Towards a minimal solution for the relative pose between axial cameras

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

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 rotation $t$ 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 algorithm [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 $\hat{\lambda}_i$ that are a linear combination of 5 base lines aligned with the axes $x, y, z, \hat{x}, \hat{z}$ in Fig. 1(a).

Given a set of intersecting ray correspondences $(\hat{\lambda}_i, \hat{\lambda}_i')$, we can establish linear relations with the form

$$\hat{\lambda}_i^T \Phi \hat{\lambda}_i' = 0$$

with $\Phi$ being a $5 \times 5$ matrix that encodes the 4 essential matrices displayed in Fig. 1(b)

$$\Phi = \begin{pmatrix} E_1 & E_2 \\ E_3 & E_4 \end{pmatrix}$$

$$E_1 = \Phi^{1:3,1:3} = [t]_x R$$

$$E_2 = \Phi^{1:3,3:5} = [Rv + t]_x RW$$

$$E_3 = \Phi^{3:5,1:3} = [W^T(t-v)]_x W^TR$$

$$E_4 = \Phi^{3:5,3:5} = [W^T(Rv + t-v)]_x W^TRW$$

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

Additionally, the following family of matrices

$$E_i = \alpha E_i + \beta E_2W^T + \gamma WE_3 + \delta WE_4W^T, \quad \forall \alpha, \beta, \gamma, \delta \in \mathbb{R}$$

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

$$2E_iE_i^T - tr(E_iE_i^T)E_i = 0$$

$$\det E_i = 0$$

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

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.


