This CVPR2015 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

A Metric Parametrization for Trifocal Tensors with Non-Colinear Pinholes

Spyridon Leonardos, Roberto Tron and Kostas Daniilidis
GRASP Laboratory, University of Pennsylvania, Philadelphia, PA 19104, USA
{spyridon,tron}@seas.upenn.edu, kostas@cis.upenn.edu

Abstract

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.

1. Introduction

The trifocal tensor was first introduced in the context of calibrated geometry to describe relations between projections of lines by Spetsakis and Aloimonos [25] and Weng *et al*. [32]. Later, Hartley [9, 10] generalized the trifocal tensor for the uncalibrated case and Shashua [23] investigated trilinear relations of matched points in three perspective views.

There has been numerous works on minimal parametrizations of the projective trifocal tensor [4, 16, 18, 20, 22, 28]. In most formulations, one of the three cameras local reference frame is chosen as the global reference frame. A symmetric formulation (where every camera has a similar role) was recently proposed by Ponce and Hebert [20] who minimally parametrized the trifocal tensor by providing necessary and sufficient conditions for three visual rays to converge in terms of three epipolar and one or two trifocal constraints. Symmetric trilinear constraints were also introduced in [21].

Symmetric representations for the two view counterpart of the trifocal tensor, the essential matrix, have been introduced in [7, 12, 27] and are based on the singular value decomposition (SVD) of essential matrices. Geometric insights and further properties of symmetric representations for the space of essential matrices endowed with a Riemannian manifold structure were recently presented by Tron and Daniilidis [29]. However, the study of the space of essential matrices as a Riemannian manifold can be traced back to Soatto *et al*. [24] who formulated structure from motion as a filtering problem on the essential manifold. Later, Ma *et al*. [15] proposed a Riemannian Newton algorithm on the essential manifold for the problem of structure and motion estimation which was later generalized by Vidal *et al*. [31] for multiple views. To the best of our knowledge, analogous representations and manifold structures for the space of calibrated trifocal tensors have not been investigated before.

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.

2. Definitions and notation

In this section, we briefly review several elementary facts from group theory and differential geometry. For a more detailed and rigorous treatment, we refer the reader to the standard texts [1, 5, 14, 19].

A d-dimensional *manifold* $\mathcal{M}$ can be informally defined as a set $\mathcal{M}$ that is locally homeomorphic to the Euclidean space $\mathbb{R}^d$. The tangent space $T_x\mathcal{M}$ at a point $x \in \mathcal{M}$ is the vector space consisting of all the tangents of all smooth curves in $\mathcal{M}$ passing through x. A *Riemannian manifold* is a manifold whose tangent spaces are equipped with a smoothly varying inner product, which is called a *Riemannian metric*. We use the notation $g(\xi, \zeta)$ to denote the inner product of two elements $\xi, \zeta \in T_x\mathcal{M}$ (where the point x will be clear from the context). The metric naturally induces a norm $\|\xi\| \doteq \sqrt{g(\xi, \xi)}$.

A *geodesic curve* on $\mathcal{M}$ is the generalization of a straight line (that is, a curve with zero acceleration). We denote as $\gamma_{x,\xi}(t)$ the geodesic emanating from x in the direction of $\xi \in T_x\mathcal{M}$. The *exponential map* $\exp_x \colon T_x\mathcal{M} \to \mathcal{M}$ is defined as $\exp_x \xi \doteq \gamma_{x,\xi}(1)$. The *logarithm map* $\log_x \colon \mathcal{M} \to$

$T_x\mathcal{M}$ is the inverse of the exponential map and is generally defined only in a neighborhood of x. Where defined, we have the identity $d(x,y) = \|\log_x(y)\|$, where $d(x,y)$ is the Riemannian distance of x, y induced by the metric.

Let $F\colon \mathcal{M} \to \mathcal{N}$ be a smooth map between two manifolds $\mathcal{M}$ and $\mathcal{N}$. The linear mapping $DF(x)\colon T_x\mathcal{M} \to T_{F(x)}\mathcal{M} : \xi \mapsto DF(x)[\xi]$ is called the *differential* of F at x. For any curve $\gamma(t)$ on $\mathcal{M}$ we have $DF(\gamma(t))[\dot{\gamma}(t)] = \frac{d}{dt}F(\gamma(t))$. Furthermore, given a real-valued function $f\colon \mathcal{M} \to \mathbb{R}$, the Riemannian gradient $\operatorname{grad} f(x)$ of f at a point $x \in \mathcal{M}$ is the unique element of $T_x\mathcal{M}$ satisfying

$$g(\operatorname{grad} f(x), \xi) = \frac{d}{dt} f(\gamma_{x,\xi}(t))\bigg|_{t=0} = Df(x)[\xi] \quad (1)$$

for all $\xi \in T_x\mathcal{M}$.

A group $(G, \cdot)$ is a set G along with a binary operation $\cdot : G \times G \to G$ satisfying the axioms of closure, associativity, existence of an identity element $e \in G$ and existence of inverse for each element in the group. A *Lie group* is a group that is also a manifold. If G is a group and $\mathcal{M}$ is a set, a *left action* of G on $\mathcal{M}$ is a map $G \times \mathcal{M} \to \mathcal{M}$, written as $(g,p) \mapsto g \cdot p$, satisfying $g_1 \cdot (g_2 \cdot p) = (g_1 g_2) \cdot p$, for all $g_1, g_2 \in M, p \in \mathcal{M}$ and $e \cdot p = p$ for all $p \in \mathcal{M}$. The action is said *continuous* if the corresponding map is continuous, and it is said *free* if $g \cdot p = p$ for some $p \in M$ implies that $g = e$. A group action induces an equivalence relation $\sim$ on $\mathcal{M}$: for any $x, y \in \mathcal{M}$, $x \sim y$ if $y = g \cdot x$ for some $g \in G$.

Let $\overline{\mathcal{M}}$ be a manifold equipped with an equivalence relation $\sim$. The equivalence class of a point $x \in \overline{\mathcal{M}}$ is denoted by $[x] = \{y \in \overline{\mathcal{M}} : y \sim x\}$. The *quotient space* $M = \overline{\mathcal{M}}/\!\sim$ is the set of all equivalence classes and $\overline{\mathcal{M}}$ is termed the *total space* or *ambient space*. The *canonical projection* is the map $\pi\colon \overline{\mathcal{M}} \to \mathcal{M}$ defined by $\pi(x) = [x]$. The quotient space is called a *quotient manifold* if the canonical projection is a submersion, *i.e.*, the differential of π at every point is surjective. Consider any $x \in \mathcal{M}$ and let $\overline{x} \in \pi^{-1}(x) \subseteq \overline{\mathcal{M}}$. The *vertical space* $\mathcal{V}_{\overline{x}} = T_{\overline{x}}(\pi^{-1}(x))$ at $\overline{x}$ is the tangent space to the equivalence class $\pi^{-1}(x)$. The *horizontal space* $\mathcal{H}_{\overline{x}}$ is the orthogonal complement of $\mathcal{V}_{\overline{x}}$ in $T_{\overline{x}}\overline{\mathcal{M}}$, that is, $\mathcal{V}_{\overline{x}} \oplus \mathcal{H}_{\overline{x}} = T_{\overline{x}}\overline{\mathcal{M}}$. Given any and $\xi \in T_x\mathcal{M}$, there exists exactly one *horizontal lift* $\overline{\xi}_{\overline{x}} \in \mathcal{H}_{\overline{x}}$ satisfying $D\pi(\overline{x})[\overline{\xi}] = \xi$.

In the context of this work, we will frequently use the Lie group of three dimensional rotations $SO(3) = \{R \in \mathbb{R}^{3\times 3} : R^T R = I, \det(R) = 1\}$. The tangent space at a point $R \in SO(3)$ is given by $T_R SO(3) = \{R\Omega : \Omega \in \mathfrak{so}(3)\}$, where $\mathfrak{so}(3)$ denotes the vector space of 3×3 skew-symmetric matrices. The standard metric of $SO(3)$ at a point $R \in SO(3)$ is given by $g(\xi_1, \xi_2) = \frac{1}{2}\operatorname{tr}(\xi_1^T \xi_2) = \frac{1}{2}\operatorname{tr}(\Omega_1^T \Omega_2)$ where $\xi_i = R\Omega_i \in T_R SO(3)$.

For modeling the translational part of the tensor, we will use Kendall's shape space [13]. Following the corresponding notation, we define $\mathbb{S}_2^3 = \{X \in \mathbb{R}^{2\times 2} : \|X\|_F = 1\}$ as the space of triangles in 2-D. The tangent space at a point $X \in \mathbb{S}_2^3$ is $T_X\mathbb{S}_2^3 = \{\xi \in \mathbb{R}^{2\times 2} : \operatorname{tr}(X^T\xi) = 0\} = X^\perp$, the metric is the usual Euclidean inner product. We also introduce the space $\mathbb{S}_2^{3*} = \{X \in \mathbb{S}_2^3 : \operatorname{rank}(X) = 2\} \subset \mathbb{S}_2^3$, which is the space of non-degenerate triangles. For reasons that will be made clear later, we will embed $\mathbb{S}_2^3$ and $\mathbb{S}_2^{3*}$ in $\mathbb{R}^{3\times 2}$ by simply adding a third row paved with zeros.

3. Derivation of the trifocal tensor

In this section, we review the derivation of the trifocal tensor that relates lines seen in three views. This derivation generalizes the one from standard textbooks [11] by not assuming that one of the camera frames coincides with the global reference frame.

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. Then, we have that the matrix $N = \begin{bmatrix} n_1 & n_2 & n_3 \end{bmatrix} \in \mathbb{R}^{4\times 3}$ is rank-deficient. Hence, also the following matrix is rank-deficient:

$$N' = \begin{bmatrix} R_1^T & 0 \\ T_1^T & 1 \end{bmatrix} N = \begin{bmatrix} l_1 & R_1^T R_2 l_2 & R_1^T R_3 l_3 \\ 0 & T_{12}^T R_2 l_2 & T_{13}^T R_3 l_3 \end{bmatrix}, \quad (2)$$

where $T_{ij} = T_i - T_j$. Hence, there must be coefficients α and β such that

$$\begin{bmatrix} l_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1^T R_2 l_2 & R_1^T R_3 l_3 \\ T_{12}^T R_2 l_2 & T_{13}^T R_3 l_3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (3)$$

From the last row, we can choose $\alpha = -T_{13}^T R_3 l_3$ and $\beta = T_{12}^T R_2 l_2$. In this way, we get

$$l_1 = l_2^T R_2^T T_{12} R_1^T R_3 l_3 - l_3^T R_3^T T_{13} R_1^T R_2 l_2 \quad (4)$$

We define the *canonical tensor centered on camera* 1 as

$$\mathcal{T}_i = R_2^T T_{12} e_i^T R_1^T R_3 - R_2^T R_1 e_i T_{13}^T R_3 \quad (5)$$

for $i \in \{1,2,3\}$, where e_i denotes the i-th standard basis vector in $\mathbb{R}^3$. Then, equation (4) becomes

$$(l_1)_i = l_2^T \mathcal{T}_i l_3 \quad (6)$$

where $(l_1)_i$ stands for the i-th component of vector l_1.

4. The normalized trifocal space

In this section, we define the normalized canonical decomposition of the trifocal tensor. Under the assumption of non-colinear cameras, we choose a global reference frame such that the z-axis is aligned with the normal of the plane on which the three cameras lie. Then, we define the normalized trifocal space and parametrize it with a quotient manifold.

4.1. The normalized canonical decomposition

First of all, we will define the canonical decomposition for a trifocal tensor in the following proposition.

Proposition 4.1. *Any trifocal tensor admits the canonical decomposition*

$$\mathcal{T}_i = R_2^T T_{12} e_i^T R_1^T R_3 - R_2^T R_1 e_i T_{13}^T R_3 \tag{7}$$

where $(T_{12})_3 = (T_{13})_3 = 0$ *and* $\|T_{12}\|_2^2 + \|T_{13}\|_2^2 = 1$.

Proof. Since T_{12}, T_{13} are invariant to global translations and since $\|RT_{12}\|_2^2 + \|QT_{13}\|_2^2 = \|T_{12}\|_2^2 + \|T_{13}\|_2^2$ for any $Q, R \in SO(3)$, it follows that the global scale and the global reference frame can be chosen independently. Under the assumption that the three camera centers do not coincide, we can always choose a global scale such that $\|T_{12}\|_2^2 + \|T_{13}\|_2^2 = 1$. Then, given a tensor in the form (5), pick any $R_0 \in SO(3)$ that aligns the z-axis with the vector $T_{12} \times T_{13}$. Then one can verify $(R_0 T_{12})_3 = (R_0 T_{13})_3 = 0$. In conclusion, the tensor can be written as

$$\mathcal{T}_i = (R_0 R_2)^T (R_0 T_{12}) e_i^T (R_0 R_1)^T (R_0 R_3) \tag{8}$$

$$-(R_0 R_2)^T (R_0 R_1) e_i (R_0 T_{13})^T (R_0 R_3) \tag{9}$$

which is in the form (7).

Note that the choice of R_0 in the proof is not unique: if R_0 is a rotation that satisfies the requirements, then any rotation $R_z R_0$, where R_z denotes a rotation around z-axis, will also satisfy the requirements.

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. For any two vectors $T_{12}, T_{13} \in \mathbb{R}^3$ such that $(T_{12})_3 = (T_{13})_3 = 0$ and $\|T_{12}\|_2^2 + \|T_{13}\|_2^2 = 1$, we will write $(T_{12}, T_{13}) \in \mathbb{S}_2^3$. If the camera centers are not colinear, then we have $(T_{12}, T_{13}) \in \mathbb{S}_2^{3*}$.

Definition 4.2. *We define the normalized trifocal space* $\mathcal{M}_\mathcal{T}$ *as the image of the mapping* $\mathcal{T}\colon SO(3)^3 \times \mathbb{S}_2^3 \to \mathbb{R}^{3\times3\times3}$ *defined by*

$$(R_1, R_2, R_3, (T_{12}, T_{13})) \mapsto [\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3] \tag{10}$$

with $\mathcal{T}_i$ *as defined in* (7)*. Since this mapping is surjective by Proposition 4.1, the space* $\mathcal{M}_\mathcal{T}$ *corresponds to the space of all trifocal tensors.*

4.2. Ambiguities of the canonical form

The purpose of this section is to describe the ambiguities of the previously derived canonical form. In the proof of Proposition 4.1, we saw that the mapping from $SO(3)^3 \times \mathbb{S}_2^3$ to $\mathbb{R}^{3\times3\times3}$ as defined in (10) is not injective. We now state the conditions under which two configurations yield the same canonical trifocal tensor. Let $X_a, X_b \in SO(3)^3 \times \mathbb{S}_2^3$. We define the equivalence relation " $\sim$ " on $SO(3)^3 \times \mathbb{S}_2^3$ as

$$X_a \sim X_b \quad \text{iff} \quad \mathcal{T}_a = \mathcal{T}_b. \tag{11}$$

Then, we have the following proposition regarding the equivalence class of a point $X \in SO(3)^3 \times \mathbb{S}_2^{3*}$.

Proposition 4.3. *Define the groups*

$$H_z = \{(R_z(\theta), R_z(\theta), R_z(\theta), R_z(\theta)) : \theta \in (-\pi, \pi]\} \tag{12}$$

$$H_{x\pi} = \{(I_3, I_3, I_3, I_3), (R_x(\pi), R_x(\pi), R_x(\pi), R_x(\pi))\} \tag{13}$$

$$H_{z\pi} = \{(I_3, I_3, I_3, I_3), (I_3, I_3, I_3, R_z(\pi))\} \tag{14}$$

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*

$$[X] = \{S_z S_{x\pi} S_{z\pi} X : S_z \in H_z, S_{x\pi} \in H_{x\pi}, S_{z\pi} \in H_{z\pi}\} \tag{15}$$

A detailed proof of Proposition 4.3 is presented in the supplementary material. 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 [32]. 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 (15) has four components, each one isomorphic to $SO(2)$. We will use $S_z = (S_{z1}, S_{z2}, S_{z3}, S_{z4})$, $S_{z\pi} = (S_{z\pi1}, S_{z\pi2}, S_{z\pi3}, S_{z\pi4})$ and $S_{x\pi} = (S_{x\pi1}, S_{x\pi2}, S_{x\pi3}, S_{x\pi4})$ to denote points in H_z, $H_{z\pi}$ and $H_{x\pi}$ respectively. Based on Proposition 4.3, we propose to parametrize the space $\mathcal{M}_\mathcal{T}$ with the quotient space

$$\mathcal{M}_\mathcal{T} = (SO(3)^3 \times \mathbb{S}_2^3)/(H_z \times H_{x\pi} \times H_{z\pi}) \tag{16}$$

Remark: Proposition 4.3 does not hold for colinear configurations, that is, for points X in the complement of $(SO(3) \times \mathbb{S}_2^{3*})$ in $(SO(3) \times \mathbb{S}_2^3)$. This is because, for these points, the equivalence class $[X]$ contains additional elements given by the rotation around the colinearity axis. Nonetheless, the quotient space in (16) covers all valid tensors $\mathcal{T}$. The only difficulty is that, for and only for colinear tensors, distinct points in $X_a, X_b \in \mathcal{T}$ might yield the same tensor. However, these points constitute a set of measure zero and, in practice, as we will see, this does not prevent (16) (and its signed version, which we introduce later) from being a useful parametrization.

5. The signed trifocal manifold parametrization

In this section, we use the cheirality constraint to fix the mirror image ambiguity. Intuitively, this corresponds to selecting two of the four components of each equivalence class. Then, we show that the resulting space is a Riemannian quotient manifold. Finally, we introduce geodesics, the exponential map and an efficient algorithm for computing the logarithm map.

5.1. Three view depth estimation

Let $X_w \in \mathbb{R}^3$ denote the coordinates of a point p in the world reference frame, $x_1, x_2, x_3 \in \mathbb{R}^3$ the normalized coordinates of point p in each of the three views and $\lambda_1, \lambda_2, \lambda_3 > 0$ the corresponding depths. Then, $X_w = \lambda_i R_i x_i + T_i$ for $i \in \{1, 2, 3\}$. By substituting the first equation into the other two, and by taking into account H_z, $H_{x\pi}$ and $H_{z\pi}$, we have

$$\lambda_2 x_2 = \lambda_1 R_2^T R_1 x_1 + R_2^T S_{z\pi 4} T_{12}, \tag{17}$$

$$\lambda_3 x_3 = \lambda_1 R_3^T R_1 x_1 + R_3^T S_{z\pi 4} T_{13}, \tag{18}$$

where the actions of H_z and $H_{x\pi}$ cancel out. Then, the following proposition follows naturally.

Proposition 5.1. *There is only one choice of $S_{z\pi}$ for which $\lambda_1, \lambda_2, \lambda_3 > 0$.*

Proof. We know that there always exists a choice of $S_{z\pi} \in H_{z\pi}$ (the true one) such that all depths are positive. Denote this solution by $(\lambda_1, \lambda_2, \lambda_3)$. Then, the depths $(-\lambda_1, -\lambda_2, -\lambda_3)$ satisfy (17) and (18) if we now choose $S'_{z\pi} = R_z(\pi) S_{z\pi}$ (that is, the other element in $H_{z\pi}$). Thus, only one choice corresponds to positive depths.

5.2. The signed trifocal manifold parametrization

In view of Proposition 5.1, 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:

$$\mathcal{M}_{\mathcal{T}} = (SO(3)^3 \times \mathbb{S}_2^3)/(H_z \times H_{x\pi}) = \overline{\mathcal{M}}_{\mathcal{T}}/(H_z \times H_{x\pi}) \tag{19}$$

This space admits a smooth manifold structure, as shown next.

Proposition 5.2. *The canonical projection $\pi \colon \overline{\mathcal{M}}_{\mathcal{T}} \to \mathcal{M}_{\mathcal{T}}$ is a smooth submersion and $\mathcal{M}_{\mathcal{T}}$ is a manifold of dimension 11.*

Proof. Since $H_z \times H_{x\pi}$ is a compact Lie group and the action is continuous, it follows that the action is proper. Moreover, $S_{z1} S_{x\pi 1} R_1 = R_1$ implies $S_{z1} = I_3$ and $S_{x\pi 1} = I_3$. As a result, for any $X \in SO(3)^3 \times \mathbb{S}_2^3$, we have $S_z S_{x\pi} X = X$ implies that $S_z S_{x\pi}$ is the identity element of the group. Thus, the action is also free. Finally, the action is trivially smooth and we conclude $\mathcal{M}_{\mathcal{T}}$ is a manifold of dimension:

$$\dim \mathcal{M}_{\mathcal{T}} = \dim\big(SO(3)^3 \times \mathbb{S}_2^3\big) - \dim\big(H_z \times H_{x\pi}\big) = 11 \tag{20}$$

where we remind the reader that $H_{x\pi}$ is discrete.

Proposition 5.2 implies that the tangent space at a point $X \in SO(3)^3 \times \mathbb{S}_2^3$ admits the decomposition into vertical and horizontal spaces

$$T_X \overline{\mathcal{M}}_{\mathcal{T}} = \mathcal{V}_X \oplus \mathcal{H}_X. \tag{21}$$

We can give a closed form expression for the vertical space:

Proposition 5.3. *The vertical space at a point $X = (R_1, R_2, R_3, T) \in \overline{\mathcal{M}}_{\mathcal{T}}$ is given by*

$$\mathcal{V}_X = \{\lambda \widehat{e_z} \cdot (R_1, R_2, R_3, T) : \lambda \in \mathbb{R}\} \tag{22}$$

where $\cdot$ denotes component-wise multiplication, $e_z = (0, 0, 1)^T$ and $\widehat{(\cdot)}$ denotes the usual hat operator of a vector in $\mathbb{R}^3$.

Proof. Let $X \in \overline{\mathcal{M}}_{\mathcal{T}}$ and note that since $H_z \times H_{x\pi}$ is one dimensional, also $\mathcal{V}_X$ is one dimensional. Let $\gamma(t) = R(t) \cdot X$ be a curve in the equivalence class of X with $R(t) \in H_z$ for all t and $R(0) = I$. Then, $\dot{\gamma}(0) = \widehat{e_z} \cdot X = \widehat{e_z} X$. It follows that $\mathcal{V}_X$ is spanned by the vector $\widehat{e_z} X$.

At this point, we will endow $T_X \overline{\mathcal{M}}_{\mathcal{T}}$ with a Riemannian metric which is necessary for defining the orthogonal projection of a vector onto the vertical and horizontal spaces, and of course for defining a metric on the signed trifocal manifold. Let any $X = (R_1, R_2, R_3, T) \in \overline{\mathcal{M}}_{\mathcal{T}}$. A Riemannian metric $\overline{g}$ for $T_X \overline{\mathcal{M}}_{\mathcal{T}}$ can be naturally defined as

$$\overline{g}(\xi, \zeta) = \frac{1}{2} \sum_{i=1}^{3} \operatorname{tr}(\xi_i^T \zeta_i) + \operatorname{tr}(\xi_4^T \zeta_4) \tag{23}$$

where $\xi = (\xi_1, \xi_2, \xi_3, \xi_4)$, $\zeta = (\zeta_1, \zeta_2, \zeta_3, \zeta_4)$ with $\xi_i, \zeta_i \in T_{R_i} SO(3)$ for $i \in \{1, 2, 3\}$ and $\xi_4, \zeta_4 \in T_T \mathbb{S}_2^3$. Now, the orthogonal projection of a tangent vector $\xi \in T_X \overline{\mathcal{M}}_{\mathcal{T}}$ onto the vertical space $\mathcal{V}_X$ is given by

$$P_X^v \xi = \widehat{e_z} X \frac{\overline{g}(\xi, \widehat{e_z} X)}{\overline{g}(\widehat{e_z} X, \widehat{e_z} X)} = \frac{1}{4} \overline{g}(\xi, \widehat{e_z} X) \widehat{e_z} X \tag{24}$$

and the corresponding orthogonal projection of a tangent vector $\xi \in T_X \overline{\mathcal{M}}_{\mathcal{T}}$ onto the horizontal $\mathcal{H}_X$ is simply given by $P_X^h \xi = \xi - P_X^v \xi$.

Next, we will endow $\mathcal{M}_{\mathcal{T}}$ with a Riemannian metric. We will need the following proposition relating the horizontal lifts of the same tangent vector of the quotient space at two distinct points in the same equivalence class.

Proposition 5.4. *Let $X \in \overline{\mathcal{M}}_{\mathcal{T}}$ and $\xi \in T_{[X]}\mathcal{M}_{\mathcal{T}}$. Then*

$$\overline{\xi}_{RX} = R\overline{\xi}_X \tag{25}$$

for all $R \in H_z \times H_{x\pi}$, where $\overline{\xi}_X$ denotes the horizontal lift of a tangent vector ξ at X.

The proof can be found in the supplementary material. It is based on the proof of Absil *et al.* [1] for the case of the real projective space. We then arrive at the desired result.

Proposition 5.5. *The signed trifocal manifold $\mathcal{M}_{\mathcal{T}}$ admits a structure of a Riemannian quotient manifold with the Riemannian metric*

$$g_{[X]}(\xi, \zeta) \doteq \overline{g}_X(\overline{\xi}_X, \overline{\zeta}_X) \tag{26}$$

Proof. Let $\overline{g}_X$ be the Riemennian metric of $\overline{\mathcal{M}}_{\mathcal{T}}$ at $X \in \overline{\mathcal{M}}_{\mathcal{T}}$ defined in (23). Then, we have

$$\begin{aligned}
&\overline{g}_{RX}(\overline{\xi}_{RX}, \overline{\zeta}_{RX}) = \overline{g}_{RX}(R\overline{\xi}_X, R\overline{\zeta}_X) \\
&= \frac{1}{2}\sum_{i=1}^{3} \operatorname{tr}((R\xi_{X,i})^T(R\zeta_{X,i})) + \operatorname{tr}((R\xi_{X,4})^T(R\zeta_{X,4})) \\
&= \frac{1}{2}\sum_{i=1}^{3} \operatorname{tr}(\xi_{X,i}^T\zeta_{X,i}) + \operatorname{tr}(\xi_{X,4}^T\zeta_{X,4}) = \overline{g}_X(\overline{\xi}_X, \overline{\zeta}_X)
\end{aligned}$$

The metric (26) does not depend on the choice of the representative of each equivalence class and thus, it is a well-defined Riemannian metric.

5.3. Geodesics and the exponential map

In this section, we show how to obtain geodesics for $\mathcal{M}_{\mathcal{T}}$ from geodesics in the ambient space $\overline{\mathcal{M}}_{\mathcal{T}}$ with horizontal tangent. The idea has been repeatedly used in [6] to obtain geodesics for the Stiefel and Grassmann manifold from geodesics of the orthogonal group and in [29] to obtain geodesics of the Essential manifold from geodesics of $SO(3)^2$. Since the projection $\pi_{\mathcal{M}_{\mathcal{T}}} : \overline{\mathcal{M}}_{\mathcal{T}} \to \mathcal{M}_{\mathcal{T}}$ is a Riemannian submersion, *i.e.* a submersion that preserves the metric, we have the following proposition [17]:

Proposition 5.6. *Let $\gamma(t)$ be a geodesic on $\overline{\mathcal{M}}_{\mathcal{T}}$ such that $\dot{\gamma}(t) \in \mathcal{H}_{\gamma(t)}$ for all t. Then, $\pi_{\mathcal{M}_{\mathcal{T}}}(\gamma(t)) = [\gamma(t)]$ is a geodesic on $\mathcal{M}_{\mathcal{T}}$.*

Moreover, we have the following proposition for geodesics with horizontal initial tangents.

Proposition 5.7. *Let $\gamma_{X,\xi}(t)$ be a geodesic on $\overline{\mathcal{M}}_{\mathcal{T}}$ emanating from $X = \gamma_{X,\xi}(0)$ with initial velocity $\xi = \dot{\gamma}_{X,\xi}(0)$. If $\xi = \dot{\gamma}_{X,\xi}(0) \in \mathcal{H}_X$, then $\dot{\gamma}_{X,\xi}(t) \in \mathcal{H}_{\gamma_{X,\xi}(t)}$ for all t.*

The proof can be found in the supplementary material and is quite elementary. The above result combined with Proposition 5.6 shows that if $\gamma(t)$ is a geodesic on $\overline{\mathcal{M}}_{\mathcal{T}}$ with $\dot{\gamma}(0) \in \mathcal{H}_{\gamma(0)}$, *i.e.* initial tangent belonging to the horizontal space, then $[\gamma(t)]$ is a geodesic in the quotient space $\mathcal{M}_{\mathcal{T}}$. Thus, the exponential map $\exp\colon T_{[X_a]}\mathcal{M}_{\mathcal{T}} \to \mathcal{M}_{\mathcal{T}}$ is defined as $[X_b] = \exp_{[X_a]}(\xi)$ and can be computed by $X_b = \exp_{X_a}(\overline{\xi}_{X_a})$, where $\overline{\xi}_{X_a}$ is the horizontal lift of ξ at X_a and $\exp_{X_a}$ is the exponential map of $\overline{\mathcal{M}}_{\mathcal{T}} = SO(3)^3 \times \mathbb{S}_2^3$.

5.4. The logarithm map and Riemannian distance

In this section, we will determine the logarithm map for the signed trifocal manifold from its ambient space, and describe an efficient algorithm for computing it. The Riemannian distance is then given by the norm of the logarithm map. Intuitively, given two points in $\overline{\mathcal{M}}_{\mathcal{T}}$, we will 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. This change of representative will yield a horizontal vector as we will show in Proposition 5.8.

Let $X_a, X_b \in \overline{\mathcal{M}}_{\mathcal{T}}$, $R_{x\pi} \in H_{x\pi}$ and $R_z(t)$ denote a rotation around z axis of angle t. Moreover, let $\theta_i(t) = \arccos\left((\operatorname{tr}(R_{ai}^T R_z(t) R_{x\pi} R_{bi}) - 1)/2\right)$ for $i \in \{1, 2, 3\}$ and $\theta_4(t) = \arccos(\operatorname{tr}(T_a^T R_z(t) R_{x\pi} T_b))$ (these are the geodesic distances in $SO(3)$ and in $\mathbb{S}_2^3$). We then have:

Proposition 5.8. *Define the cost function*

$$\begin{aligned}
f(t) &= \sum_{i=1}^{4} f_i(t) = \sum_{i=1}^{4} \frac{1}{2}\theta_i^2(t) \\
&= \frac{1}{2}\sum_{i=1}^{3} d^2(R_{ai}, R_z(t)R_{x\pi}R_{bi}) + \frac{1}{2}d^2(T_a, R_z(t)R_{x\pi}T_b).
\end{aligned} \tag{27}$$

Moreover, let $t_{opt} = \operatorname{argmin}_t f(t)$. Then, the logarithm $\log_{X_a}(R_z(t_{opt})R_{x\pi}X_b)$ is a horizontal vector in $\mathcal{H}_{X_a}\overline{\mathcal{M}}_{\mathcal{T}}$.

Proof. By differentiating the functions f_i we obtain

$$\dot{f}_i(t) = -\frac{\theta_i(t)}{2\sin\theta_i(t)}\operatorname{tr}(R_{ai}^T \widehat{e_z} R_z(t) R_{x\pi} R_{bi}) \tag{28}$$

$$= \frac{1}{2}\operatorname{tr}\left((\widehat{e_z}R_{ai})^T \log_{R_{ai}}(R_z(t)R_{x\pi}R_{bi})\right) \tag{29}$$

for $i \in \{1, 2, 3\}$ and for $i = 4$ we have

$$\dot{f}_4(t) = -\frac{\theta_4(t)}{\sin\theta_4(t)}\operatorname{tr}(T_a^T \hat{e}_z R_z(t) R_{x\pi} T_b) \tag{30}$$

$$= \operatorname{tr}\left((\widehat{e_z}T_a)^T \log_{T_a}(R_z(t)R_{x\pi}T_b)\right) \tag{31}$$

Comparing the condition $\dot{f}(t_{\text{opt}}) = \sum_{i=1}^{4} \dot{f}_i(t_{\text{opt}}) = 0$ with the basis of $\mathcal{V}_{X_a}$ in Proposition 5.3 and the definition of $\mathcal{H}_{X_a}$, we deduce that the logarithm $\log_{X_a}\left(R_z(t_{\text{opt}})R_{x\pi}X_b\right)$ must be an horizontal vector.

At this point, we will describe an algorithm for computing the logarithm map. At a high level, we need to optimize globally $f(t)$ twice, once for each of the two choices for $R_{x\pi}$ and choose the solution that gives the lower cost. For simplicity, we describe here only the optimization for the component that corresponds to $R_{x\pi} = I_3$ (the optimization for the other component is analogous).

Although global optimization is generally hard, we can exploit the special structure of f to efficiently compute its global minimizer t_{opt}. First, the cost function f is continuous and 2π-periodic, since $R_z(t + 2\pi) = R_z(t)$, but it is not everywhere smooth. For the first three terms in $f(t)$, the derivative $\dot{f}_i(t)$, $i \in \{1, 2, 3\}$, is not defined when $\cos\theta_i(t) = -1$. This correspond to discontinuity points t_{di}, which can be computed in closed form (see [29, Proposition 5.6]). It can also be shown [29] that $f_i(t)$, $i \in \{1, 2, 3\}$ is convex between discontinuity points.

It remains to analyze the behavior of $f_4(t)$. We have two distinct cases:

Case 1: $T_a = R_z(t_0)T_b$ for some $t_0 \in \mathbb{R}$. In this special case $f_4(t)$ is simply given by $f_4(t) = \frac{1}{2}(\arccos(\cos(t - t_0)))^2$. The derivative $\dot{f}_4(t)$ is not defined for $t = t_0 + (2k + 1)\pi$, $k \in \mathbb{Z}$ and $\ddot{f}_4(t) = 1$ when defined. So, in this case f_4 is piecewise convex and thus, f is also piecewise convex. The four points of discontinuity of $\dot{f}$ can be computed in closed form and thus, projected Newton method [2] can be applied to each of four resulting intervals in a way similar to [29]. Then, the global optimum of $f(t)$ can be computed as the minimum of the four local minima.

Case 2: $T_a \neq R_z(t_0)T_b$ for all $t_0 \in \mathbb{R}$. This is the more general case. By differentiating $\dot{f}_4(t)$ we can obtain a closed form expression for $\ddot{f}_4$. Unfortunately, $\ddot{f}_4$ can take negative values and thus, $f_4(t)$ is not convex. However, each period can be divided into two intervals, one at which $f_4(t)$ is convex (thus easy to optimize) and one at which $f_4(t)$ is concave (for which we use a branch-and-bound search). First, we need to identify these intervals. Let

$$c_{14} = (T_b T_a^T)_{1,1} + (T_b T_a^T)_{2,2}, \tag{32}$$

$$c_{24} = (T_b T_a^T)_{1,2} - (T_b T_a^T)_{2,1}. \tag{33}$$

Then, $\dot{f}_4(t) = 0$ for $t = \arctan(c_{24}/c_{14})$. It can be immediately seen that $\dot{f}_4(t) = 0$ has two solutions: one corresponding to the minimum over a period and one corresponding to the maximum. Let t_{max} and t_{min} denote these two values. Since $\ddot{f}_4(t)$ is continuous, it follows that $\ddot{f}_4(t_{min}) > 0$ and $\ddot{f}_4(t_{max}) < 0$. As a consequence, f_4 is convex in an interval (t_{c1}, t_{c2}) around t_{min}, and concave in an interval $(t_{c2}, t_{c1} + 2\pi)$ around t_{max}. The values t_{c1} and t_{c2} can be computed from t_{max} and t_{min} using the bisection method for $\ddot{f}_4(t) = 0$. For the interval (t_{c1}, t_{c2}) we have that $f(t)$ is continuous, convex with up to three discontinuity points of the first derivative. Thus, a projected Newton method as in [29] can be again applied to each of the subintervals. For the interval $(t_{c2}, t_{c1} + 2\pi)$, $f(t)$ is generally neither convex nor concave, and we implement a branch-and-bound search. Since we already have a good initial guess from the interval where $f(t)$ is convex, most of the subintervals are quickly rejected. Moreover, a lower bound for f in an interval $[a, b]$, on which f_4 is concave, can be efficiently estimated by minimizing the following convex underestimate of $f(t)$ using the Newton method

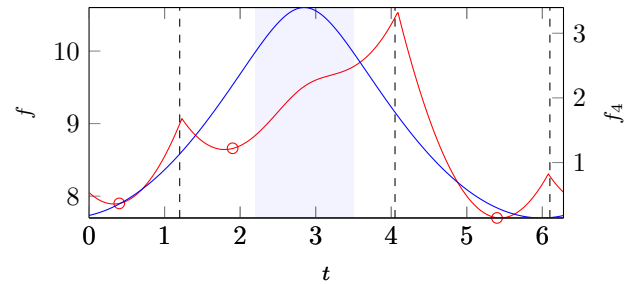

Figure 1: An instance of the cost $f(t)$ (in red) and $f_4(t)$ (in blue) for one period. Black dashed lines correspond to the three discontinuity points t_{di}. Shaded region corresponds to the interval on which f_4 is concave and the red circles to the local minimizers of $f(t)$.

$$f_l(t) = \sum_{i=1}^{3} f_i(t) + \frac{f_4(b) - f_4(a)}{b - a}(t - a) + f_4(a) \tag{34}$$

This underestimate is simply the sum of the three piecewise convex function f_1, f_2, f_3 with a linear underestimate of f_4.

In conclusion, using the above described method, the computation of the logarithm $\log_{[X_a]}[X_b]$ between two equivalence classes $[X_a], [X_b] \in \mathcal{M}_{\mathcal{T}}$ can be efficiently carried out. In our experiments, the average running time is approximately 5 ms with our preliminary Matlab implementation.

6. Optimization on the trifocal manifold

In this section, we describe how to minimize a cost function that takes as input a trifocal tensor. For example, assume we have n point-line-line correspondences $x_{1p} \leftrightarrow l_{2p} \leftrightarrow l_{3p}$, $p = 1, 2, \ldots, n$. Such a cost function is the sum of squared algebraic errors which is given by

$$f_a(\mathcal{T}) = \sum_{p=1}^{n} \sum_{i=1}^{3} \left((x_{1p})_i l_{2p}^T \mathcal{T}_i l_{3p}\right)^2. \tag{35}$$

Another example is the Sampson error [11, 15], *i.e.*, the first order approximation to the geometric error:

$$f_s(\mathcal{T}) = \sum_{p=1}^{n} \sum_{i=1}^{3} \frac{\left((x_{1p})_i l_{2p}^T \mathcal{T}_i \, l_{3p}\right)^2}{J_p(\mathcal{T}) J_p(\mathcal{T})^T} \tag{36}$$

where $J_p(\mathcal{T})$ is the Jacobian of the expression $(x_{1p})_i l_{2p}^T \mathcal{T}_i l_{3p}$ with respect to x_{1p}, l_{2p} and l_{3p} (note that this is a row vector).

More generally, given a real-valued function $f\colon \mathbb{R}^{3\times3\times3} \to \mathbb{R}$, let $f_{\overline{\mathcal{M}}_\mathcal{T}}\colon \overline{\mathcal{M}}_\mathcal{T} \to \mathbb{R}$ defined by $f_{\overline{\mathcal{M}}_\mathcal{T}} = f \circ \mathcal{T}$ where $\mathcal{T}$ is the map given in Definition 4.2. Moreover, let $X_a, X_b \in \overline{\mathcal{M}}_\mathcal{T}$ such that $X_a \sim X_b$. Since $f_{\overline{\mathcal{M}}_\mathcal{T}}$ is constant in each equivalence class, it induces a unique function $f_{\mathcal{M}_\mathcal{T}}$ on $\mathcal{M}_\mathcal{T}$ such that $f_{\overline{\mathcal{M}}_\mathcal{T}} = f_{\mathcal{M}_\mathcal{T}} \circ \pi$.

In order to combine the parametrization of $\mathcal{M}_\mathcal{T}$ given by the exponential map with the trust-region methods described in [1] we need to compute $\operatorname{grad} f_{\mathcal{M}_\mathcal{T}}([X])$ and $\operatorname{Hess} f_{\mathcal{M}_\mathcal{T}}([X])[\xi]$. Tangent vectors to quotient manifolds are represented in a computer program by their horizontal lifts. In detail, $\operatorname{grad} f_{\mathcal{M}_\mathcal{T}}([X])$ is represented by its horizontal lift $\operatorname{grad} f_{\overline{\mathcal{M}}_\mathcal{T}}(X)$ at X. Note that $\operatorname{grad} f_{\overline{\mathcal{M}}_\mathcal{T}}(X)$ is guaranteed to be an horizontal vector. Thus, we just have to compute the (Riemannian) gradient of $f_{\overline{\mathcal{M}}_\mathcal{T}}$. In the rest of this section, we show how to obtain the expression for the Riemannian gradient by its Euclidean counterpart.

Let $X(t)$ be a geodesic curve of the form $X(t) = (R_1(t), R_2(t), R_3(t), T(t))$. Let $\mathcal{T}(t) \doteq \mathcal{T}(X(t))$. The tangent of $\mathcal{T}(t)$ is given by

$$\begin{aligned}\dot{\mathcal{T}}_i = {} & R_2^T T_{12} e_i^T \dot{R}_1^T R_3 - R_2^T \dot{R}_1 e_i T_{13}^T R_3 + \dot{R}_2^T R_2 \mathcal{T}_i \\ & + \mathcal{T}_i R_3^T \dot{R}_3 + R_2^T \dot{T}_{12} e_i^T R_1^T R_3 - R_2^T R_1 e_i \dot{T}_{13}^T R_3\end{aligned} \tag{37}$$

for $i = 1, 2, 3$. Now, consider the function $f_{\overline{\mathcal{M}}_\mathcal{T}}(X(t)) = f(\mathcal{T}(t))$. At $t = 0$ we have

$$\overline{g}\big(\dot{X}, \operatorname{grad} f_{\overline{\mathcal{M}}_\mathcal{T}}(X)\big) = \langle \dot{\mathcal{T}}, \operatorname{grad} f(\mathcal{T}) \rangle \tag{38}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product. To alleviate the notation, let $G = \operatorname{grad} f(\mathcal{T}) \in \mathbb{R}^{3\times3\times3}$ and let G_i be i-th slice of G for $i = 1, 2, 3$. Then $\langle \dot{\mathcal{T}}, G \rangle = \sum_{i=1}^3 \operatorname{tr}\big(G_i^T \dot{\mathcal{T}}_i\big)$ and

$$\begin{aligned}\operatorname{tr}\big(G_i^T \dot{\mathcal{T}}_i\big) = {} & \operatorname{tr}\big(\dot{R}_1^T (R_3 G_i^T R_2^T T_{12} e_i^T - R_2 G_i R_3^T T_{13} e_i^T)\big) \\ & + \operatorname{tr}\big(\dot{R}_2^T R_2 \mathcal{T}_i G_i^T\big) + \operatorname{tr}\big(\dot{R}_3^T R_3 \mathcal{T}_i^T G_i\big) \\ & + \operatorname{tr}\big(\dot{T}_{12}^T R_2 G_i R_3^T R_1 e_i\big) - \operatorname{tr}\big(\dot{T}_{13}^T R_3 G_i^T R_2^T R_1 e_i\big)\end{aligned}$$

Since the manifold $\overline{\mathcal{M}}_\mathcal{T}$ is a submanifold of a Euclidean space, it follows that any Euclidean vector $\zeta \in T_X \overline{\mathcal{M}}_\mathcal{T}$ can be uniquely orthogonally decomposed as $\zeta = \zeta' + \zeta''$ with $\zeta' \in T_X \overline{\mathcal{M}}_\mathcal{T}$ and $\zeta'' \in (T_X \overline{\mathcal{M}}_\mathcal{T})^\perp$. Hence $\langle \xi, \zeta \rangle = \langle \xi, \zeta' \rangle$ for any $\xi \in T_X \overline{\mathcal{M}}_\mathcal{T}$. Using this fact, we have that $\operatorname{grad} f_{\overline{\mathcal{M}}_\mathcal{T}}(X) = (\xi_1, \xi_2, \xi_3, \xi_4)$, where

$$\begin{aligned}\xi_1 &= \textstyle\sum_{i=1}^3 P_{R_1}\left(\left(R_3 G_i^T R_2^T T_{12} - R_1^T R_2 G_i R_3^T T_{13}\right) e_i^T\right) \\ \xi_2 &= \textstyle\sum_{i=1}^3 P_{R_2}\left(R_2 \mathcal{T}_i G_i^T\right) \qquad \xi_3 = \textstyle\sum_{i=1}^3 P_{R_3}\left(R_3 \mathcal{T}_i^T G_i\right) \\ \xi_4 &= \textstyle\sum_{i=1}^3 P_T\left(R_2 G_i R_3^T R_1 e_i, -R_3 G_i^T R_2^T R_1 e_i\right)\end{aligned}$$

where $P_R \xi = R \operatorname{skew}(R^T \xi)$, skew denotes the skew-symmetric part of a matrix, and $P_X \xi = \xi - X \operatorname{tr}(X^T \xi)$ denotes the orthogonal projection of a vector $\zeta \in \mathbb{R}^{3\times2}$ onto the tangent space $T_T \mathbb{S}_2^3$. It is possible to also compute the Riemannian Hessian of $f_{\overline{\mathcal{M}}_\mathcal{T}}(X)$, which encodes second-order information about the function. However, its derivation is rather lengthy, and we refer the reader to [1] for details.

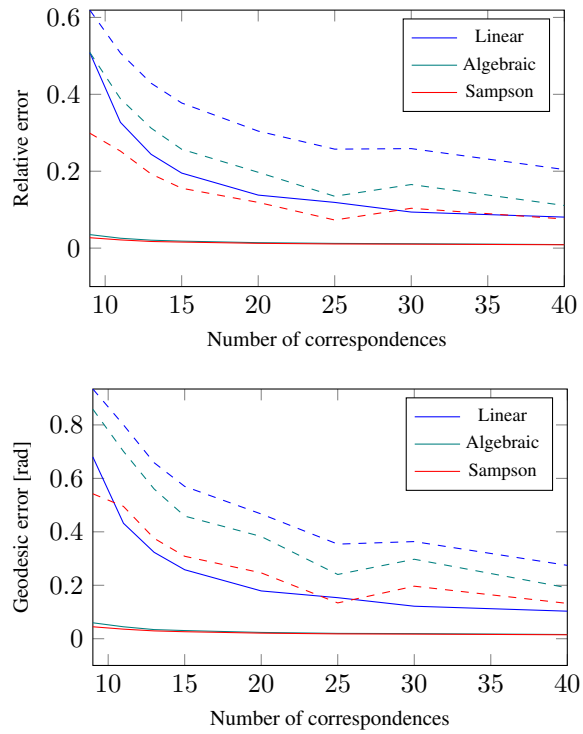

Figure 2: Relative (top) and geodesic (in $SO(3)^3 \times \mathbb{S}^5$, bottom) mean (dashed) and median (solid) errors before and after non-linear minimization of either the algebraic or Sampson cost.

We evaluate our implementation on the fountain-P11 dataset from [26] which includes the ground-truth camera poses. We extract SIFT features [30] to obtain point correspondences across different image triplets. We keep only image triplets with more than 50 point correspondences. To obtain an initial estimate of the trifocal tensor, we use the linear algorithm described in [32]. For optimization on the signed trifocal manifold, we use the trust-region solver of [3]. As error metric, we use the relative error between the estimated tensor and the ground-truth in the Frobenius norm sense. We also use the geodesic distance of $SO(3)^3 \times \mathbb{S}^5$ since the trifocal tensor can be parametrized by three camera orientations and two relative translations of unit total length. To compare with the ground-truth, we align the estimated rotations and relative translations using orthogonal Procrustes analysis. We vary the number of point correspondences from 9 to 40. For each image triplet and each number of correspondences, we repeat our experiment 50 times by randomly

selecting a different set of correspondences each time.

The results are presented in Figure 2. We compare the solution of the linear algorithm [32] with the solutions obtained by minimizing the algebraic (35) and the Sampson (36) errors on the trifocal manifold. The proposed optimization significantly outperforms the linear algorithm and produces very accurate estimates even with only few correspondences.

7. Pose averaging and the Weiszfeld algorithm

The Weiszfeld algorithm has been traditionally used for computing the l_1-mean (geometric median) of a set of points in $\mathbb{R}^n$. Recently, Hartley *et al.* [8] proposed the use of the Weiszfeld algorithm for the purpose of rotation averaging under the l_1-norm. Instead of using RANSAC for outlier rejection, they obtain multiple estimates of the relative rotations from the corresponding essential matrices, and average them using the Weiszfeld algorithm on $SO(3)$. This idea has been extended in [29] for averaging essential matrices.

A generic form of the Weiszfeld algorithm for an arbitrary Riemannian manifold $\mathcal{M}$ is presented in Algorithm 1. The new iterate $x(t+1)$ is obtained by taking the exponential map of a weighted average of directions on the tangent space of the current iterate. The weights are inversely proportional to the geodesic distance between the current iterate and each sample. Intuitively, points away from the current estimate have little impact on the update and thus, the algorithm is robust to outliers.

In this work, we use the Weiszfeld algorithm to average estimates of trifocal tensors seen as points on the signed trifocal manifold $\mathcal{M}_{\mathcal{T}}$. As in [29], the initial estimate is chosen as the midpoint of the two points having the lower cost. The sample trifocal tensors were obtained by the linear algorithm described in [32]. Unfortunately, this method does not perform well for a small number of point correspondences, resulting in noisy samples. In our experiments we observed that the algorithm converged in 10 to 15 iterations. We compare our approach with RANSAC and with the Weizsfeld algorithm on the manifold $SO(3)^2 \times \mathbb{S}^5$, *i.e.* a manifold parametrization of the trifocal tensor with $R_1 = I_3$. The purpose of this experiment is to show the advantage of the quotient versus non-quotient parametrization when using distances between trifocal tensors. We vary the number of samples from 10 to 50. The results are shown in Figure 3. 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 like RANSAC. Also, the Weiszfeld algorithm performs much better on the quotient manifold, as anticipated.

Algorithm 1 Weiszfeld averaging

Require: Points $x_1, x_2, \ldots, x_n \in \mathcal{M}$, initial estimate $x(0)$

1: **for** $t = 0, \ldots, N$ **do**

2: $\quad w_i(t) = d(x(t), x_i)^{-1}$

3: $\quad x(t+1) = \exp_x \left(\frac{\sum_i w_i(t) \log_x(x_i)}{\sum_i w_i(t)} \right)$

4: **end for**

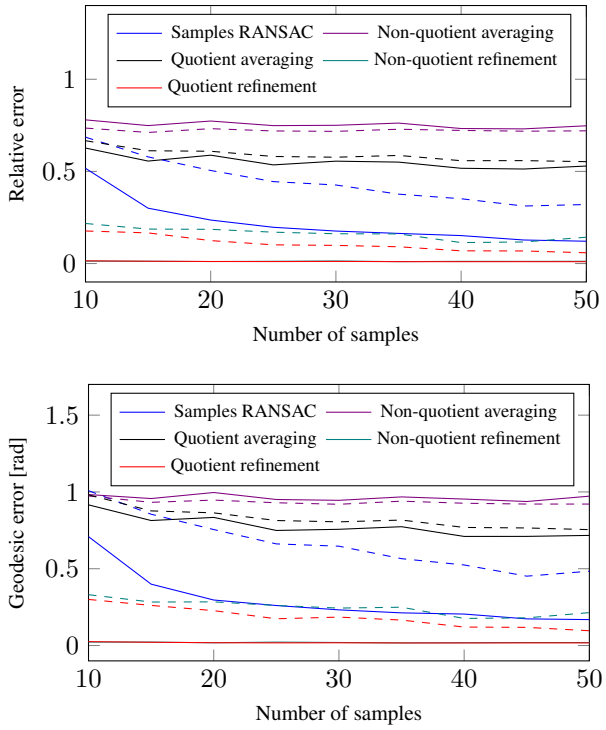

Figure 3: Relative (top) and geodesic (bottom) mean (dashed) and median (solid) errors for the Weiszfeld (quotient and non-quotient parametrizations) and RANSAC algoritms, without and with non-linear minimization.

8. Conclusions and future work

In this work, we investigated a novel parametrization of the trifocal tensor for calibrated cameras with non-colinear pinholes obtained from a quotient Riemannian manifold. We incorporated techniques for optimization on manifolds and pose averaging in our approach and we showed that the resulting distance is meaningful. In future work, we will focus on a more general parametrization that encompasses both degenerate and non-degenerate cases.

Acknowledgments: The authors are grateful for support through the following grants: NSF-DGE-0966142, NSF-IIS-1317788, NSF-IIP-1439681, NSF-IIS-1426840, ARL MAST-CTA W911NF-08-2-0004, ARL RCTA W911NF-10-2-0016, and ONR N000141310778.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. 1, 5, 7

[2] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1999. 6

[3] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014. 7

[4] N. Canterakis. A minimal set of constraints for the trifocal tensor. In *European Conference on Computer Vision*, pages 84–99, 2000. 1

[5] M. do Carmo. *Riemannian Geometry*. Mathematics (Boston, Mass.). Birkhäuser, 1992. 1

[6] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353. 5

[7] C. Geyer and K. Daniilidis. Mirrors in motion: Epipolar geometry and motion estimation. In *International Conference on Computer Vision*, pages 766–773, 2003. 1

[8] R. Hartley, K. Aftab, and J. Trumpf. L1 rotation averaging using the weiszfeld algorithm. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3041–3048. IEEE, 2011. 8

[9] R. I. Hartley. Projective reconstruction from line correspondences. In *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 903–907, 1994. 1

[10] R. I. Hartley. Lines and points in three views and the trifocal tensor. *Int. J. Comput. Vision*, 22(2):125–140, Mar. 1997. 1

[11] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 2, 6

[12] U. Helmke, K. Hüper, P. Y. Lee, and J. Moore. Essential matrix estimation using gauss-newton iterations on a manifold. *International Journal of Computer Vision*, 74(2):117–136, 2007.

[13] D. G. Kendall. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, 16:81–121, 1984. 2

[14] J. M. Lee. Introduction to smooth manifolds. 2000. 1

[15] Y. Ma, J. Koseck, and S. Sastry. Optimization criteria and geometric algorithms for motion and structure estimation. *International Journal of Computer Vision*, 44(3):219–249, 2001. 1, 6

[16] K. Nordberg. The key to three-view geometry. *Int. J. Comput. Vision*, 94(3):282–294, Sept. 2011. 1

[17] B. O'Neill. The fundamental equations of a submersion. *The Michigan Mathematical Journal*, 13(4):459–469, 12 1966. 5

[18] T. Papadopoulo and O. Faugeras. A new characterization of the trifocal tensor. In *European Conference on Computer Vision*, pages 109–123, 1998. 1

[19] P. Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics. Springer, 2006. 1

[20] J. Ponce and M. Hebert. Trinocular geometry revisited. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 1

[21] J. Ponce, K. McHenry, T. Papadopoulo, M. Teillaud, and B. Triggs. On the absolute quadratic complex and its application to autocalibration. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 780–787. IEEE, 2005. 1

[22] C. Ressl. A minimal set of constraints and a minimal parameterization for the trifocal tensor. In *International Archives of Photogrammetry and Remote Sensing*, pages 277–282, Volume XXXIV / 3A, 2002. talk: Symposium der ISPRS-Comm. III, Graz; 2002-09-09 – 2002-09-13. 1

[23] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:779–789, 1994. 1

[24] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *Automatic Control, IEEE Transactions on*, 41(3):393–413, 1996. 1

[25] M. E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990. 1

[26] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 7

[27] R. Subbarao, Y. Genc, and P. Meer. Robust unambiguous parametrization of the essential manifold. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 1

[28] P. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997. 1

[29] R. Tron and K. Daniilidis. On the quotient representation for the essential manifold. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1574–1581, 2014. 1, 5, 6, 8

[30] A. Vedaldi and B. Fulkerson. VLFeat - an open and portable library of computer vision algorithms. In *ACM International Conference on Multimedia*, 2010. 7

[31] R. Vidal, Y. Ma, S. Hsu, and S. Sastry. Optimal motion estimation from multiview normalized epipolar constraint. In *ICCV*, pages 34–41, 2001. 1

[32] J. Weng, T. S. Huang, and N. Ahuja. Motion and structure from line correspondences; closed-form solution, uniqueness, and optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):318–336, 1992. 1, 3, 7, 8