A Metric Parametrization for Trifocal Tensors with Non-Colinear Pinholes

Spyridon Leonardos[1], Roberto Tron[1] and Kostas Daniilidis[1]
[1]GRASP Laboratory, University of Pennsylvania.

The trifocal tensor, which describes the relation between projections of points and lines in three views, is a fundamental entity of geometric computer vision. In this work, we investigate a new parametrization of the trifocal tensor for calibrated cameras with non-colinear pinholes obtained from a quotient Riemannian manifold. We incorporate this formulation into state-of-the art methods for optimization on manifolds, and show, through experiments in pose averaging, that it produces a meaningful way to measure distances between trifocal tensors.

The trifocal tensor was first introduced in the context of calibrated geometry to describe relations between projections of lines by Spetsakis and Aloimonos [5] and Weng *et al.* [6]. Later, Hartley [2, 3] generalized the trifocal tensor for the uncalibrated case and Shashua [4] investigated trilinear relations of matched points in three perspective views. There has been numerous works on minimal parametrizations of the projective trifocal tensor. In most formulations, one of the three cameras local reference frame is chosen as the global reference frame.

In this work, we propose a parametrization of the trifocal tensor for calibrated cameras with non-colinear pinholes based on a quotient Riemannian manifold. This parametrization is almost symmetric (we use a preferred camera only for the translations), and is derived from a particular choice of the global reference frame. We show how it can be used for refining estimates of the tensor from image data through state-of-the-art techniques for optimization on manifolds [1]. In addition, the Riemannian structure provides a notion of distance between trifocal tensors. We show that this distance can be computed efficiently, and that it produces meaningful results in a sample Structure from Motion problem.

Let $g_i = (R_i, T_i) \in SE(3)$ be the pose of the i-th camera such that the camera center in the global reference frame is simply given by the translation T_i. Assuming that the cameras are calibrated, the corresponding projection matrices are given by $P_i = \begin{bmatrix} R_i^T & -R_i^T T_i \end{bmatrix} \in \mathbb{R}^{3\times 4}$. Now, let $\{l_i\}_{i=1}^3$ be a set of images of three lines intersecting in 3-D. The intersection of the pre-images of the lines, that is, the three planes with normals $n_i = P_i^T l_i$, $i \in \{1,2,3\}$ is not empty. By defining the *canonical tensor centered on camera* 1 as $\mathcal{T}_i = R_2^T T_{12} e_i^T R_1^T R_3 - R_2^T R_1 e_i T_{13}^T R_3$, the incidence constraint becomes

$$(l_1)_i = l_2^T \mathcal{T}_i l_3 \tag{1}$$

where $(l_1)_i$ stands for the i-th component of vector l_1. We show that any trifocal tensor admits the canonical decomposition

$$\mathcal{T}_i = R_2^T T_{12} e_i^T R_1^T R_3 - R_2^T R_1 e_i T_{13}^T R_3 \tag{2}$$

with $(T_{12})_3 = (T_{13})_3 = 0$ and $\|T_{12}\|^2 + \|T_{13}^2\| = 1$. We write $(T_{12}, T_{13}) \in \mathbb{S}_2^3$ for T_{12}, T_{13} satisfying the previous conditions and we use $\mathbb{S}_2^{3^*}$ if T_{12} and T_{13} are not colinear. Intuitively, the change of world coordinates corresponds to aligning the z-axis with the normal to the plane defined by the three cameras (which is given by $T_{12} \times T_{13}$). This plane is then parallel to the xy-plane, thus the third components of the translations become zero.

We define the *normalized trifocal space* $\mathcal{M}_\mathcal{T}$ as the image of the function $\mathcal{T} : SO(3)^3 \times \mathbb{S}_2^3 \to \mathbb{R}^{3\times 3\times 3}$ that maps the camera poses to the trifocal tensor. Since this mapping is surjective, the space $\mathcal{M}_\mathcal{T}$ corresponds to the space of all trifocal tensors. However, this mapping is not injective. We define the groups

$$H_z = \{(R_z(\theta), R_z(\theta), R_z(\theta), R_z(\theta)) : \theta \in (-\pi, \pi]\} \tag{3}$$

$$H_{x\pi} = \{(I_3, I_3, I_3, I_3), (R_x(\pi), R_x(\pi), R_x(\pi), R_x(\pi))\} \tag{4}$$

$$H_{z\pi} = \{(I_3, I_3, I_3, I_3), (I_3, I_3, I_3, R_z(\pi))\} \tag{5}$$

acting on the left on $SO(3)^3 \times \mathbb{S}_2^{3*}$ by component-wise multiplication. Then, given a point $X \in SO(3)^3 \times \mathbb{S}_2^{3*}$, its equivalence class with respect to " $\sim$ " is given by

$$[X] = \{S_z S_{x\pi} S_{z\pi} X : S_z \in H_z, S_{x\pi} \in H_{x\pi}, S_{z\pi} \in H_{z\pi}\} \tag{6}$$

The above result is in accordance with the mirror image ambiguity according to which, without using the cheirality constraint, the translational parts of the trifocal tensor can be estimated only up to a sign [6]. This ambiguity corresponds to the action of $H_{z\pi}$ and it is intrinsic to the tensor estimation process. In addition, rotating the global reference frame about x-axis by an angle of π results in a z-axis which is still perpendicular to the plane defined by the three cameras. This ambiguity is related to the action of $H_{x\pi}$ and it is an artifact of the particular choice of the global reference frame. As a result, $[X]$ in (6) has four components, each one isomorphic to $SO(2)$. We propose to parametrize the space $\mathcal{M}_\mathcal{T}$ with the quotient space

$$\mathcal{M}_\mathcal{T} = (SO(3)^3 \times \mathbb{S}_2^3)/(H_z \times H_{x\pi} \times H_{z\pi}) \tag{7}$$

Given a point $X \in \mathcal{M}_\mathcal{T}$ we can always pick two of the four components of $[X]$ (the ones corresponding to the positive depths). Thus, if $\overline{\mathcal{M}}_\mathcal{T} = SO(3)^3 \times \mathbb{S}_2^3$, we define the signed trifocal parametrization as:

$$\mathcal{M}_\Upsilon = (SO(3)^3 \times \mathbb{S}_2^3)/(H_z \times H_{x\pi}) = \overline{\mathcal{M}}_\mathcal{T}/(H_z \times H_{x\pi}) \tag{8}$$

This space admits a smooth manifold structure. The tangent space at a point $X \in SO(3)^3 \times \mathbb{S}_2^3$ admits the decomposition into vertical and horizontal spaces $T_X \overline{\mathcal{M}}_\mathcal{T} = \mathcal{V}_X \oplus \mathcal{H}_X$. The signed trifocal manifold $\mathcal{M}_\Upsilon$ admits a structure of a Riemannian quotient manifold with the Riemannian metric

$$g_{[X]}(\xi, \zeta) \doteq \overline{g}_X(\overline{\xi}_X, \overline{\zeta}_X) \tag{9}$$

where $\overline{g}_X$ is the metric of the ambient space and $\overline{\xi}_X$ denotes the horizontal lift of the tangent vector ξ at X.

Next, we show how to obtain geodesics for $\mathcal{M}_\Upsilon$ from geodesics in the ambient space $\overline{\mathcal{M}}_\mathcal{T}$ with horizontal tangent. In addition, we describe an efficient algorithm for computing the logarithm map for the signed trifocal manifold. Intuitively, given two points in $\overline{\mathcal{M}}_\mathcal{T}$, we move the second point to another representative of its equivalence class for which the squared Riemannian distance of $\overline{\mathcal{M}}_\mathcal{T}$ is minimized.

As a first application, we describe how to minimize a cost function that takes as input a trifocal tensor. We combine the parametrization of $\mathcal{M}_\Upsilon$ given by the exponential map with the trust-region methods described in [1]. As a second application, we use the Weiszfeld algorithm to average estimates of trifocal tensors seen as points on the signed trifocal manifold $\mathcal{M}_\Upsilon$. Although it is hard to beat RANSAC, the Weizsfeld algorithm can be used to obtain a sufficiently good initial estimate of the trifocal tensor without the need of tuning a threshold as in RANSAC. Also, the Weiszfeld algorithm performs much better on the quotient manifold compared on a non-quotient parametrization, as anticipated.

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] Richard I. Hartley. Projective reconstruction from line correspondences. In *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 903–907, 1994.

[3] Richard I. Hartley. Lines and points in three views and the trifocal tensor. *Int. J. Comput. Vision*, 22(2):125–140, March 1997.

[4] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:779–789, 1994.

[5] Minas E. Spetsakis and John Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990.

[6] Juyang Weng, Thomas S. Huang, and Narendra Ahuja. Motion and structure from line correspondences; closed-form solution, uniqueness, and optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):318–336, 1992.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.