Algebraic Characterization of Essential Matrices and Their Averaging in Multiview Settings

Supplementary Material

Yoni Kasten* Amnon Geifman* Meirav Galun Ronen Basri
Weizmann Institute of Science
{yoni.kasten,amnon.geifman,meirav.galun,ronen.basri}@weizmann.ac.il

1. Counter example

In the introduction of the paper we claim that a consistent n-view fundamental matrix whose 3×3 blocks form essential matrices does not necessarily form a consistent n-view essential matrix. We justify this argument by a constructing a counter example for the case $n=3$.

First note that the following observation is true. If $\mathbf{t}_2, \mathbf{t}_3, \mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ and $R_2, R_3 \in SO(3)$ and we set

$$V_1 = I_{3\times 3}, V_2 = R_2^T, V_3 = R_3^T + \mathbf{a}\mathbf{t}_3^T, \mathbf{t}_1 = 0_{3\times 1}$$

and

$$F_{ij} = V_i([\mathbf{t}_i - \mathbf{t}_j]_\times)V_j^T, \quad i,j = 1,2,3$$

then by construction

$$F = \begin{bmatrix} 0 & F & F_{13} \\ F_{12}^T & 0 & F_{23} \\ F_{13}^T & F_{23}^T & 0 \end{bmatrix}$$

is a consistent n-view fundamental matrix, unless the outer product $\mathbf{a}\mathbf{t}_3^T$ is such that it reduces the full rank of V_3 to 2.

Examine each of the block matrices F_{ij}:

$$F_{12} = V_1[\mathbf{t}_1 - \mathbf{t}_2]_\times V_2^T = [-\mathbf{t}_2]_\times R_2$$
$$F_{13} = V_1[\mathbf{t}_1 - \mathbf{t}_3]_\times V_3^T = [-\mathbf{t}_3]_\times (R_3 + \mathbf{t}_3\mathbf{a}^T) = [-\mathbf{t}_3]_\times R_3$$
$$F_{23} = V_2[\mathbf{t}_2 - \mathbf{t}_3]_\times V_3^T = R_2^T[\mathbf{t}_2 - \mathbf{t}_3]_\times (R_3 + \mathbf{t}_3\mathbf{a}^T)$$

It follows from this derivation that F_{12} and F_{13} are essential matrices. Next, we use the identity $[R\mathbf{t}]_\times = R[\mathbf{t}]_\times R^T$, which holds for any $R \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$, to show that for any $\mathbf{b} \in \mathbb{R}^3$

$$\begin{aligned} F_{23} &= [R_2^T(\mathbf{t}_2 - \mathbf{t}_3)]_\times R_2^T(R_3 + \mathbf{t}_3\mathbf{a}^T) \\ &= [R_2^T(\mathbf{t}_2 - \mathbf{t}_3)]_\times (R_2^T(R_3 + \mathbf{t}_3\mathbf{a}^T) + R_2^T(\mathbf{t}_2 - \mathbf{t}_3)\mathbf{b}^T) \\ &= [R_2^T(\mathbf{t}_2 - \mathbf{t}_3)]_\times R_2^T(R_3 + \mathbf{t}_3\mathbf{a}^T + (\mathbf{t}_2 - \mathbf{t}_3)\mathbf{b}^T). \end{aligned}$$

*Equal contributors

The trivial choice of $\mathbf{a} = \mathbf{b} = 0$ yields an essential matrix F_{23} which is consistent with the essential matrices F_{12} and F_{13}. Our aim is to make a specific choice of $R_2, R_3, \mathbf{t}_2, \mathbf{t}_3, \mathbf{a}, \mathbf{b}$ such that F_{23} will be an essential matrix that is inconsistent with the essential matrices F_{13} and F_{12}.

1.1. Choosing $R_2, R_3, \mathbf{t}_2, \mathbf{t}_3, \mathbf{a}, \mathbf{b}$

We first give some intuition for the way we set the values of $R_2, R_3, \mathbf{t}_2, \mathbf{t}_3, \mathbf{a}, \mathbf{b}$. We look at the term

$$R_2^T(R_3 + \mathbf{t}_3\mathbf{a}^T + (\mathbf{t}_2 - \mathbf{t}_3)\mathbf{b}^T) \tag{1}$$

and wish to set the values such that this term does not collapse to $R_2^T R_3$, but is still in the form of $R_2^T R^*$, for some $R^* \in SO(3)$. To that end, we look for two rotation matrices R^* and R^{**} such that $M = R^* - R^{**}$ has rank 2, and set the values such that the term in (1) will be equal to $R_2^T R^*$.

Following the construction of M, the SVD of M is of the form

$$M = U \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T$$

and we set

$$\mathbf{t}_3 = \mathbf{u}_1\sigma_1, \mathbf{a} = \mathbf{v}_1, \mathbf{b} = \mathbf{v}_2, \mathbf{t}_2 = \mathbf{t}_3 - \mathbf{u}_2\sigma_2. \tag{2}$$

Therefore,

$$R^* - R^{**} = M = \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \mathbf{u}_2\sigma_2\mathbf{v}_2^T = \mathbf{t}_3\mathbf{a}^T + (\mathbf{t}_3 - \mathbf{t}_2)\mathbf{b}^T.$$

Now we set R_2 to be some rotation matrix and $R_3 = R^{**}$. Clearly,

$$\mathbf{t}_3\mathbf{a}^T + (\mathbf{t}_3 - \mathbf{t}_2)\mathbf{b}^T = R_2 R_2^T R^* - R_3$$

which means

$$R_2^T R^* = R_2^T(R_3 + \mathbf{t}_3\mathbf{a}^T + (\mathbf{t}_3 - \mathbf{t}_2)\mathbf{b}^T),$$

yielding

$$F_{23} = [R_2^T(\mathbf{t}_2 - \mathbf{t}_3)]_\times R_2^T R^*,$$

which is an essential matrix that in general is inconsistent with F_{12}, F_{13}.

1.2. Technical details of the code provided for demonstrating a counter example

We first randomly sampled two rotation matrices R^*, R^{**} until we obtained $M = R^* - R^{**}$ of rank 2, and we set $R_3 = R^{**}$. Then, we sampled randomly a rotation matrix, and we set R_2. These selections are stored in a file called "counter_data.mat".

In the code we assign values to $\mathbf{t}_2, \mathbf{t}_3, R_3, \mathbf{a}, \mathbf{b}$, according to (2). We verify that indeed F_{12}, F_{13} and F_{23} are essential matrices. By construction, the 3-view fundamental matrix F is consistent.

In order to verify that the essential matrices are not consistent we extract the relative rotations from them. Each essential matrix defines two possible relative rotations. We evaluate the relation $R_{12}R_{23}R_{31}$ for each of the 8 choices of triplet of relative rotations, and verify that none of them closes a loop, i.e the following always holds

$$R_{12}R_{23}R_{31} \neq I.$$

An interesting observation. In addition, we verify in the code that for any choice of signs for the eigenvectors of F, X, Y, it turns out that $\sqrt{0.5}(X+Y)$ is indeed not a block rotation matrix. Interestingly, it holds that $\Sigma_+ = -\Sigma_-$, which shows that by itself this condition, in this case, is not sufficient for defining an appropriate consistent essential matrix. To conclude, requiring from a set of essential matrices to fulfill the sufficient conditions for consistent n-view fundamental matrix does not provide sufficient conditions for generating from this set, a consistent essential matrix. The counter example demonstrates the inconsistency of the essential matrices, as well as the violation of our conditions on the consistency of an n-view essential matrix.

2. Proof of Lemma 5

Below we prove Lemma 5 from the paper (which we rename here to be Lemma 1).

Lemma 1. *Let $E \in \mathbb{S}^{3n}$ of rank(6), and $\Sigma \in \mathbb{R}^{3\times3}$, a diagonal matrix, with positive elements on the diagonal. Let $X, Y, U, V \in \mathbb{R}^{3n\times3}$, and we define the mapping $(X, Y) \leftrightarrow (U, V) : X = \sqrt{0.5}(\hat{U} + \hat{V})$, $Y = \sqrt{0.5}(\hat{V} - \hat{U})$, $\hat{U} = \sqrt{0.5}(X - Y)$, $\hat{V} = \sqrt{0.5}(X + Y)$.*

Then, the (thin) SVD of E is of the form

$$E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$$

if and only if the (thin) spectral decomposition of E is of the form

$$E = [X, Y] \begin{pmatrix} \Sigma & \\ & -\Sigma \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}$$

Proof. $(\Rightarrow)$

$$\begin{aligned} E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix} &= \hat{U}\Sigma\hat{V}^T + \hat{V}\Sigma\hat{U}^T = \\ 0.5 \cdot (\hat{U} + \hat{V})\Sigma(\hat{U} + \hat{V})^T - 0.5 \cdot (\hat{V} - \hat{U})\Sigma(\hat{V} - \hat{U})^T &= \\ 0.5 \left[\hat{U} + \hat{V}, \hat{V} - \hat{U}\right] \begin{pmatrix} \Sigma & \\ & -\Sigma \end{pmatrix} \left[\hat{U} + \hat{V}, \hat{V} - \hat{U}\right]^T &= \\ [X, Y] \begin{pmatrix} \Sigma & \\ & -\Sigma \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix} \end{aligned} \tag{3}$$

where $X = \sqrt{0.5}(\hat{U} + \hat{V})$ and $Y = \sqrt{0.5}(\hat{V} - \hat{U})$. Since, $\begin{bmatrix} \hat{U}^T \\ \hat{V}^T \end{bmatrix} \left[\hat{U}, \hat{V}\right] = I_{6\times6}$, it yields $\begin{bmatrix} X^T \\ Y^T \end{bmatrix} [X, Y] = I_{6\times6}$, concluding that the last term in (3) is indeed (thin) spectral decomposition of E.

$(\Leftarrow)$

$$\begin{aligned} E = [X, Y] \begin{pmatrix} \Sigma & \\ & -\Sigma \end{pmatrix} [X, Y]^T &= X\Sigma X^T - Y\Sigma Y^T = \\ 0.5(X + Y)\Sigma(X - Y)^T + 0.5(X - Y)\Sigma(X + Y)^T &= \\ 0.5[X - Y, X + Y] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} [X + Y, X - Y]^T &= \\ [\hat{U}, \hat{V}] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix} \end{aligned} \tag{4}$$

where $\hat{U} = \sqrt{0.5}(X - Y)$ and $\hat{V} = \sqrt{0.5}(X + Y)$. The same argument for orthogonality works here, showing that indeed that last term in (4) is SVD of E. $\square$

3. Handling scaled rotations

Our optimization enforces the consistency of camera triplets, while allowing the essential matrices to be scaled arbitrarily. This is possible, because our theory can be generalized to handle scaled rotations. Below we generalize Definition 4 from the paper to allow essential matrices of the form $E_{ij} = \alpha_i R_i^T([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times)R_j\alpha_j$ and prove that the main theorem, i.e., Theorem 3 from the paper, holds for this generalization as well. This argument is then used to justify our treatment of camera triplets.

Definition 1. *An n-view essential matrix E is called* **scaled consistent** *if there exist n rotation matrices $\{R_i\}_{i=1}^n$, n vectors $\{\mathbf{t}_i\}_{i=1}^n$ and n non-zero scalars $\{\alpha_i\}_{i=1}^n$ such that $E_{ij} = \alpha_i R_i^T([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times)R_j\alpha_j$.*

The following theorem is a generalized version of Thm. 2 from the paper, and the derivations here are inspired by the derivations made in [19].

Theorem 2. *Let E be a scaled consistent n-view essential matrix, associated with scaled rotation matrices $\{\alpha_i R_i\}_{i=1}^n$, $\alpha_i \neq 0$ and camera centers $\{\mathbf{t}_i\}_{i=1}^n$. E satisfies the following conditions*

1. *E can be formulated as $E = A + A^T$ where $A = UV^T$ and $U, V \in \mathbb{R}^{3n\times 3}$*

$$V = \begin{bmatrix} \alpha_1 R_1^T \\ \vdots \\ \alpha_n R_n^T \end{bmatrix} \quad U = \begin{bmatrix} \alpha_1 R_1^T T_1 \\ \vdots \\ \alpha_n R_n^T T_n \end{bmatrix}$$

   with $T_i = [\mathbf{t}_i]_\times$ and w.l.o.g $\sum_{i=1}^n \alpha_i^2 \mathbf{t}_i = 0$.

2. *Each column of U is orthogonal to each column of V, i.e., $V^T U = 0_{3\times 3}$*

3. *rank(V)=3*

4. *If not all $\{\mathbf{t}_i\}_{i=1}^n$ are collinear, then rank(U) and rank(A) = 3. Moreover, if the (thin) SVD of A is $A = \hat{U}\Sigma\hat{V}^T$, with $\hat{U}, \hat{V} \in \mathbb{R}^{3n\times 3}$ and $\Sigma \in \mathbb{R}^{3\times 3}$ then the (thin) SVD of E is*

$$E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$$

   implying rank(E) = 6.

Proof. 1. The decomposition is a straightforward result from Def. 1. Moreover, any global translation of all the camera centers, will not change the values of the entries of E. In particular, if we denote the camera centers by $\{\tilde{\mathbf{t}}_i\}_{i=1}^n$ and they are translated to their new position $\mathbf{t}_i = \tilde{\mathbf{t}}_i - \frac{\sum \alpha_i^2 \tilde{\mathbf{t}}_i}{\sum \alpha_i^2}$, then $\sum \alpha_i^2 \mathbf{t}_i = 0$.

2. By the decomposition above, we have that

$$V^T U = \sum_{i=1}^{i=n} \alpha_i^2 R_i R_i^T T_i = [\sum_{i=1}^{i=n} \alpha_i^2 \mathbf{t}_i]_\times = 0$$

   which concludes that each column of U is orthogonal to each column of V.

3. $rank(V) = 3$ since each block of V is of rank 3 and V has 3 columns.

4. Assume by contradiction that $rank(U) < 3$. Then, $\exists \mathbf{t} \in \mathbb{R}^3, \mathbf{t} \neq 0$, s.t. $U\mathbf{t} = 0$. This implies that $\mathbf{t}_i \times \mathbf{t} = 0$ for all $i = 1, \ldots, n$. This implies that all the $\mathbf{t}_i$'s are parallel to $\mathbf{t}$, violating our assumption that not all camera locations are collinear. Consequently $rank(U) = 3$ and therefore also $rank(A) = 3$. Finally, let $A = \hat{U}\Sigma\hat{V}^T$ the SVD of A. Since $A = UV^T$ we get that

$$Span(U) = Span(\hat{U}), Span(V) = Span(\hat{V}).$$

   Then, since $E = A + A^T$, we get

$$E = \hat{U}\Sigma\hat{V}^T + \hat{V}\Sigma\hat{U}^T = \begin{bmatrix}\hat{U} & \hat{V}\end{bmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}.$$

   Following the result that the columns of U are orthogonal to those of V, it turns out that $\begin{bmatrix}\hat{U} & \hat{V}\end{bmatrix}$ is column orthogonal, concluding that the form above is the SVD of E, and $rank(E) = 6$.

□

Next, we show that Thm. 3 in the paper is also applicable with the generalized definition, i.e., Def. 1.

Theorem 3. *Let $E \in \mathbb{S}^{3n}$ be a consistent n-view fundamental matrix with a set of n cameras whose centers are not all collinear. We denote by $\Sigma_+, \Sigma_- \in \mathbb{R}^{3\times 3}$ the diagonal matrices with the 3 positive and 3 negative eigenvalues of E, respectively. The following conditions are equivalent:*

1. *E is a scaled consistent n-view essential matrix*

2. *The (thin) SVD of E can be written in the form*

$$E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma_+ & \\ & \Sigma_+ \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$$

   with $\hat{U}, \hat{V} \in \mathbb{R}^{3n\times 3}$ such that each 3×3 block of $\hat{V}$, $\hat{V}_i$, $i = 1, \ldots, n$, is a scaled rotation matrix, i.e., $\hat{V}_i = \hat{\alpha}_i \hat{R}_i$, where $\hat{R}_i \in SO(3)$ and $\hat{\alpha}_i \neq 0$. We say that $\hat{V}$ is a scaled block rotation matrix.

3. *$\Sigma_+ = -\Sigma_-$ and the (thin) spectral decomposition of E is of the form*

$$E = [X, Y] \begin{pmatrix} \Sigma_+ & \\ & \Sigma_- \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}$$

   such that $\sqrt{0.5}(X + Y)$ is a scaled block rotation matrix.

Proof. (1)⇒(2) Assume that E is a scaled consistent n-view essential matrix. Then, according to Thm. 2, $E = A + A^T$ with $A = UV^T$ and $U, V \in \mathbb{R}^{3n\times 3}$ with

$$V = \begin{bmatrix} \alpha_1 R_1^T \\ \vdots \\ \alpha_n R_n^T \end{bmatrix} \quad U = \begin{bmatrix} \alpha_1 R_1^T T_1 \\ \vdots \\ \alpha_n R_n^T T_n \end{bmatrix}$$

where $T_i = [\mathbf{t}_i]_\times$. Since $A = UV^T$ and $rank(A) = 3$, then $A^TA = VU^TUV$ and $A^TA \succeq 0$ with $rank(A^TA) = 3$ (A and A^TA share the same null space). First, we construct a spectral decomposition to A^TA, relying on the special properties of U and V. We have $rank(U) = 3$, and therefore U^TU, which is a 3×3, symmetric positive semi-definite matrix, is of full rank. Its spectral decomposition is of the form $U^TU = PDP^T$, where $P \in SO(3)$. (Spectral decomposition guarantees that $P \in O(3)$. However, P can be replaced by $-P$ if $\det(P) = -1$.) $D \in \mathbb{R}^{3\times3}$ is a diagonal matrix consisting of the (positive) eigenvalues of U^TU. This spectral decomposition yields the following decomposition

$$A^TA = VPDP^TV^T. \tag{5}$$

Now, note that

$$\begin{aligned} P^TV^TVP &= P^T \begin{bmatrix}\alpha_1R_1 & \dots & \alpha_nR_n\end{bmatrix} \begin{bmatrix}\alpha_1R_1^T \\ \vdots \\ \alpha_nR_n^T\end{bmatrix} P \\ &= P^T(\sum_{i=1}^n \alpha_i^2)I_{3\times3}P = (\sum_{i=1}^n \alpha_i^2)I_{3\times3}. \end{aligned}$$

Let $\alpha = \sum_{i=1}^n \alpha_i^2$. By a simple manipulation (5) becomes a spectral decomposition

$$A^TA = \frac{1}{\sqrt{\alpha}}VP(\alpha D)P^TV^T\frac{1}{\sqrt{\alpha}}. \tag{6}$$

On the other hand, the (thin) SVD of A is of the form $A = \hat{U}\Sigma\hat{V}^T$, where $\hat{U}, \hat{V} \in \mathbb{R}^{3n\times3}$, $\Sigma \in \mathbb{R}^{3\times3}$. This means that

$$A^TA = \hat{V}\Sigma^2\hat{V}^T. \tag{7}$$

Due to the uniqueness of the eigenvector decomposition, (6) and (7) collapse to the same eigenvector decomposition, up to some global rotation, $H \in SO(3)$, that is $\frac{1}{\sqrt{\alpha}}VP = \hat{V}H$, which means that

$$\hat{V}_i = \frac{\alpha_i}{\sqrt{\alpha}}R_i^TPH^T. \tag{8}$$

Since $R_i^T, P, H^T \in SO(3)$, then setting $\hat{\alpha}_i =: \frac{\alpha_i}{\sqrt{\alpha}}$ and $\hat{R}_i =: R_i^TPH^T \in SO(3)$, shows that $\hat{V}$ is a scaled block rotation matrix. Finally, by Thm. 2, the (thin) SVD of E is of the form

$$E = \begin{bmatrix}\hat{U}, \hat{V}\end{bmatrix} \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix} \tag{9}$$

and according to Lemma 1, the eigenvalues of E are Σ and $-\Sigma$. Since the elements on the diagonal of Σ are positive, and E is symmetric with exactly 3 positive eigenvalues Σ_+ and 3 negative eigenvalues Σ_-, it follows that $\Sigma = \Sigma_+$ and $-\Sigma = \Sigma_-$ concluding the proof.

(2)⇒(1) Let E be a consistent n-view fundamental matrix that satisfies condition (2). We would like to show that E is a scaled consistent n-view essential matrix. By condition (2) E can be written as

$$E = \hat{U}\Sigma_+\hat{V}^T + \hat{V}\Sigma_+\hat{U}^T = \bar{U}\hat{V}^T + \hat{V}\bar{U}^T \tag{10}$$

where $\bar{U} = \hat{U}\Sigma_+$ with $\hat{V}_i = \hat{\alpha}_i\hat{R}_i$, $\hat{R}_i \in SO(3)$. By definition $E_{ii} = 0$, and this implies that $\bar{U}_i\hat{V}_i^T$ is a skew symmetric matrix. Using Lemma 4 in the paper, $\bar{U}_i = \hat{V}_i\hat{T}_i$ for some skew symmetric matrix $\hat{T}_i = [\hat{\mathbf{t}}_i]_\times$. Plugging $\bar{U}_i$ and $\hat{V}_i$ in (10) yields

$$\begin{aligned} E_{ij} = \bar{U}_i\hat{V}_j^T + \hat{V}_i\bar{U}_j^T &= \hat{\alpha}_i\hat{\alpha}_j\hat{R}_i\hat{T}_i\hat{R}_j^T - \hat{\alpha}_i\hat{\alpha}_j\hat{R}_i\hat{T}_j\hat{R}_j^T \\ &= \hat{\alpha}_i{R_i}^T([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times)R_j\hat{\alpha}_j \end{aligned}$$

where $R_i = \hat{R}_i^T$, $\alpha_i = \hat{\alpha}_i$ and $\mathbf{t}_i = \hat{\mathbf{t}}_i$, concluding the proof. Finally, the derivation of the equivalence $(2) \Leftrightarrow (3)$ is exactly as in Thm. 3 in the paper.

Corollary 1. *Let E be a scaled consistent n-view matrix, then*

1. *The scale of each block $\hat{V}_i$ in the scaled block rotation matrix $\hat{V}$ can be calculated using by $\hat{\alpha}_i = (det(\hat{V}_i))^{\frac{1}{3}}$*
2. *Let E be a scaled consistent n-view essential matrix. Then, the transformation $diag(\frac{1}{\hat{\alpha_1}}I_{3\times3}, \dots, \frac{1}{\hat{\alpha_n}}I_{3\times3}) \cdot E \cdot diag(\frac{1}{\hat{\alpha_1}}I_{3\times3}, \dots, \frac{1}{\hat{\alpha_n}}I_{3\times3})$ transform E to be a consistent n-view essential matrix.*

Corollary 2. *A scaled consistent 3-view essential matrix is invariant to pairwise scaling.*

Proof. This proof is inspired by the proof of Corollary 2, presented in [15]. Let E be a scaled consistent 3-view essential matrix whose blocks are defined as $E_{ij} = \alpha_iR_i^T(T_i - T_j)R_j\alpha_j$, and let $\widetilde{E}$ be a 9×9 matrix whose blocks are defined to be $\widetilde{E}_{ij} = s_{ij}E_{ij}$ where $s_{ij} \neq 0$ are arbitrary pairwise scale factors. Without loss of generality we can assume that the number of negative scale factors is even (otherwise we can multiply the entire matrix by -1). Therefore, $s_1 = (\frac{s_{12}s_{13}}{s_{23}})^{\frac{1}{2}}$, $s_2 = (\frac{s_{23}s_{12}}{s_{13}})^{\frac{1}{2}}$, and $s_3 = (\frac{s_{13}s_{23}}{s_{12}})^{\frac{1}{2}}$ determine real values such that $s_1s_2 = s_{12}$, $s_1s_3 = s_{13}$, and $s_2s_3 = s_{23}$. Let $\widetilde{\alpha}_i = s_i\alpha_i$ for $i = 1, 2, 3$, we get that

$$\widetilde{E}_{ij} = s_{ij}E_{ij} = \tilde{\alpha}_iR_i^T(T_i - T_j)R_j\tilde{\alpha}_j \tag{11}$$

Hence, $\tilde{E}$ is a scaled consistent 3-view essential matrix. $\square$

4. Uniqueness of consistent camera matrices

This part deals with the argument stated in Corollary 1 in the paper, claiming that the recovery of camera matrices from a consistent multiview essential matrix is unique up to a global similarity transformation. More formally, we claim

Theorem 4. *Let E be a consistent n-view essential matrix and let $P_1, \dots, P_n$ be a set of non-collinear camera matrices which is consistent with E, then this set of camera matrices is unique up to some global similarity transformation.*

We first justify the argument for 3 cameras, exemplify its extension for 4 cameras and finally derives an induction for n cameras.

A calibrated camera matrix P_i is represented as

$$P_i(R_i, \mathbf{t}_i) = [R_i^T| - R_i^T t_i] \in \mathbb{R}^{3\times 4}$$

where R_i and $\mathbf{t}_i$ are the orientation and location of the i'th camera, respectively. In addition, we represent a similarity transformation $S(s, R, \mathbf{t})$ by a matrix of the form

$$S = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

Then, applying a similarity transformation on a camera matrix P_i yields

$$S(P_i) = PS = [sR_i^T R|R_i^T (t - t_i)]$$

which is

$$\left[R_i^T R | \frac{R_i^T (t - t_i)}{s}\right]$$

in a calibrated format.

3 cameras. We consider the case of 3 non-collinear cameras. Note that every point correspondence between two of the views can be extended to the third view by intersecting epipolar lines. Consequently, we can produce any number of correspondences across the three views. Uniqueness of reconstruction then follows from [12] who proved that 4 points in three views generally yield unique camera recovery.

4 cameras. We consider a consistent 4-view essential matrix E with 4 corresponding cameras

$$P_1(R_1, \mathbf{t}_1), P_2(R_2, \mathbf{t}_2), P_3(R_3, \mathbf{t}_3), P_4(R_4, \mathbf{t}_4).$$

Suppose we have another set of cameras which is consistent with E,

$$P_1^*(R_1^*, \mathbf{t}_1^*), P_2^*(R_2^*, \mathbf{t}_2^*), P_3^*(R_3^*, \mathbf{t}_3^*), P_4^*(R_4^*, \mathbf{t}_4^*).$$

We show that these two sets are the same up to a global similarity transformation. Without loss of generality we assume that the third camera is not collinear with the rest of cameras. From the consistency of E_{123}, there exists a similarity transformation $\bar{S}$ such that

$$\bar{S}(P_1) = P_1^*, \bar{S}(P_2) = P_2^*, \bar{S}(P_3) = P_3^*$$

and from the consistency of E_{234} there exists a similarity transformation $\hat{S}$ such that

$$\hat{S}(P_2) = P_2^*, \hat{S}(P_3) = P_3^*, \hat{S}(P_4) = P_4^*.$$

Now, since $P_2^* = \bar{S}(P_2) = \hat{S}(P_2)$, it yields

$$\left[R_2^T \bar{R} | \frac{R_2^T (\bar{t} - t_2)}{\bar{s}}\right] = \left[R_2^T \hat{R} | \frac{R_2^T (\hat{t} - t_2)}{\hat{s}}\right]$$

and consequently

$$R_2^T \bar{R} = R_2^T \hat{R} \Rightarrow \bar{R} = \hat{R}.$$

Then we have

$$\frac{R_2^T (\bar{t} - t_2)}{\bar{s}} = \frac{R_2^T (\hat{t} - t_2)}{\hat{s}} \Rightarrow \frac{\bar{t}}{\bar{s}} - \frac{t_2}{\bar{s}} = \frac{\hat{t}}{\hat{s}} - \frac{t_2}{\hat{s}}$$

and similarly from the relation $P_3^* = \bar{S}(P_3) = \hat{S}(P_3)$ we get that

$$\frac{R_3^T (\bar{t} - t_2)}{\bar{s}} = \frac{R_3^T (\hat{t} - t_2)}{\hat{s}} \Rightarrow \frac{\bar{t}}{\bar{s}} - \frac{t_3}{\bar{s}} = \frac{\hat{t}}{\hat{s}} - \frac{t_3}{\hat{s}}.$$

By subtracting the last two equations we have

$$\frac{t_3 - t_2}{\bar{s}} = \frac{t_3 - t_2}{\hat{s}} \Rightarrow \bar{s} = \hat{s} \Rightarrow \bar{t} = \hat{t},$$

implying that $\bar{S} = \hat{S}$.

Proof. The proof is by induction. The argument was proved for 3 cameras and now we suppose that we have a consistent n-view essential matrix. From the induction assumption we know that since $E_{1,\dots,(n-1)}$ is consistent, the recovery of the camera matrices is unique and similarly the recovery of the camera matrices from $E_{2,\dots,n}$ is unique. As in 4 cameras we assume without loss of generality that the third camera is not collinear with the rest of the cameras, and a similar derivation yields that the recovery of the cameras is unique up to some global similarity transformation. $\square$