Towards a minimal solution for the relative pose between axial cameras

Francisco Vasconcelos
fpv@isr.uc.pt
João P. Barreto
jpbar@deec.uc.pt

Institute for Systems and Robotics
University of Coimbra
Coimbra, Portugal

An axial camera is a particular case of a non-central camera where every back-projection ray intersects a line in 3D (the axis). The axial camera can be used to model vision systems and imaging situations of practical interest. Examples include any catadioptric system that combines a revolution mirror with a central camera for which the viewpoint is aligned with the mirror axis (e.g. a pinhole looking at a spherical mirror) [8]; the situation of a perspective camera looking through multiple flat refractive mediums [1]; or a multi-camera rig composed by two or more pinhole cameras with collinear optical centers [3].

This paper addresses the problem of estimating the translation $\mathbf{t}$ and the rotation R between two axial cameras using point correspondences. Pless showed that this problem can be linearly solved from a minimum of 17 point correspondences using a DLT like approach [4]. Later in [3, 7] it was observed that for the case of axial cameras this linear estimation could be accomplished from 16 point correspondences.

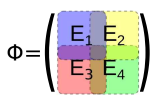
The relative pose problem has 6 unknowns meaning that in theory 6 point correspondences provide enough information for determining the relative rotation and translation of the axial camera. Stewenius et al. proposed in [6] a minimal solution for the relative pose between generalized cameras. However, their algorithm is complex, provides a large number of possible solutions (up to 64), and, as reported in [3], it degenerates for most axial camera configurations. This article does not provide a minimal solution for the relative pose between axial cameras, but shows how the motion can be computed using as few as 10 point correspondences. Our 10-point method is an advance with respect to the previous 16-point [3, 7].

Given that all back-projection rays of an axial camera intersect its axis, they belong to a linear line congruent of dimension 4 [5]. This means that all rays can be represented by 5 dimensional coordinate vectors λ_i that are a linear combination of 5 base lines aligned with the axes $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ in Fig. 1(a).

Given a set of intersecting ray correspondences (λ_i, λ_i'), we can establish linear relations with the form

$$\lambda_i{}^{\mathsf{T}} \Phi \lambda_i' = 0 \tag{1}$$

with Φ being a 5×5 matrix that encodes the 4 essential matrices displayed in Fig. 1(b)



$$\mathsf{E}_1 = \Phi^{\{1:3,1:3\}} = [\mathbf{t}]_\times \mathsf{R} \tag{2}$$
$$\mathsf{E}_2 = \Phi^{\{1:3,3:5\}} = [\mathsf{R}\mathbf{v} + \mathbf{t}]_\times \mathsf{RW} \tag{3}$$
$$\mathsf{E}_3 = \Phi^{\{3:5,1:3\}} = [\mathsf{W}^{\mathsf{T}}(\mathbf{t} - \mathbf{v})]_\times \mathsf{W}^{\mathsf{T}}\mathsf{R} \tag{4}$$
$$\mathsf{E}_4 = \Phi^{\{3:5,3:5\}} = [\mathsf{W}^{\mathsf{T}}(\mathsf{R}\mathbf{v} + \mathbf{t} - \mathbf{v})]_\times \mathsf{W}^{\mathsf{T}}\mathsf{RW} \tag{5}$$

The matrix Φ has 17 free parameters, and therefore can be linearly estimated from 16 correspondences.

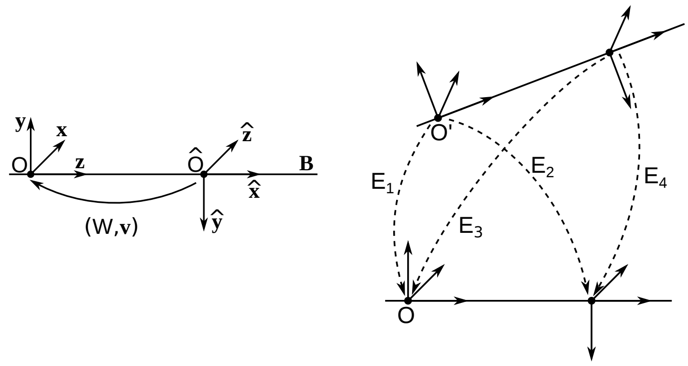
Additionally, the following family of matrices

$$\mathsf{E}_i = \alpha \mathsf{E}_1 + \beta \mathsf{E}_2 \mathsf{W}^{\mathsf{T}} + \gamma \mathsf{W}\mathsf{E}_3 + \delta \mathsf{W}\mathsf{E}_4\mathsf{W}^{\mathsf{T}}, \quad \forall \alpha, \beta, \gamma, \delta \in \Re \tag{6}$$

must have the properties of an essential matrix, and therefore verify the following nonlinear constraints

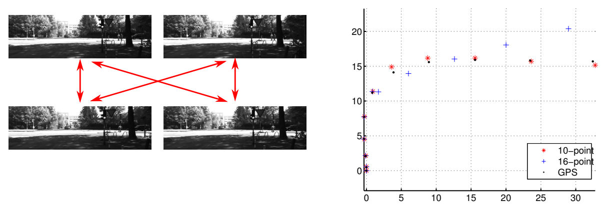
$$2\mathsf{E}_i\mathsf{E}_i{}^{\mathsf{T}}\mathsf{E}_i - tr(\mathsf{E}_i\mathsf{E}_i{}^{\mathsf{T}})\mathsf{E}_i = 0 \tag{7}$$
$$\det \mathsf{E}_i = 0 \tag{8}$$

This makes us able to solve the problem using just 10 correspondences (λ_i, λ_i') by first generating a 7 dimensional linear subspace for Φ and then solving a system of cubic equations in 6 variables, with the action matrix technique [2].



(a) Representation of reference frames along the camera axis $\mathbf{B}$ (b) The four 3×3 essential matrices encoded by Φ

Figure 1: A new parameterization for axial cameras



(a) Input data (b) Trajectory

Figure 2: Performance comparison between 10-point algorithm and 16-point algorithm with real data.

The algorithm is validated and compared against the 16-point algorithm [3] for estimating the relative pose between stereo camera pairs, using both synthetic and real input data (Fig. 2), and showing that our algorithm has a superior performance.

Our long-term goal, however, is to reach a 6-point minimal algorithm, which will require a more in-depth study of the non-linear relations between the essential matrices described in this paper.

[1] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari. A theory of multi-layer flat refractive geometry. In *CVPR*, 2012.

[2] Martin Byröd, Klas Josephson, and Kalle Åström. Fast and stable polynomial equation solving and its application to computer vision. *Int. J. Comput. Vision*, 84(3):237–256, September 2009. ISSN 0920-5691.

[3] Jae-Hak Kim, Hongdong Li, and Richard Hartley. Motion estimation for nonoverlapping multicamera rigs: Linear algebraic and L_∞ geometric solutions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(6): 1044–1059, June 2010.

[4] R. Pless. Using many cameras as one. In *CVPR*, 2003.

[5] H Pottmann and J Wallner. *Computational line geometry*. Springer Verlag, Berlin, 1 edition, 2001.

[6] H. Stewénius, D. Nistér, M. Oskarsson, and K. Åström. Solutions to minimal generalized relative pose problems. In *OMNIVIS*, 2005.

[7] P. Sturm. Multi-view geometry for general camera models. In *CVPR*, 2005.

[8] Peter Sturm, Srikumar Ramalingam, Jean-Philippe Tardif, Simone Gasparini, and João Barreto. Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(1-2):1–183, 2011.