falitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. *International Journal of Computer Vision* 65, 43–72.
- Molnár, J., Chetverikov, D., 2014. Quadratic transformation for planar mapping of implicit surfaces. *Journal of Mathematical Imaging and Vision* 48, 176–184.
- Molnár, J., Frohlich, R., Chetverikov, D., Kató, Z., 2014a. 3D Reconstruction of Planar Patches Seen by Omnidirectional Cameras, in: *International Conf. on Digital Image Computing: Techniques and Applications*, pp. 1–8.
- Molnár, J., Huang, R., Kató, Z., 2014b. 3D Reconstruction of Planar Surface Patches: A Direct Solution, in: *ACCV Workshop on Big Data in 3D Computer Vision*, pp. 286–300.
- Molton, N., Davison, A., Reid, I., 2004. Locally planar patch features for real-time structure from motion, in: *Proceedings of the British Machine Vision Conference*, BMVA Press, pp. 90.1–90.10.
- Morel, J.M., Yu, G., 2009. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* 2, 438–469.
- Oxford University, KU Leuven, INRIA, CMP, 2007. *Affine Covariant Features*. [www.robots.ox.ac.uk/vgg/research/affine/](http://www.robots.ox.ac.uk/vgg/research/affine/).
- Perd'och, M., Matas, J., Chum, O., 2006. Epipolar geometry from two correspondences, in: *Proc. International Conf. on Pattern Recognition*, pp. 215–219.
- Riggi, F., Toews, M., Arbel, T., 2006. Fundamental matrix estimation via TIP-transfer of invariant parameters, in: *Proc. International Conf. on Pattern Recognition*, pp. 21–24.
- Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J., 2007. Segmenting, modeling, and matching video clips containing multiple moving objects. *IEEE Trans. Pattern Analysis and Machine Intelligence* 29, 477–491.
- Scaramuzza, D., Martinelli, A., Siegwart, R., 2006. A flexible technique for accurate omnidirectional camera calibration and structure from motion, in: *IEEE International Conference on Computer Vision Systems*, 2006 ICVS'06., IEEE, pp. 45–45.
- Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms, in: *Conf. on Computer Vision and Pattern Recognition*, pp. 519–528.
- Sonka, M., Hlavac, V., Boyle, R., 2008. *Image Processing, Analysis, and Machine Vision*. Thomson.
- Svoboda, T., Pajdla, T., 2002. Epipolar geometry for central catadioptric cameras. *International Journal of Computer Vision* 49, 23–37.
- Tanacs, A., Majdik, A., Molnar, J., Rai, A., Kato, Z., 2014. Establishing correspondences between planar image patches, in: *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, pp. 1–7.
- Tuytelaars, T., Mikolajczyk, K., 2008. Local invariant feature detectors: a survey, in: *Foundations and Trends in Computer Graphics and Vision*. Now Publishers Inc., volume 3, pp. 177–280.
- University of Szeged, Institute for Computer Science and Control, 2008. Patch-Based Multiview Reconstruction of Planar Surfaces. [inf.u-szeged.hu/projectdirs/rgvc/collabreconba2015/](http://inf.u-szeged.hu/projectdirs/rgvc/collabreconba2015/).
- Xu, Y., Monasse, P., Géraud, T., Najman, L., 2014. Tree-based morse regions: A topological approach to local feature detection. *IEEE Transactions on Image Processing* 23, 5612–5625.
- Ying, X., Mei, X., Yang, S., Wang, G., Rong, J., Zha, H., 2014. Imposing differential constraints on radial distortion correction, in: *Asian Conference on Computer Vision*, Springer, pp. 384–398.
- Zickler, T., Belhumeur, P., Kriegman, D., 2002. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision* 49, 215–227.

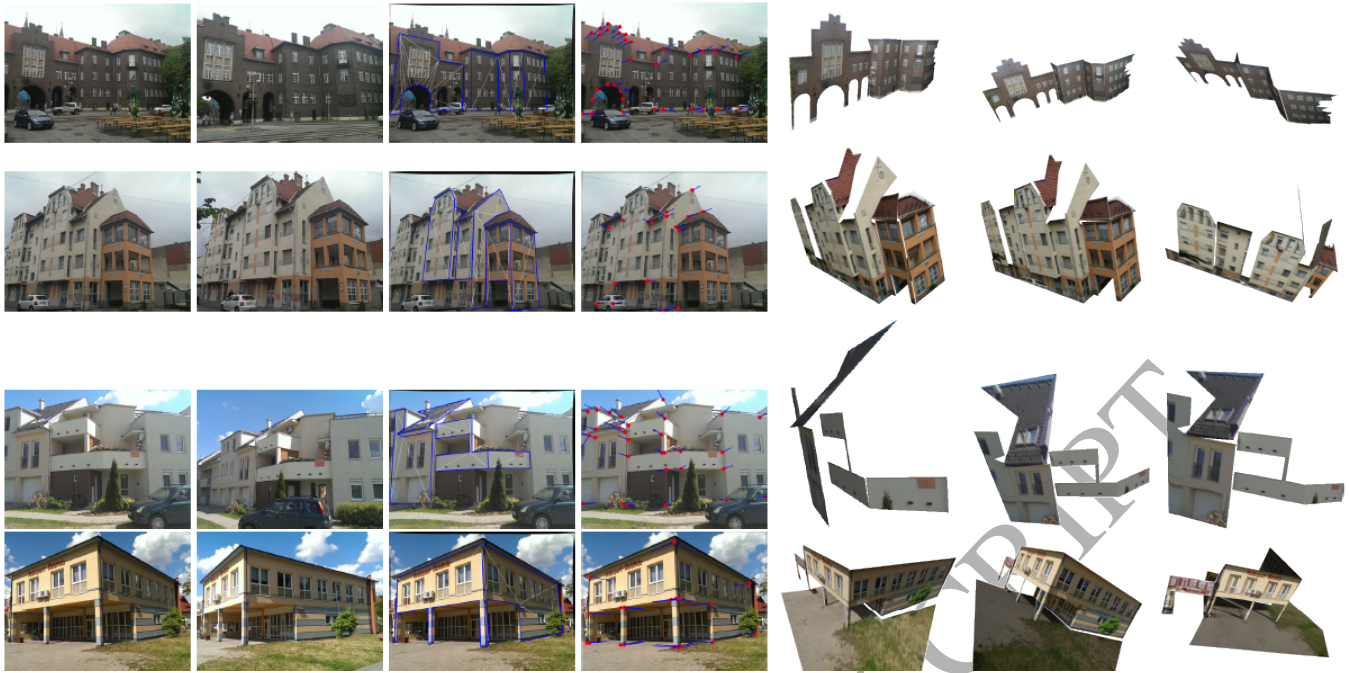


Fig. 9. Surface reconstruction for building facades.



**József Molnár** was born in 1964. He graduated from the Budapest University of Technology and Economics in 1988 (MSc in mechanical engineering). After 20 years of software development he began his PhD studies in 2008, defended his thesis in 2011 at the Eötvös Loránd University, Budapest. His research interests are in the fields of 3D reconstruction, differential geometry, variational image processing methods including segmentation, optical flow estimation and shape analysis.



**Iván Eichhardt** was born in 1989. He received his MSc degree in Computer Science in 2014 from the Eötvös Loránd University, Budapest, Hungary. He is a student of the PhD School of Computer Science at this university. He is also a member of the Systems and Control Lab at the Institute for Computer Science and Control (MTA SZTAKI). His main fields of research are computer vision, image processing, structure-from-motion, 3D reconstruction and sensor fusion.

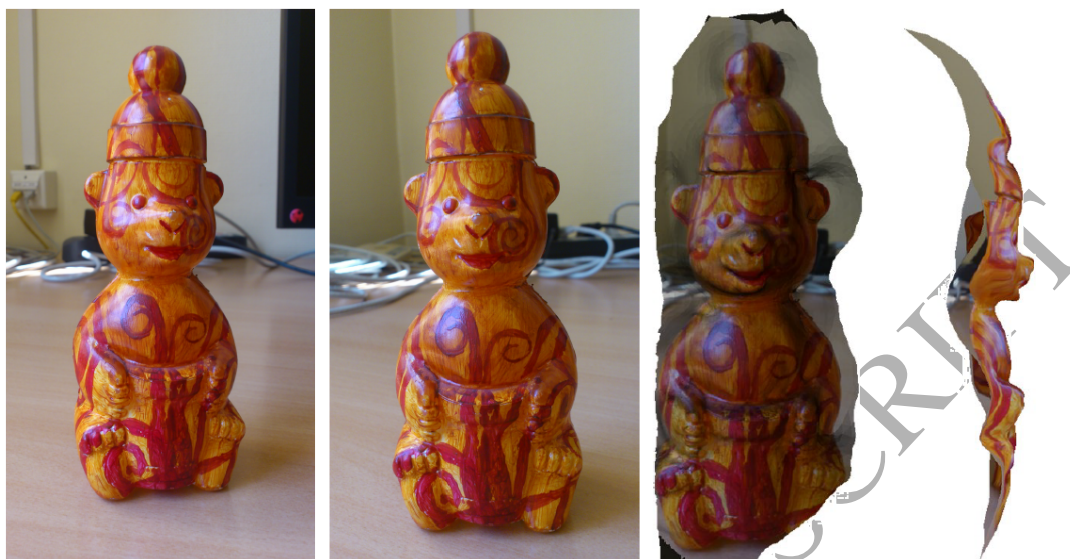


Fig. 10. Surface reconstruction of the Bear. Input images and two views of the reconstructed surface.

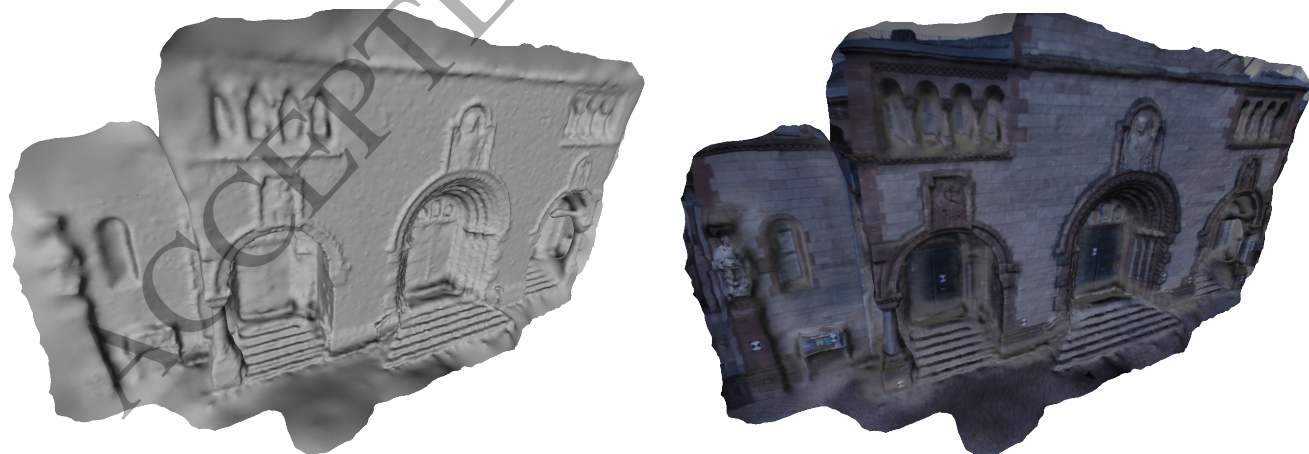


Fig. 11. Surface reconstruction of 'Herz-Jesu-P8'.



Fig. 12. Surface reconstruction (Barath and Eichhardt, 2016) of 'fountain-P11'.