Image and Vision Computing 34 (2015) 45–50

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Generalized essential matrix: Properties of the singular value decomposition☆

Pedro Miraldo *, Helder Araujo

Institute for Systems and Robotics and the Department of Electrical and Computer Engineering, University of Coimbra, Portugal

ARTICLE INFO

Article history:
Received 31 December 2013
Received in revised form 4 November 2014
Accepted 5 November 2014
Available online 29 November 2014

Keywords:
Generalized epipolar geometry
Relative pose
Singular value decomposition
Rigid transformation of lines

ABSTRACT

When considering non-central imaging devices, the computation of the relative pose requires the estimation of the rotation and translation that transform the 3D lines from one coordinate system to the second. In most of the state-of-the-art methods, this transformation is estimated by the computing a 6×6 matrix, known as the generalized essential matrix. To allow a better understanding of this matrix, we derive some properties associated with its singular value decomposition.

1. Introduction

Relative pose estimation is one of the main problems in computer vision, which has been studied for more than a century [1]. The goal is to estimate the rigid transformation between two cameras (or the same camera in two different positions) using matching between pixels that are images of the same 3D point in the world. The cameras (or camera) are considered calibrated. As a result, for each image pixel, we know the corresponding 3D projection line in the world. Thus, by computing the 3D projection lines associated with each match of pixels, the problem can be seen as finding the rotation and translation that align the 3D projection lines to ensure that they intersect in the world, as shown in Fig. 1. One of the most important applications is its use in robotics navigation, in methods such as visual odometry [2].

When considering conventional perspective cameras there are several solutions for the relative pose. We note that there are minimal (5-point algorithms) and non-minimal solutions. One of the goals of minimal solutions is to allow the determination of outliers from a large data-set, to build a robust data-set. On the other hand, the goal of non-minimal solutions is to estimate directly an accurate solution, from a given data-set. A common procedure is to run first the minimal solutions using RANSAC [3,4], followed by iterative refinement, using non-minimal methods. In most of the approaches, authors used the essential matrix [5]. Let us consider a rotation matrix $\mathbf{R} \in \mathcal{SO}(3)$ and a translation vector $\mathbf{t} \in \mathbb{R}^3$, from the epipolar geometry one has

$$\mathbf{d}_i^{(2)^T} \underbrace{s(\mathbf{t})\mathbf{R}}_{\mathbf{E}}\, \mathbf{d}_i^{(1)} = 0, \quad \text{where} \quad s(\mathbf{t}) \subset \mathbb{R}^{3\times 3} \doteq \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}, \quad (1)$$

where $\mathbf{d}_i^{(1)}$ and $\mathbf{d}_i^{(2)}$ denote the inverse projection of two pixels that are the images of the same 3D points with distinct cameras with different external parameters — see Fig. 1(a). Matrix $\mathbf{E} \subset \mathbb{R}^{3\times 3}$ is known as the essential matrix. Some properties associated with the singular value decomposition of $\mathbf{E}$ were derived in [6–8]:

Proposition 1. *The essential matrix* $\mathbf{E}$ *is such that* $\mathbf{E}\mathbf{E}^T$ *only depends on the translation vector* $\mathbf{t}$ *and the singular value decomposition of* $\mathbf{E}\mathbf{E}^T$ *has one singular value equal to zero and other two singular values are equal.*

Based on properties of its singular value decomposition, one can define the following constraints:

Proposition 2. $\mathbf{E}$ *is an essential matrix (which means that it can be decomposed into rotation and translation) if and only if*

$$\det(\mathbf{E}) = 0 \quad \text{and} \quad \frac{1}{2} tr\left(\mathbf{E}\mathbf{E}^T\right)^2 - tr\left(\left(\mathbf{E}\mathbf{E}^T\right)^2\right) = 0.$$

☆ This paper has been recommended for acceptance by Peter Sturm.

* Corresponding author.
E-mail addresses: miraldo@isr.uc.pt (P. Miraldo), helder@isr.uc.pt (H. Araujo).
URL: http://www.deec.uc.pt/miraldo (P. Miraldo).

http://dx.doi.org/10.1016/j.imavis.2014.11.003
0262-8856/

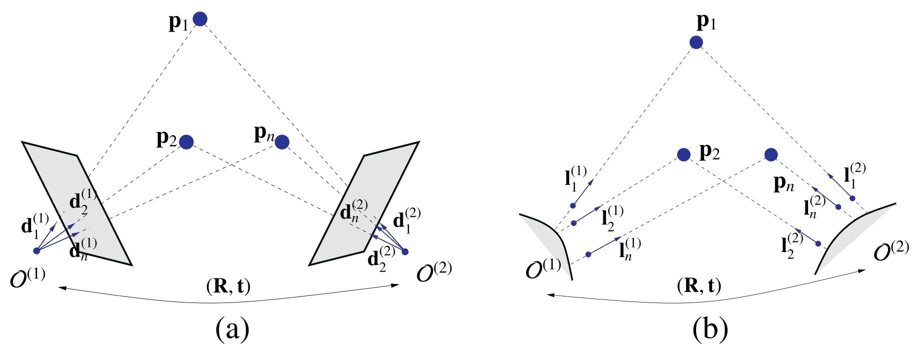

Fig. 1. Representation of the relative pose problem for both central (a) and non-central cases (b).

In addition, the following constraint can also be derived

$$\frac{1}{2}tr\left(\mathbf{E}\mathbf{E}^T\right)\mathbf{E}-\mathbf{E}\mathbf{E}^T\mathbf{E}=0.$$

These constraints (which ensure that $\mathbf{E}$ can be decomposed into rotation and translation in the way shown in Eq. (1)) were used in most of the algorithms for the minimal 5-point relative pose of perspective cameras, for example [9–12]. We note that other solutions (that do not explicitly use these properties) were derived, for example [13].

However, and mainly to get wide field of views, new imaging devices have been developed — for example multiple perspective camera systems, catadioptric cameras or cameras with complex optical systems. In most of these cases, camera models are non-central. As a result, all of these methods for relative pose cannot be used and new algorithms have to be developed.

To deal with general cases (central and non-central camera models) Pless [14] proposed the concept of the generalized epipolar constraint. He considered that a camera can be represented by the general camera model (proposed by Grossberg and Nayar at [15]), which basically assumes that all pixels are mapped into 3D straight lines in the world. Similarly to the case described in the first paragraph, the match of image pixels is mapped into 3D straight lines and the goal is to estimate the rigid transformation that aligns these 3D lines to ensure that they intersect. To represent lines, Pless used *Plücker* coordinates — a line is represented by $\mathbf{l} \doteq (\mathbf{d}, \mathbf{m}) \subset \mathbb{R}^6$ [16] where $\mathbf{d}, \mathbf{m} \in \mathbb{R}^3$ are the direction and moment of the lines respectively. Under this framework, Pless defined the generalized epipolar constraint as:

$$\mathbf{l}_i^{(2)T}\underbrace{\begin{pmatrix} s(\mathbf{t})\mathbf{R} & \mathbf{R} \\ \mathbf{R} & \mathbf{0}\end{pmatrix}}_{\mathcal{E}\subset\mathbb{R}^{6\times 6}}\mathbf{l}_i^{(1)}=0, \tag{2}$$

where $\mathcal{E}$ is denoted as generalized essential matrix. From Eq. (2), one can see that 17 corresponding 3D lines can be used to compute $\mathcal{E}$ linearly. Sturm at [17] studied the properties of the generalized essential matrix when the underlying camera model belongs to central, axial and xslit cameras, which included the minimum number of correspondences between projection rays required for computing essential matrices using linear equations, for each case. Li et al. at [18] show that despite the rank deficiency in the generalized essential matrix for different camera models, it is possible to compute the rotation and translation between two views for different configurations and demonstrate real results on multi-camera configurations. Kim and Kanade at [19] decomposed the generalized essential matrix to study the degenerate cases for a specific type of ray geometry.

To conclude, we note that several algorithms for the relative pose under the framework of generalized camera models have been developed: Lhuillier [20] proposed a generic structure-from-motion method based on an angular error; Schweighofer and Pinz [21] proposed a globally convergent solution to the structure and motion estimation; and Stewenius et al. [10] proposed a solution for the minimal 6-point relative pose problem.

In the case of central cameras the essential matrix has been extensively used to estimate relative pose. The generalized epipolar constraint has been less frequently employed to estimate the relative pose. One of the reasons may be linked to the fact the generalized essential matrix has not been analyzed with the same level of detail as the essential matrix for central cameras. One of the goals of this paper is to derive some properties of the generalized essential matrix allowing a deeper understanding of its structure. In particular we derive some properties of the singular value decomposition of $\mathcal{E}$ (which can be compared to the result of Proposition 1 in the case of $\mathbf{E}$) that should be helpful for the applications of the generalized essential matrix in relative pose applications (specially for the minimal case). We start by considering the following proposition:

Proposition 3. *Matrix $\mathcal{E}$ is full-rank and its determinant is* $\det(\mathcal{E})=1$.

Proof. Since $\mathcal{E}$ is a block triangular matrix and from [22], $\text{rank}(\mathcal{E}) = \text{rank}(\mathbf{R}) + \text{rank}(\mathbf{R})$ and since $\mathbf{R}\in\mathcal{SO}(3)$ implies rank $(\mathbf{R})=3$, one can conclude that rank $(\mathcal{E})=6$ and that the matrix has *full-rank*. Again, since $\mathcal{E}$ is a block triangular matrix we may write $\det(\mathcal{E})=\det(\mathbf{R})\det(\mathbf{R})$ and, since $\mathbf{R}\in\mathcal{SO}(3)$ implies $\det(\mathbf{R})=1$, $\det(\mathcal{E})=1$, proving the proposition. ■

Let us consider the decomposition denoted as

$$\mathcal{E}\doteq\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad \text{where} \quad \mathbf{U},\mathbf{V}\in\mathcal{SO}(6) \quad \text{which implies} \quad \mathbf{U}\mathbf{U}^T=\mathbf{V}\mathbf{V}^T=\mathbf{I} \tag{3}$$

and $\mathbf{\Sigma}$ is a six-dimensional diagonal matrix. For notation, let us consider the representation of the singular value decomposition, such that:

$$\mathcal{E}\mathbf{V}=\mathbf{U}\mathbf{\Sigma} \quad \Rightarrow \quad \mathcal{E}\mathbf{v}_i=\sigma_i\mathbf{u}_i, \quad \forall i=1,\ldots,6, \tag{4}$$

where: σ_i, $\mathbf{v}_i$ and $\mathbf{u}_i$ are called the ith singular value, and right and left singular vectors respectively. $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of $\mathbf{U}$ and $\mathbf{V}$.

The main contributions of the paper are derived in the following section. We propose three theorems, namely: the eigen decomposition of $\mathcal{E}\mathcal{E}^T$ (Theorem 1), the singular value decomposition of $\mathcal{E}$ (Theorem 2), and the sufficient conditions to ensure that a singular value decomposition represents an essential matrix (Theorem 3).

2. Singular value decomposition of $\mathcal{E}$

From Eq. (3), let us consider the following results:

$$\mathcal{E}\mathcal{E}^T=\mathbf{U}\mathbf{\Sigma}\underbrace{\mathbf{V}^T\mathbf{V}}_{\mathbf{I}}\mathbf{\Sigma}\mathbf{U}^T=\mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T \quad \text{and} \quad \mathcal{E}^T\mathcal{E}=\mathbf{V}\mathbf{\Sigma}\underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{I}}\mathbf{\Sigma}\mathbf{V}^T=\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T, \tag{5}$$