A minimal parameterization of the trifocal tensor

Klas Nordberg
Computer Vision Laboratory
Linköping University
klas@isy.liu.se

Abstract

The paper describes a minimal set of 18 parameters that can represent any trifocal tensor consistent with the internal constraints. 9 parameters describe three orthogonal matrices and 9 parameters describe 10 elements of a sparse tensor $\tilde{\mathcal{T}}$ with 17 elements in well-defined positions equal to zero. Any valid trifocal tensor is then given as some specific $\tilde{\mathcal{T}}$ transformed by the orthogonal matrices in the respective image domain. The paper also describes a simple approach for estimating the three orthogonal matrices in the case of a general $3 \times 3 \times 3$ tensor, i.e., when the internal constraints are not satisfied. This can be used to accomplish a least squares approximation of a general tensor to a tensor that satisfies the internal constraints. This type of constraint enforcement, in turn, can be used to obtain an improved estimate of the trifocal tensor based on the normalized linear algorithm, with the constraint enforcement as a final step. This makes the algorithm more similar to the corresponding algorithm for estimation of the fundamental matrix. An experiment on synthetic data shows that the constraint enforcement of the trifocal tensor produces a significantly better result than without enforcement, expressed by the positions of the epipoles, given that the constraint enforcement is made in normalized image coordinates.

1. Notation and concepts

Some explanation of the notation used in the paper may be needed. Superscripts are used on matrices to denote multiple matrix multiplications, e.g., $\mathbf{A}^2 = \mathbf{A}\mathbf{A}$. Equality between elements of projective spaces, i.e., vector equality up to a multiplication by a non-zero scalar, is denoted by $\sim$. A *fiber* is a generalization of a row or column in matrices to higher order tensors, i.e., a fiber is a "one-dimensional" sub-array. Einstein's summation rule is applied; summation is implied over the appropriate range for all indices that are repeated twice in an expression. For an introduction to tensor notation in terms of covariant and contravariant indices, see Section 15.2 and Appendix 1 in [4].

In this presentation, we will use tensors to represent multi-linear mappings on vector spaces, in this particular case on $\mathbb{R}^3$ which embeds homogeneous coordinates in three images. This means that a third order tensor, such as the trifocal tensor $\mathcal{T}$, can be linearly combined with a triplet of $\mathbb{R}^3$ vectors $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ to produce a real number. This triplet is here denoted $\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$, which represents the outer or Kronecker product of all combinations of elements in the three vectors. This product is also a third order tensor, and the multi-linear mapping of $\mathcal{T}$ with the three vectors is equal to reshaping both $\mathcal{T}$ and $\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$ to 27-dimensional vectors and then take their inner product. This notation generalize to the case when $\mathcal{T}$ is applied to a triplet of 3×3 matrices $\mathbf{U}, \mathbf{V}, \mathbf{W}$. The matrix triplet is here represented as $\mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W}$, the outer or Kronecker product between the matrices, corresponding to a 27×27 matrix. This matrix represents the mapping of each of the three $\mathbb{R}^3$ spaces, before their elements are applied to the tensor $\mathcal{T}$, as one single object.

2. Introduction

The trifocal tensor is a well explored object in computer vision where it can be used, for example, to determine correspondences between points and lines in three views [9]. The standard derivation of the trifocal tensor [4] is based on the multi-view pinhole camera model: $\mathbf{y}_k \sim \mathbf{C}_k \, \mathbf{x}$ where $\mathbf{x}$ and $\mathbf{y}_k$ are the homogeneous coordinates of a 3D point and its projection onto view k, and $\mathbf{C}_k$ is the camera projection matrix for view k. Assuming a singular value decomposition of camera matrix 1, $\mathbf{C}_1 = \mathbf{L} \, (\, \mathbf{S} \mid \mathbf{0} \,) \, \mathbf{H}^T$, where $\mathbf{L}$ is a 3×3 orthogonal matrix, $\mathbf{H}$ is a 4×4 orthogonal matrix, $(\, \mathbf{S} \mid \mathbf{0} \,)$ is a 3×4 matrix with $\mathbf{S}$ a diagonal matrix that holds the singular values, the three camera matrices can be normalized in accordance to

$$\mathbf{C}'_1 = \mathbf{S}^{-1} \, \mathbf{L}^T \, \mathbf{C}_1 \, \mathbf{H} = (\, \mathbf{I} \mid \mathbf{0} \,), \tag{1}$$

$$\mathbf{C}'_2 = \mathbf{C}_2 \, \mathbf{H} = (\, \mathbf{A} \mid \mathbf{a}_4 \,), \tag{2}$$

$$\mathbf{C}'_3 = \mathbf{C}_3 \, \mathbf{H} = (\, \mathbf{B} \mid \mathbf{b}_4 \,). \tag{3}$$

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

This corresponds to a common coordinate transformation of the 3D space represented by $\mathbf{H}$ and a transformation of the 2D coordinate system in view 1 represented by $\mathbf{S}^{-1}\mathbf{L}^T$. $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{H}$ define the 3×3 matrices $\mathbf{A}, \mathbf{B}$ and vectors $\mathbf{a}_4, \mathbf{b}_4 \in \mathbb{R}^3$ that appear in Equations (2) and (3). We will use a prime sign ' to denote objects (vectors, matrices, tensors) transformed to these coordinate systems in the first image and the 3D space.

The trifocal tensor $\mathcal{T}'$ related to cameras $\mathbf{C}'_1, \mathbf{C}'_2, \mathbf{C}'_3$ can be defined in terms of the three matrices

$$\mathcal{T}' = \{\mathbf{T}'_1, \mathbf{T}'_2, \mathbf{T}'_3\} \quad \text{where} \quad \mathbf{T}'_i = \mathbf{a}_i \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_i^T \tag{4}$$

and it defines an incidence relation between a point $\mathbf{x}'_1$ in view 1 and lines $\mathbf{l}_2, \mathbf{l}_3$ in views 2 and 3 in accordance to

$$\mathcal{T}' \, (\mathbf{x}'_1 \otimes \mathbf{l}_2 \otimes \mathbf{l}_3) = (x'_1)^i \, (\mathbf{l}_2^T \mathbf{T}'_i \, \mathbf{l}_3) = 0 \tag{5}$$

when the corresponding point and lines intersect in 3D space. To obtain the tensor corresponding to the original cameras we only need to compensate for the transformation $\mathbf{D} = \mathbf{S}^{-1}\mathbf{L}^T$ in view 1, the tensor is invariant to the common transformation $\mathbf{H}$ of the 3D space. We have $\mathbf{x}' = \mathbf{D}\mathbf{x}$ and the above incidence relation can be rewritten as

$$\begin{aligned} \mathcal{T}' \, ((\mathbf{D}\mathbf{x}) \otimes \mathbf{l}_2 \otimes \mathbf{l}_3) &= D^i_j (x_1)^j \, (\mathbf{l}_2^T \mathbf{T}'_i \, \mathbf{l}_3) = \\ &= (x_1)^j \, (\mathbf{l}_2^T [\mathbf{T}'_i \, D^i_j] \, \mathbf{l}_3) = 0. \end{aligned} \tag{6}$$

This implies that

$$\mathcal{T} = \{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3\} \quad \text{where} \quad \mathbf{T}_j = \mathbf{T}'_i D^i_j \tag{7}$$

and leads to

$$\mathcal{T} \, (\mathbf{x}_1 \otimes \mathbf{l}_2 \otimes \mathbf{l}_3) = x_{1i} \, (\mathbf{l}_2^T \mathbf{T}_i \, \mathbf{l}_3) = 0 \tag{8}$$

when the point $\mathbf{x}_1$ and the lines $\mathbf{l}_1, \mathbf{l}_2$ correspond to a 3D point and 3D lines that intersect. Consequently, (7) defines the trifocal tensor related to camera matrices $\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3$. The relation between $\mathcal{T}$ and $\mathcal{T}'$ is

$$\mathcal{T}' = \mathcal{T} \, (\mathbf{D} \otimes \mathbf{I} \otimes \mathbf{I}) \tag{9}$$

where $\mathbf{I}$ is the 3×3 identity matrix. $\mathcal{T}$ or $\mathcal{T}'$ can be represented either as three 3×3 matrices, as $3 \times 3 \times 3$ arrays of numbers, or as 27-dimensional vectors.

The above presentation shows how $\mathcal{T}$ can be computed from three general camera matrices, but not any set of 27 real numbers constitutes a valid trifocal tensor. A trifocal tensor must satisfy a set of *internal constraints* that guarantee that it can be related to some set of three camera matrices. Disregarding a multiplication by a scalar, each $\mathbf{C}_k$ has 11 degrees of freedom, in total 33 degrees of freedom for three cameras. The trifocal tensor is invariant to a common transformation $\mathbf{H}$ of the 3D coordinates. There are 15 degrees of freedom for $\mathbf{H}$ which leaves $33 - 15 = 18$ degrees of freedom for $\mathcal{T}$. Since $\mathcal{T}$ has 27 elements it must satisfy a set of 8 internal constraints to account for the 18 degrees of freedom plus the overall scalar multiplication.

The existence of internal constraints is important when the trifocal tensor is estimated from real and noisy data. A straight-forward approach for determining the tensor from point or line correspondences can be based on the direct linear transformation method that describes the tensor as the singular vector of a $N \times 27$ matrix corresponding to the smallest (at best zero valued) singular value, see Algorithm 16.1 in [4]. In the case of noisy data, however, the resulting trifocal tensor does not satisfy the internal constraints. Consequently, it is not consistent to a particular set of cameras and may not be expected to describe three-view point and line correspondences correctly.

Papadopolo and Faugeras [6] presented the first set of internal constraints for $\mathcal{T}$. It consists of 12 constraints, and includes some dependencies, although in a non-trivial way. Canterakis [1] presented a minimal set of 8 internal constraints that are both necessary and sufficient for a set of 27 numbers to constitute a valid trifocal tensor. Both [6] and [1] are based on extracting a set of matrices or polynomials from the tensor and require that certain properties such as the rank of the matrices or number of distinct roots of the polynomials meet certain criteria for a valid trifocal tensor.

Both [6] and [1] have some disadvantages. First, if a potential trifocal tensor does not meet the constraints, there is no straight-forward way to enforce the constraints. This problem is treated by Ressl [8] where iterative adjustments of the tensor are derived from the Gauss-Helmert model of the constraints. Second, even if a constraint enforcement method can be devised, there is no explicit metric for the constraints which makes it possible to determine the closest proper trifocal tensor to an arbitrary tensor.

2.1. This paper

In this paper the internal constraints are addressed indirectly by proving the existence of a minimal parameterization of $\mathcal{T}$. This set has 9 parameters that determine three orthogonal matrices and 9 parameters that determine a sparse representation $\tilde{\mathcal{T}}$ after $\mathcal{T}$ has been transformed by the orthogonal matrices. $\tilde{\mathcal{T}}$ consists of 17 elements in well-defined positions that are zero and the remaining elements can vary freely. The transformations are well-defined in the case that the camera matrices are known, but they can be estimated from the tensor itself in the practical case when it has been perturbed by noise. The transformations are then estimated by solving a non-linear optimization problem using standard numerical methods. This approach is similar to the way the normalized 8-point algorithm deals with internal constraints for the fundamental matrix $\mathbf{F}$; after a transformation of $\mathbf{F}$ certain elements of the resulting matrix should vanish. In that case, the transformations are based on the singular value decomposition. In the case of the trifocal tensor, however, the resulting transformations does not seem to be possible to relate to higher order variants of SVD.

2.2. The cross product operator

Many of the results derived in the following section are based on basic properties of the vector cross product. This operation can conveniently be represented in terms of the *cross product operator*, $[\mathbf{a}]_\times$, a 3×3 anti-symmetric matrix such that $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b}$. A set of useful equations related to this operator are presented below, and will be used later in Section 3.

From $\mathbf{a} \times \mathbf{b} \sim \mathbf{b} \times \mathbf{a}$ follows

$$[\mathbf{a}]_\times \mathbf{b} \sim [\mathbf{b}]_\times \mathbf{a} \quad \text{or} \quad \mathbf{b}^T [\mathbf{a}]_\times \sim \mathbf{a}^T [\mathbf{b}]_\times. \tag{10}$$

The cross product between a vector and itself vanishes:

$$[\mathbf{a}]_\times \mathbf{a} = \mathbf{0} \quad \text{or} \quad \mathbf{a}^T [\mathbf{a}]_\times = \mathbf{0}^T. \tag{11}$$

For all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, $\mathbf{a} \times \mathbf{b}$ is perpendicular to $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{a}^T [\mathbf{b}]_\times \mathbf{a} = 0. \tag{12}$$

The anti-symmetry of $[\mathbf{a}]_\times$ allows us to write $[\mathbf{a}]_\times^T \sim [\mathbf{a}]_\times$ and, combined with Equations (10) and (12), it leads to

$$\mathbf{a}^T [\mathbf{b}]_\times [\mathbf{b}]_\times [\mathbf{a}]_\times \mathbf{b} = \mathbf{b}^T [\mathbf{a}]_\times^T [\mathbf{b}]_\times [\mathbf{a}]_\times \mathbf{b} = 0. \tag{13}$$

The third power of an anti-symmetric matrix is a scalar times the matrix:

$$[\mathbf{a}]_\times^3 = -\|\mathbf{a}\|^2 [\mathbf{a}]_\times \sim [\mathbf{a}]_\times. \tag{14}$$

If $\mathbf{A}$ is a full rank 3×3 matrix we get

$$\mathbf{A}^T [\mathbf{a}]_\times \mathbf{A} = \det \mathbf{A} \cdot [\mathbf{A}^{-1}\mathbf{a}]_\times \sim [\mathbf{A}^{-1}\mathbf{a}]_\times \tag{15}$$

which can be derived by applying (11) together with algebraic properties of the inverse of a 3×3 matrix.

3. $\mathcal{T}'$ transformed to a new basis

Define three transformation matrices

$$\begin{aligned} \mathbf{U}_0 &= (\mathbf{A}^{-1}\mathbf{a}_4, [\mathbf{A}^{-1}\mathbf{a}_4]_\times^2 \mathbf{B}^{-1}\mathbf{b}_4, [\mathbf{A}^{-1}\mathbf{a}_4]_\times \mathbf{B}^{-1}\mathbf{b}_4), \\ \mathbf{V}_0 &= (\mathbf{a}_4, [\mathbf{a}_4]_\times \mathbf{A}\mathbf{B}^{-1}\mathbf{b}_4, [\mathbf{a}_4]_\times^2 \mathbf{A}\mathbf{B}^{-1}\mathbf{b}_4), \\ \mathbf{W}_0 &= (\mathbf{b}_4, [\mathbf{b}_4]_\times \mathbf{B}\mathbf{A}^{-1}\mathbf{a}_4, [\mathbf{b}_4]_\times^2 \mathbf{B}\mathbf{A}^{-1}\mathbf{a}_4). \end{aligned} \tag{16}$$

To simplify the expressions in the following derivations, the last two matrices can also be written

$$\mathbf{V}_0 = (\mathbf{a}_4, [\mathbf{a}_4]_\times \mathbf{r}, [\mathbf{a}_4]_\times \mathbf{r}'), \tag{17}$$

$$\mathbf{W}_0 = (\mathbf{b}_4, [\mathbf{b}_4]_\times \mathbf{s}, [\mathbf{b}_4]_\times \mathbf{s}'), \tag{18}$$

for $\mathbf{r} = \mathbf{A}\mathbf{B}^{-1}\mathbf{b}_4$, $\mathbf{r}' = [\mathbf{a}_4]_\times \mathbf{r}$ and $\mathbf{s} = \mathbf{B}\mathbf{A}^{-1}\mathbf{a}_4$, $\mathbf{s}' = [\mathbf{b}_4]_\times \mathbf{s}$. The three matrices $\mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_0$ have orthogonal columns, which can be verified based on the properties of the cross product operator presented in Section 2.2. This means that a suitable scaling of each column of these three matrices produce three orthogonal matrices:

$$\begin{aligned} &\mathbf{U} = \mathbf{U}_0(\mathbf{U}_0^T \mathbf{U}_0)^{-\frac{1}{2}}, \quad \mathbf{V} = \mathbf{V}_0(\mathbf{V}_0^T \mathbf{V}_0)^{-\frac{1}{2}}, \\ &\mathbf{W} = \mathbf{W}_0(\mathbf{W}_0^T \mathbf{W}_0)^{-\frac{1}{2}}. \end{aligned} \tag{19}$$

We will now transform $\mathcal{T}'$ in accordance to

$$\tilde{\mathcal{T}'}_i^{jk} = \mathcal{T'}_m^{pq} U^m_{\ i} V_p^{\ j} W_q^{\ k}. \tag{20}$$

This implies that each element of $\tilde{\mathcal{T}}'$ is given by multiplying a triplet of columns from each of $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$ onto the elements of $\mathcal{T}'$. In the following, we will use a tilde sign to denote objects defined for the transformed image coordinates given by transformations $\mathbf{U}, \mathbf{V}, \mathbf{W}$ (the prime sign still means that we have made this transformation in the 3D and 2D coordinates given by the cameras in Equations (1) to (3), this will be taken care of in Section 4). Next, we will show that certain of these triplets always result in zero when $\mathcal{T}'$ is formed in accordance to (4). More precisely, we will show that certain pairs of columns from $\mathbf{U}, \mathbf{V}$ or $\mathbf{W}$ give a zero valued fiber when multiplied onto $\mathcal{T}'$. The fact that fibers, corresponding to 3-dimensional vectors, vanish after multiplication by two of the three orthogonal matrices means that the result is still zero after multiplication by any column in the third matrix. To simplify the derivations, we can both disregard the scaling factors introduced in (19) to make $\mathbf{U}, \mathbf{V}, \mathbf{W}$ orthogonal (since they only scale each column in each of the matrices) and also disregard the final multiplication by the third matrix (since it only multiplies an orthogonal matrix onto a zero vector). We will use $\overset{\circ}{=}$ to denote equality under these assumptions. $\mathbf{v}_k$ and $\mathbf{w}_k$ are columns of $\mathbf{V}$ or $\mathbf{W}$, respectively, and $s_i = \mathbf{b}_i^T [\mathbf{b}_4]_\times \mathbf{s}$, $r_i = \mathbf{a}_i^T [\mathbf{a}_4]_\times \mathbf{r}$ and correspondingly for primed variables. Here we go!

$$\tilde{\mathcal{T}'}_i^{22} \overset{\circ}{=} \mathbf{v}_2^T \mathbf{T}'_i \mathbf{w}_2 = \mathbf{r}^T [\mathbf{a}_4]_\times^T (\mathbf{a}_i \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_i^T)[\mathbf{b}_4]_\times \mathbf{s} = \overset{\text{eq (11)}}{=} r_i \cdot 0 - 0 \cdot s_i = 0_i \tag{21}$$

$$\tilde{\mathcal{T}'}_i^{23} \overset{\circ}{=} \mathbf{v}_2^T \mathbf{T}'_i \mathbf{w}_3 = \mathbf{r}^T [\mathbf{a}_4]_\times^T (\mathbf{a}_i \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_i^T)[\mathbf{b}_4]_\times \mathbf{s}' = \overset{\text{eq (11)}}{=} r_i \cdot 0 - 0 \cdot s'_i = 0_i \tag{22}$$

$$\tilde{\mathcal{T}'}_i^{32} \overset{\circ}{=} \mathbf{v}_3^T \mathbf{T}'_i \mathbf{w}_2 = \mathbf{r'}^T [\mathbf{a}_4]_\times^T (\mathbf{a}_i \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_i^T)[\mathbf{b}_4]_\times \mathbf{s} = \overset{\text{eq (11)}}{=} r'_i \cdot 0 - 0 \cdot s_i = 0_i \tag{23}$$

$$
\begin{aligned}
\tilde{\mathcal{T}}'^{33}_{i} &\overset{\circ}{=} \mathbf{v}_3^T \mathbf{T}'_i \mathbf{w}_3 = \mathbf{r}'^T [\mathbf{a}_4]_\times^T (\mathbf{a}_i \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_i^T)[\mathbf{b}_4]_\times \mathbf{s}' = \\
&\overset{\text{eq }(11)}{=} r'_i \cdot 0 - 0 \cdot s'_i = 0_i
\end{aligned} \tag{24}
$$

$$
\begin{aligned}
\tilde{\mathcal{T}}'^{j2}_{1} &\overset{\circ}{=} U^m{}_1 \mathbf{T}'_m \mathbf{w}_2 = U^m{}_1 (\mathbf{a}_m \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_m^T)[\mathbf{b}_4]_\times \mathbf{s} \sim \\
&\overset{\text{eq }(11)}{\sim} (\mathbf{A}^{-1}\mathbf{a}_4)_m \mathbf{a}_4 \mathbf{b}_m^T [\mathbf{b}_4]_\times \mathbf{s} = \\
&= \mathbf{a}_4 \mathbf{a}_4^T \mathbf{A}^{-T} \mathbf{B}^T [\mathbf{b}_4]_\times \mathbf{B} \mathbf{A}^{-1} \mathbf{a}_4 \overset{\text{eq }(12)}{=} \mathbf{a}_4 \cdot 0 = \mathbf{0}
\end{aligned} \tag{25}
$$

$$
\begin{aligned}
\tilde{\mathcal{T}}'^{j2}_{2} &\overset{\circ}{=} U^m{}_2 \mathbf{T}'_m \mathbf{w}_2 = U^m{}_2 (\mathbf{a}_m \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_m^T)[\mathbf{b}_4]_\times \mathbf{s} \sim \\
&\sim ([\mathbf{A}^{-1}\mathbf{a}_4]_\times^2 \mathbf{B}^{-1}\mathbf{b}_4)_m \mathbf{a}_4 \mathbf{b}_m^T [\mathbf{b}_4]_\times \mathbf{s} = \\
&= \mathbf{a}_4 \mathbf{b}_4^T \mathbf{B}^{-T} [\mathbf{A}^{-1}\mathbf{a}_4]_\times^2 \mathbf{B}^T [\mathbf{b}_4]_\times \mathbf{B} \mathbf{A}^{-1} \mathbf{a}_4 \sim \\
&\overset{\text{eq }(15)}{\sim} \mathbf{a}_4 \mathbf{b}_4^T \mathbf{B}^{-T} [\mathbf{A}^{-1}\mathbf{a}_4]_\times^2 [\mathbf{B}^{-T}\mathbf{b}_4]_\times \mathbf{A}^{-1} \mathbf{a}_4 = \\
&\overset{\text{eq }(13)}{=} \mathbf{a}_4 \cdot 0 = \mathbf{0}
\end{aligned} \tag{26}
$$

$$
\begin{aligned}
\tilde{\mathcal{T}}'^{2k}_{1} &\overset{\circ}{=} \mathbf{v}_2^T \mathbf{T}'_m U^m{}_1 = \mathbf{r}^T [\mathbf{a}_4]_\times^T (\mathbf{a}_m \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_m^T) U^m{}_1 \sim \\
&\overset{\text{eq }(11)}{\sim} \mathbf{r}^T [\mathbf{a}_4]_\times \mathbf{a}_m \mathbf{b}_4^T (\mathbf{A}^{-1}\mathbf{a}_4)_m = \\
&= \mathbf{r}^T [\mathbf{a}_4]_\times \mathbf{A}\mathbf{A}^{-1} \mathbf{a}_4 \mathbf{b}_4^T = \mathbf{r}^T [\mathbf{a}_4]_\times \mathbf{a}_4 \mathbf{b}_4^T \overset{\text{eq }(11)}{=} \mathbf{0}^T
\end{aligned} \tag{27}
$$

$$
\begin{aligned}
\tilde{\mathcal{T}}'^{3k}_{1} &\overset{\circ}{=} \mathbf{v}_3^T \mathbf{T}'_m U^m{}_1 = \mathbf{r}'^T [\mathbf{a}_4]_\times^T (\mathbf{a}_m \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_m^T) U^m{}_1 = \\
&= \text{same as eq (27)} = \mathbf{0}^T
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
\tilde{\mathcal{T}}'^{2k}_{2} &\overset{\circ}{=} \mathbf{v}_2^T \mathbf{T}'_m U^m{}_3 = \mathbf{r}^T [\mathbf{a}_4]_\times^T (\mathbf{a}_m \mathbf{b}_4^T - \mathbf{a}_4 \mathbf{b}_m^T) U^m{}_3 \sim \\
&\overset{\text{eq }(11)}{\sim} \mathbf{r}^T [\mathbf{a}_4]_\times \mathbf{a}_m \mathbf{b}_4^T ([\mathbf{A}^{-1}\mathbf{a}_4]_\times^2 \mathbf{B}^{-1}\mathbf{b}_4)_m = \\
&= \mathbf{b}_4 \mathbf{B}^{-T} \mathbf{A}^T [\mathbf{a}_4]_\times \mathbf{A} [\mathbf{A}^{-1}\mathbf{a}_4]_\times^2 \mathbf{B}^{-1} \mathbf{b}_4 \mathbf{b}_4^T = \\
&\overset{\text{eq }(15)}{=} \mathbf{b}_4 \mathbf{B}^{-T} [\mathbf{A}^{-1}\mathbf{a}_4]_\times^3 \mathbf{B}^{-1} \mathbf{b}_4 \mathbf{b}_4^T \sim \\
&\overset{\text{eq }(14)}{\sim} \mathbf{b}_4 \mathbf{B}^{-T} [\mathbf{A}^{-1}\mathbf{a}_4]_\times \mathbf{B}^{-1} \mathbf{b}_4 \mathbf{b}_4^T \sim \\
&\overset{\text{eq }(10)}{\sim} \mathbf{a}_4 \mathbf{A}^{-T} [\mathbf{B}^{-1}\mathbf{b}_4]_\times \mathbf{B}^{-1} \mathbf{b}_4 \mathbf{b}_4^T \overset{\text{eq }(11)}{=} \mathbf{0}^T
\end{aligned} \tag{29}
$$

The transformed trifocal tensor $\tilde{\mathcal{T}}'$ can, in the same way as the original tensor be described as a set of 3×3 matrices:

$$
\tilde{\mathcal{T}}' = \{\tilde{\mathbf{T}}'_1, \tilde{\mathbf{T}}'_2, \tilde{\mathbf{T}}'_3\}. \tag{30}
$$

Based on Equations (21) to (29) these three matrices can now be characterized as

$$
\tilde{\mathbf{T}}'_1 = \begin{pmatrix} \times & 0 & \times \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \tilde{\mathbf{T}}'_2 = \begin{pmatrix} \times & 0 & \times \\ 0 & 0 & 0 \\ \times & 0 & 0 \end{pmatrix},
$$
$$
\tilde{\mathbf{T}}'_3 = \begin{pmatrix} \times & \times & \times \\ \times & 0 & 0 \\ \times & 0 & 0 \end{pmatrix}, \tag{31}
$$

where the $\times$-signs represent some numbers that can be determined from $\mathcal{T}$. The important observation is that 17 of the elements in $\tilde{\mathcal{T}}$ are zero. We summarize this result as:

Lemma 1 *There exist orthogonal matrices $\mathbf{U}, \mathbf{V}, \mathbf{W}$ that transform the trifocal tensor $\mathcal{T}'$ to the sparse form presented in (31).*

4. Generalization to un-normalized cameras

Lemma 1 is derived under the assumption that the camera matrices have been transformed such that the camera matrices are described by Equations (1) to (3) and it remains to show that this result is valid also in the general case. Equation (20) can be written as

$$
\tilde{\mathcal{T}}' = \mathcal{T}' (\mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W}) \tag{32}
$$

which combined with (9) gives

$$
\tilde{\mathcal{T}}' = \mathcal{T} (\mathbf{DU} \otimes \mathbf{V} \otimes \mathbf{W}). \tag{33}
$$

This means that if $\mathcal{T}$, the trifocal tensor related to the original cameras, is transformed by $(\mathbf{DU})$, $\mathbf{V}$, and $\mathbf{W}$ the result is the sparse form $\tilde{\mathcal{T}}'$, described in (31). This set of transformations, however, includes $\mathbf{DU}$ which in general is not orthogonal. The missing piece is to compensate for this non-orthogonality.

Based on the so-called QR-factorization [3], $\mathbf{DU}$ can be expressed as

$$
\mathbf{DU} = \mathbf{QR} \tag{34}
$$

where $\mathbf{R}$ is an upper triangular matrix and $\mathbf{Q}$ is orthogonal. Inserted into (33) this gives

$$
\tilde{\mathcal{T}}' = \mathcal{T} (\mathbf{Q} \otimes \mathbf{V} \otimes \mathbf{W}) (\mathbf{R} \otimes \mathbf{I} \otimes \mathbf{I}) \tag{35}
$$

or

$$
\tilde{\mathcal{T}} = \tilde{\mathcal{T}}' (\mathbf{R}^{-1} \otimes \mathbf{I} \otimes \mathbf{I}) = \mathcal{T} (\mathbf{Q} \otimes \mathbf{V} \otimes \mathbf{W}). \tag{36}
$$

As previously, $\tilde{\mathcal{T}}$ can be represented as three matrices

$$
\tilde{\mathcal{T}} = \{\tilde{\mathbf{T}}_1, \tilde{\mathbf{T}}_2, \tilde{\mathbf{T}}_3\} \tag{37}
$$

and (36) states that these matrices are given by linear combinations with the matrices $\tilde{\mathbf{T}}'_k$ and the elements in $\mathbf{R}^{-1}$. $\mathbf{R}$ is upper triangular and this property is inherited by its inverse $\mathbf{R}^{-1}$. *Thus*, $\tilde{\mathbf{T}}_1$ is only a scalar multiplication of $\tilde{\mathbf{T}}'_1$, $\tilde{\mathbf{T}}_2$ is a linear combination of $\tilde{\mathbf{T}}'_1$ and $\tilde{\mathbf{T}}'_2$, and $\tilde{\mathbf{T}}_3$ is a linear combination of $\tilde{\mathbf{T}}'_1$, $\tilde{\mathbf{T}}'_2$ and $\tilde{\mathbf{T}}'_3$. *Thus*, the matrices $\tilde{\mathbf{T}}_k$ has the same structure as the matrices $\tilde{\mathbf{T}}'_k$ in (31). We summarize this result as:

Lemma 2 *There exist orthogonal matrices $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ that transform the tensor $\mathcal{T}$ to the sparse form presented in (31).*

By inverting (36) into

$$\mathcal{T} = \tilde{\mathcal{T}}(\mathbf{Q}^T \otimes \mathbf{V}^T \otimes \mathbf{W}^T) \quad (38)$$

the above result can also be formulated as

Lemma 3 *$\mathcal{T}$ can be parameterized by means of 3 orthogonal matrices, in total 9 parameters, and the elements marked by $\times$ in the three matrices $\tilde{\mathbf{T}}_k$, also 9 parameters since a common multiplication by a scalar can be disregarded. In total this gives 18 parameters for $\mathcal{T}$, which is equal to its stipulated number of degrees of freedom.*

5. Determining the orthogonal transformations

The result derived in the previous section only says that there exist three orthogonal transformations that reduce $\mathcal{T}$ to the sparse form in (31), and if the tensor is consistent with a known set of camera matrices then the transformations can be computed. In practice, however, it is neither the case that the tensor is exactly consistent with some cameras nor are the cameras known. This situation typically arises when $\mathcal{T}$ is estimated from noisy data.

On the other hand, the description of the internal constraints presented here allows us to search for the three orthogonal transformation matrices $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ that produce the sparse form in (31). In the practical case we cannot expect that the 17 elements marked as “0” vanish completely. Instead we can try to determine $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ that minimize the norm of these elements. For example, this can be done by measuring the sum of squares of these 17 elements and minimize this over the transformation matrices. More formally, the trifocal tensor lives in a 27-dimensional real vector space T and there is a 17-dimensional subspace $Z \subset T$ such that $\tilde{T} \perp Z$ for the transformed tensor $\tilde{\mathcal{T}}$. Let $\mathbf{P}_Z$ be the projection operator from Z to $\mathbb{R}^{17}$. The problem to be solved is then to find orthogonal $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ such that

$$\varepsilon = \|\mathbf{P}_Z[\mathcal{T}(\mathbf{Q} \otimes \mathbf{V} \otimes \mathbf{W})]\|_2 \quad (39)$$

is minimized. If $\mathcal{T}$ satisfies the internal constraints this minimum is $\varepsilon = 0$, otherwise $\varepsilon > 0$ to a degree determined by the perturbation in $\mathcal{T}$. In the rest of this section, we will consider how to determine the transformations related to a tensor $\mathcal{T}_0$ that may not meet the internal constraints.

This minimization problem can be solved based on standard minimization methods for non-linear multi-variable functions. However, we must ensure that $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ are orthogonal which means that they need to be suitably parameterized, for example, based on the matrix exponential function of anti-symmetric matrices or on quaternions [2]. In either case, ε becomes a non-linear and non-convex function in the free parameters. This means that the minimization must be based on iterative methods which, in turn, are highly dependent on good initial estimates of the parameters. Fortunately, it is possible to determine a set of camera matrices that are consistent with $\mathcal{T}_0$, at least approximately since $\mathcal{T}_0$ is an approximation of a valid tensor. This can be done, for example, based on Algorithm 15.1 in [4]. The camera matrices that are retrieved by this algorithm are precisely of the form described in Equations (1) to (3), with the difference that the retrieved cameras refer to $\mathcal{T}$ while $\mathbf{C}'_1, \mathbf{C}'_2, \mathbf{C}'_3$ refer to $\mathcal{T}'$, (9). This observation means that if we use the retrieved camera matrices to determine $\mathbf{A}, \mathbf{B}, \mathbf{a}_4, \mathbf{b}_4$, Equations (2) and (3), from which $\mathbf{U}, \mathbf{V}, \mathbf{W}$ are given by (16)eq:defW0 and (19), then $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ with $\mathbf{Q} = \mathbf{U}$ transform $\mathcal{T}_0$ approximately to the sparse form described in (31). There is one issue, however, that needs to be solved if $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ are computed in this way. Algorithm 15.1 always produces a matrix $\mathbf{B}$ of rank 2, i.e., $\mathbf{B}^{-1}$ is not defined in (16). Since $\mathcal{T}$ is invariant to projective transformations of the 3D space, this problem can be solved by multiplying the retrieved camera matrices from right by the 4×4 matrix

$$\mathbf{H}' = \left(\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \mathbf{h}^T & 1 \end{array}\right). \quad (40)$$

The resulting cameras still conform to Equations (1) to (3), but $\mathbf{h} \in \mathbb{R}^3$ is chosen such that $\mathbf{A}$ and $\mathbf{B}$ have full rank. With this modification of the camera matrices, $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ can be computed according to the above description.

When an initial set of matrices $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ has been determined, any suitable non-linear minimization procedure can be used to find the optimal transformations that minimize ε. As long as $\mathcal{T}_0$ is not perturbed by too much noise, i.e., too far from satisfying the internal constraints, this minimization should find the global minimum.

6. Enforcement of the internal constraints

In Sections 3 and 4 we have seen that the trifocal tensor can be reduced to a sparse form, (31), by a set of orthogonal transformations, and Section 5 describes a method for determining these transformations, even in the case when $\mathcal{T}_0$ is perturbed by noise. An interesting application that combines these two results is trying to determine the closest approximation of $\mathcal{T}_0$ that is found in the set of tensors that meet the constraints. If we choose the 2-norm for this approximation, the solution is given directly by the optimal transformations $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ described in Section 5.

We want to find $\mathcal{T}$ in the set of tensors that satisfy the internal constraints such that ε' is minimized, where

$$\varepsilon'^2 = \|\mathcal{T}_0 - \mathcal{T}\|^2 = ([\mathcal{T}_0]_i^{jk} - \mathcal{T}_i^{jk})([\mathcal{T}_0]_i^{jk} - \mathcal{T}_i^{jk}). \quad (41)$$

Since $\mathcal{T}$ satisfies the internal constraints, Lemma 3 states that it can be written $\mathcal{T} = \tilde{\mathcal{T}}(\mathbf{Q}^T \otimes \mathbf{V}^T \otimes \mathbf{W}^T)$ for orthogonal matrices $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ which gives

$$\begin{aligned}\varepsilon'^2 &= ([\mathcal{T}_0]_i^{jk} - \tilde{\mathcal{T}}_l^{pq} Q_i^l V_p^j W_q^k)([\mathcal{T}_0]_i^{jk} - \tilde{\mathcal{T}}_r^{st} Q_i^r V_s^j W_t^k) \\ &= ([\mathcal{T}_0]_i^{jk} Q_i^l V_p^j W_q^k - \tilde{\mathcal{T}}_l^{pq})([\mathcal{T}_0]_i^{jk} Q_i^r V_s^j W_t^k - \tilde{\mathcal{T}}_r^{st}) \\ &= \|\mathcal{T}_0(\mathbf{Q} \otimes \mathbf{V} \otimes \mathbf{W}) - \tilde{\mathcal{T}}\|^2. \end{aligned} \quad (42)$$

Consequently, to find $\mathcal{T}$ is the same as finding $\tilde{\mathcal{T}} \in Z$ and the three orthogonal transformations $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ that minimize ε'. This, in turn, is the same as finding $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ that minimize ε, (39), and then choose $\tilde{\mathcal{T}}$ as the transformed $\mathcal{T}_0$ projected onto $Z^\perp$, i.e., as the elements marked with $\times$ in (31). The resulting $\mathcal{T}$, (38), is then the best 2-norm approximation of $\mathcal{T}_0$ in the set of valid trifocal tensors. This type of constraint enforcement is very similar to the approach used for enforcing the internal constraint $\det \mathbf{F} = 0$ for the fundamental matrix $\mathbf{F}$ in the normalized 8-point algorithm [5]. The main difference is that here the decomposition of the trifocal tensor is not based on standard SVD and does not seem to be related to any higher order variants thereof.

7. Experimental evaluation

To demonstrate the potential of the constraint enforcement described in Section 6, an experiment on synthetic data is presented here. A cube with side 0.4 units contains N points that are uniformly distributed within the cube, see Figure 1. A circle has its center located 1 unit from the center of the cube and three cameras are located with equal spacing on the circle, with their optical axes intersecting with the box center. Each camera projects the cube onto a 512×512 pixels image such that it fills the images almost entirely (focal distance = 800 units). The image coordinates of the points are perturbed by Gaussian noise, $\sigma = 1$. From the set of N corresponding image points, $\mathcal{T}_0$ is estimated using the normalized linear algorithm [4]. The internal constraints are enforced on $\mathcal{T}_0$ using the method described in Section 6, both when $\mathcal{T}_0$ is represented in standard coordinates and when it is represented in normalized image image coordinates. The two approaches differ since the metric used for the least squares approximation of $\mathcal{T}$ differ. The resulting tensors are denoted $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. From each of $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2$, a corresponding set of epipoles $\mathbf{e}_{21}, \mathbf{e}_{31}$ are retrieved using Algorithm 15.1 in [4]. Due to the symmetry of the problem, only epipole $\mathbf{e}_{21}$ is considered here.

1000 observations each for $N = 7, 10, 15, 20$ are used to estimate the error in terms of the Euclidean distance from the correct $\mathbf{e}_{21}$ to its position estimated from $\mathcal{T}_0$, $\mathcal{T}_1$, and $\mathcal{T}_2$. This distance contains "outliers" where the estimated epipole is very far from the true position, for example if the retrieved cameras or fundamental matrices are close to being degenerate. Such errors occur independently for all three types of estimated tensors. To get a better understanding of the errors, values that are above a threshold of 100 are removed and the statistics is computed from what remains.

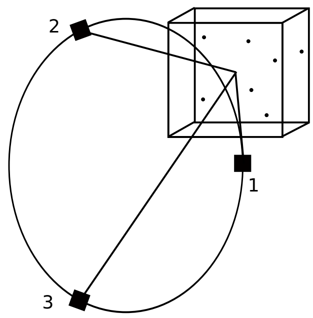

Figure 1. Three cameras equally spaced on a unit circle observe random 3D points within a box with 0.4 units sides. The circle center is one unit from the box center.

The result is presented in Table 1. It shows that the epipole based on enforcement of the internal constraints in normalized coordinates gives fewer outliers than the other two cases and also that the mean error within the inliers is lower. This behavior is similar to what happens when the fundamental matrix is estimated directly from noisy image data; constraint enforcement of the fundamental matrix in normalized coordinates gives an improved estimate [4]. With the threshold used here, the epipole estimated by enforcing the internal constraints in standard coordinates produce too much outliers to be of practical use.

N	$\mathcal{T}_0$	$\mathcal{T}_1$	$\mathcal{T}_2$
7	50 (34%)	56 (1%)	49 (38%)
10	43 (77%)	52 (1%)	40 (80%)
15	33 (95%)	52 (3%)	30 (97%)
20	25 (99%)	54 (5%)	23 (99%)
50	13 (100%)	45 (15%)	12 (100%)

Table 1. Mean distances in pixels between the true epipole $\mathbf{e}_{21}$ and the epipoles estimated from $\mathcal{T}_0$ (no constraint enforcement), $\mathcal{T}_1$ (constraints are enforced in standard coordinates), and from $\mathcal{T}_2$ (constraints are enforced in normalized coordinates). Distances above 100 are considered as outliers and are not included, the percentage of inliers are presented in parentheses.

8. Summary and discussion

The paper presents a minimal parameterization of the trifocal tensor. 9 parameters are used to define three orthogonal matrices $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ and 9 parameters are used to define a sparse tensor $\tilde{\mathcal{T}}$ described by (31) (since the tensor is an element of a projective space, only 9 parameters are needed to specify the 10 elements marked by $\times$). These 18 parameters are sufficient to represent any trifocal tensor that satisfies the internal constraints. It is also shown how $\mathbf{Q}, \mathbf{V}, \mathbf{W}$ can be determined from an arbitrary tensor $\mathcal{T}_0$ that may not satisfy the internal constraints. We can apply these transformations to $\mathcal{T}_0$ and set to zero the elements that should vanish for a correctly determined trifocal tensor, and finally transform back to the original coordinates by means of $\mathbf{Q}^T, \mathbf{V}^T, \mathbf{W}^T$. The result is a 2-norm approximation of $\mathcal{T}_0$ in the set of valid trifocal tensors.

The constraint enforcement can be used as a final step in the normalized linear algorithm for computation of the trifocal tensor [4]. This makes the algorithm similar to the standard normalized linear algorithm for computation of the fundamental matrix, where the internal constraint $\det \mathbf{F} = 0$ is enforced. Experimental data suggests that the constraint enforcement produces a better approximation of the correct tensor when the enforcement has been made in normalized image coordinates [5]. This constraint enforced tensor should therefore be a good initial solution for more sophisticated estimation procedures of the trifocal tensor such as the Gold Standard method [4].

Parameterizations of three-view geometry in general and the trifocal tensor in particular, including minimal ones, have been presented in the literature previously, for examples see [4, 10]. A minimal parameterization also appears in [6]. In summary, the minimal parameterizations that are described in the literature are relatively complex from an algebraic point of view, for example, each set of parameters may correspond to multiple solutions of the three-view geometry (i.e., of the trifocal tensor), or implies non-linear constraints on over-parameterized representations. In contrast, the minimal parameterization proposed here only requires three orthogonal 3×3 matrices, for which several standard parameterizations are known (e.g., matrix exponential of anti-symmetric matrices, quaternions, Rodrigues formula, see Section A.4.3 in [4]), together with the 10 non-zero elements of $\tilde{\mathcal{T}}$ with an arbitrary non-zero scaling which, again, can be straight-forward parameterized in several ways (e.g., setting one of the 10 elements $= 1$). To summarize, the proposed parameterization means that using a minimal parameterization of the trifocal tensor can be an *easy* task, see Section 16.4.2 in [4].

9. Acknowledgement

This work has been funded by the Swedish Research Foundation (SSF) within the VISCOS project. Thanks also to Berkant Savas at the Department of Mathematics for fruitful discussions on the presented tensor decomposition.

References

[1] N. Canterakis. A minimal set of constraints for the trifocal tensor. In *Proceedings of 6th European Conference on Computer Vision*, pages 84–99, Dublin, Ireland, 2000. LNSC 1842. 2

[2] O. Faugeras. *Three-dimensional Computer Vision*. The MIT Press, 1993. 5

[3] G. H. Golub and C. G. V. Loan. *Matrix Computations, 2nd ed*. Johns Hopkins University Press, 1989. 4

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision, 2nd ed*. Cambridge University Press, 2003. 1, 2, 5, 6, 7

[5] R. I. Hartley. In defence of the 8-point algorithm. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 19(6):580–593, 1997. 6, 7

[6] T. Papadopoulo and O. Faugeras. A new characterization of the trifocal tensor. In *Proceedings of 5th European Conference on Computer Vision*, pages 109–123, Freiburg, Germany, 1998. LNSC 1406. 2, 7

[7] L. Quan. Invariant of six points and projective reconstruction from three uncalibrated images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(1):34–46, 1995.

[8] C. Ressl. *Geometry, constraints, and computation of the trifocal tensor*. PhD thesis, Technical University, Wien, 2003. 2

[9] A. Shashua. Algebraic functions for recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995. 1

[10] P. H. S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997. 7