A closed form solution for viewing graph construction in uncalibrated vision

Carlo Colombo Marco Fanfani
University of Florence
Via Santa Marta 3, 50139, Florence, Italy
{carlo.colombo, marco.fanfani}@unifi.it

Abstract

This paper presents a closed form solution for the problem of computing a set of projective cameras from the fundamental matrices of a given viewing graph. The approach is incremental, exploits trifocal constraints, and does not rely on either image or structure points. Represented by a vector of four parameters that uniquely ensure its consistency with the local trifocal geometry, each newly computed camera is automatically coherent with the projective frame chosen as global reference, thus not needing any a posteriori synchronization. Results of experiments made under controlled conditions show that the proposed approach is relatively resilient to noise, and faster by three orders of magnitude than classical camera resectioning solutions, while reaching a comparable accuracy. This makes our closed form approach a good candidate for camera initialization in scenarios involving large-scale viewing graphs.

1. Introduction

Uncalibrated computer vision is a way to keep low the number of parameters being estimated (which is quite useful in large-scale scenarios) and to deal efficiently with missing data and outliers [12]. Several computer vision applications encompass an uncalibrated reconstruction step, in which motion and/or structure are estimated up to a projective transformation. Bundle adjustment in uncalibrated settings is known to be less prone to local minima, with respect to the calibrated case [16]. Estimates are subsequently upgraded to metric by means of self-calibration [15] [13]. Some applications such as image-based rendering and view synthesis do not even require that calibration matrices be known at all [2]. Started in the early nineties [4] [7], research on uncalibrated vision flourished at the turn of the milennium, with formulations based on fully projective concepts and representations such as fundamental matrices [10], trifocal tensors [1], and rank conditions [11]. In modern projective Structure from Motion scenarios, typically working with huge collections of images, views are often endowed with a graph structure, referred to as *viewing graph*, whose nodes are cameras and edges are fundamental matrices [9] [18]. Several works have addressed problems connected with viewing graphs, such as finding the missing edges [14] [22] or establishing an optimal configuration for the nodes [6] [3] [20].

In this paper, a closed form solution is presented for obtaining a coherent set of projective cameras from a given, possibly incomplete, viewing graph. The main idea is to leverage basic theory on trifocal geometry and linear algebra in order to calculate in advance the expression for camera parameters, thus limiting to a minimum the computations at run time. Camera set optimization is left as a subsequent step. Starting from an arbitrary pair of cameras consistent with the fundamental matrix between them, the approach incrementally computes a new camera using the two additional fundamental matrices of the triplet. At each iteration, the four parameters that make the new camera fully consistent with the local trifocal geometry and with the global projective frame chosen, are computed without relying on either 2D or 3D points. The approach is experimentally validated on a minimal graph of four images under controlled noise conditions. A comparison is also carried out with the standard camera resectioning method based on scene point reprojection and numerical minimization [21]. Results show that the proposed approach is quite noise-resilient, with comparable performance both in terms of reprojection and projective 3D reconstruction errors, and remarkably faster computational speed. This makes our closed form solution quite promising for initializing cameras in applications requiring large-scale viewing graphs.

2. Theory

2.1. The third camera problem

Given two views with camera matrices $\mathtt{P}_1$ and $\mathtt{P}_2$, the fundamental matrix between them is uniquely determined as

$$\mathtt{F}_{21} \sim [\mathbf{e}_{21}]_\times \mathtt{P}_2 \mathtt{P}_1^+ \quad , \tag{1}$$

where $\mathbf{e}_{21}$ such that $\mathtt{F}_{21}^\top \mathbf{e}_{21} = \mathbf{0}$ is the epipole on view 2. (The symbol '$\sim$' denotes equality up to scale, $\mathtt{P}^+$ is the pseudo-inverse of $\mathtt{P}$ and $[\mathbf{u}]_\times \mathbf{v} \doteq \mathbf{u} \times \mathbf{v}$.) If $\mathtt{P}_1 \doteq [\mathtt{A}_1\ \mathbf{a}_1]$ and $\mathtt{F}_{21}$ are given instead, then $\mathtt{P}_2$ is not unique, as it is determined up to a 4 degrees of freedom transformation. Its general expression is

$$\mathtt{P}_2 \sim \left[[\mathbf{e}_{21}]_\times \mathtt{F}_{21} \quad \mathbf{e}_{21}\right] \begin{bmatrix} \mathtt{A}_1 & \mathbf{a}_1 \\ {}^1\boldsymbol{\rho}_2^\top & {}^1\sigma_2 \end{bmatrix} \quad , \tag{2}$$

where ${}^1\boldsymbol{\rho}_2$ is a (possibly zero) 3-vector and ${}^1\sigma_2$ is a nonzero scalar. One can easily verify that eq. 1 is satisfied whatever the choice of the four parameters, which is equivalent to say that the matrix $\mathtt{P}_2^\top \mathtt{F}_{21} \mathtt{P}_1$ *is skew-symmetric*. The latter condition is a practical way to test the *consistency* of a camera pair and a fundamental matrix. Notice that, if the canonical camera $[\mathtt{I}\ \mathbf{0}]$ is chosen as $\mathtt{P}_1$, then eq. 2 reduces to $\mathtt{P}_2 \sim \left[[\mathbf{e}_{21}]_\times \mathtt{F}_{21} + \mathbf{e}_{21}{}^1\boldsymbol{\rho}_2^\top \quad {}^1\sigma_2 \mathbf{e}_{21}\right]$, which is the usual expression for the second camera in textbooks [5] [8].

Assume now that a third view is given, together with the fundamental matrices relating it to the first and second views, respectively. The three fundamental matrices $\mathtt{F}_{21}$, $\mathtt{F}_{31}$ and $\mathtt{F}_{32}$, are not independent, as they must meet the three *trifocal compatibility* constraints

$$\epsilon_{ijk} = \mathbf{e}_{ik}^\top \mathtt{F}_{ij} \mathbf{e}_{jk} = 0, \quad i \neq j \neq k \quad . \tag{3}$$

The problem arises of computing an expression for the third camera given the first two, such that all camera pairs are consistent with the associated fundamental matrix. Using $\mathtt{P}_1$ as reference, the third camera can be written in the same form used in eq. 2 for the second camera, thus obtaining $\mathtt{P}_3 \sim \left[[\mathbf{e}_{31}]_\times \mathtt{F}_{31}\mathtt{A}_1 + \mathbf{e}_{31}{}^1\boldsymbol{\rho}_3^\top \quad [\mathbf{e}_{31}]_\times \mathtt{F}_{31}\mathbf{a}_1 + {}^1\sigma_3 \mathbf{e}_{31}\right]$. While this form automatically ensures the consistency of the camera pair ($\mathtt{P}_1$, $\mathtt{P}_3$) with $\mathtt{F}_{31}$ whatever the choice of the four parameters, there is actually only one choice of the vector $\left[{}^1\boldsymbol{\rho}_3^\top\ {}^1\sigma_3\right]^\top$ which guarantees that also the last consistency constraint, i.e., $\mathtt{P}_3^\top \mathtt{F}_{32} \mathtt{P}_2 + \left(\mathtt{P}_3^\top \mathtt{F}_{32} \mathtt{P}_2\right)^\top = 0$, is met. A closed form solution to this problem is derived hereafter. Since $\mathbf{e}_{31}^\top \mathtt{F}_{32} \mathbf{e}_{21} = 0$ by eq. 3, some of the mixed terms cancel out and the matrix $\mathtt{P}_3^\top \mathtt{F}_{32} \mathtt{P}_2$ can be written as

$$\mathtt{P}_1^\top \mathtt{Q} \mathtt{P}_1 + \mathtt{P}_1^\top \mathbf{q} \left[{}^1\boldsymbol{\rho}_2^\top\ {}^1\sigma_2\right] + \left[{}^1\boldsymbol{\rho}_3^\top\ {}^1\sigma_3\right]^\top \mathbf{r}^\top \mathtt{P}_1 \quad , \tag{4}$$

with $\mathtt{Q} \doteq -\mathtt{F}_{31}^\top [\mathbf{e}_{31}]_\times \mathtt{F}_{32} [\mathbf{e}_{21}]_\times \mathtt{F}_{21}, \mathbf{q} \doteq -\mathtt{F}_{31}^\top [\mathbf{e}_{31}]_\times \mathtt{F}_{32} \mathbf{e}_{21}$, and $\mathbf{r}^\top \doteq \mathbf{e}_{31}^\top \mathtt{F}_{32} [\mathbf{e}_{21}]_\times \mathtt{F}_{21}$.

A geometrical interpretation of the vectors $\mathbf{q}$ and $\mathbf{r}$ is easily obtained by considering the *trifocal plane* passing through the three camera centers. In the following we will assume, as in Fig. 1, that the centers are not aligned, so that this plane is unique, and the three epipole pairs $(\mathbf{e}_{ij}, \mathbf{e}_{ik})$ are distinct. The trifocal plane intersects the image plane i in the *trifocal line* $\mathbf{l}_i \doteq [\mathbf{e}_{ij}]_\times \mathbf{e}_{ik} / \|[\mathbf{e}_{ij}]_\times \mathbf{e}_{ik}\|$, i.e., the epipolar line passing through the local epipole pair (the condition $j < k$ fixes the sign of $\mathbf{l}_i$). Notice that

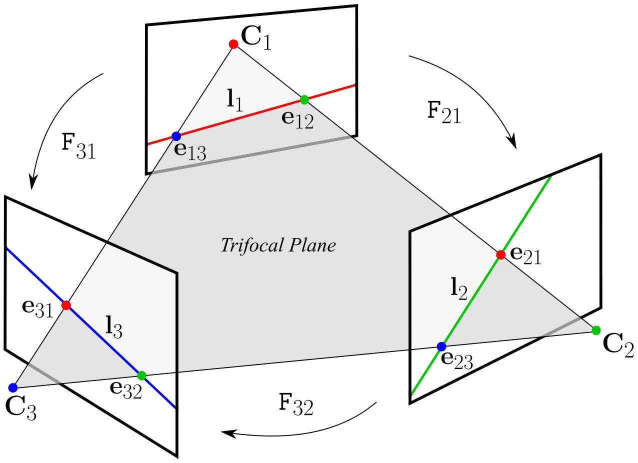


Figure 1. Trifocal geometry.

$$\mathbf{l}_i \sim \mathtt{F}_{ij} \mathbf{e}_{jk} \sim \mathtt{F}_{ij} [\mathbf{e}_{ji}]_\times \mathbf{l}_j \sim \mathtt{F}_{ij} [\mathbf{e}_{jk}]_\times \mathbf{l}_j \quad , \tag{5}$$

where the last two equalities hold since $[\mathbf{e}_{ji}]_\times \mathbf{l}_j \sim \mathbf{e}_{ji} \times \mathbf{e}_{ji} \times \mathbf{e}_{jk} = \mu \mathbf{e}_{ji} + \nu \mathbf{e}_{jk}$ for some μ and ν. Recalling that $\mathtt{F}_{ij} = \mathtt{F}_{ji}^\top$, and noting that $\mathbf{l}_i^\top \mathbf{l}_i \doteq 1$, it is clear from the above that $\mathbf{q}$ and $\mathbf{r}$ are both representations of the trifocal line $\mathbf{l}_1$, and in particular

$$\mathbf{q} = \chi \mathbf{l}_1 \quad , \qquad \chi \doteq -\mathbf{l}_1^\top \mathtt{F}_{31}^\top [\mathbf{e}_{31}]_\times \mathtt{F}_{32} \mathbf{e}_{21} \tag{6}$$

$$\mathbf{r} = \xi \mathbf{l}_1 \quad , \qquad \xi \doteq -\mathbf{l}_1^\top \mathtt{F}_{21}^\top [\mathbf{e}_{21}]_\times \mathtt{F}_{32}^\top \mathbf{e}_{31} \quad . \tag{7}$$

The matrix $\mathtt{Q} = -\mathtt{F}_{31}^\top [\mathbf{e}_{31}]_\times \mathtt{F}_{32} [\mathbf{e}_{21}]_\times \mathtt{F}_{21}$ can also be expressed in terms of the trifocal line $\mathbf{l}_1$. Indeed, it is a rank 2 matrix with $\mathbf{e}_{12}$ and $\mathbf{e}_{13}$ respectively in its right and left null spaces: $\mathtt{Q}\mathbf{e}_{12} = \mathtt{Q}^\top \mathbf{e}_{13} = \mathbf{0}$. Moreover, using again eq. 5, it is easy to show that $\mathtt{Q}\mathbf{e}_{13} \sim \mathtt{Q}^\top \mathbf{e}_{12} \sim \mathbf{l}_1$, so that $\mathtt{Q}$ is a solution of the following system, linear in $\mathtt{M}$:

$$\left\{ \begin{array}{rcl} \mathtt{M}\mathbf{e}_{12} & = & \mathbf{0} \\ \mathbf{e}_{13}^\top \mathtt{M} & = & \mathbf{0}^\top \\ \mathbf{e}_{12}^\top \mathtt{M} & = & h \mathbf{l}_1^\top \\ \mathtt{M}\mathbf{e}_{13} & = & k \mathbf{l}_1 \quad , \end{array} \right. \tag{8}$$

with scale factors $h \doteq -\mathbf{l}_1^\top \mathtt{F}_{21}^\top [\mathbf{e}_{21}]_\times \mathtt{F}_{32}^\top [\mathbf{e}_{31}]_\times \mathtt{F}_{31} \mathbf{e}_{12}$, and $k \doteq -\mathbf{l}_1^\top \mathtt{F}_{31}^\top [\mathbf{e}_{31}]_\times \mathtt{F}_{32} [\mathbf{e}_{21}]_\times \mathtt{F}_{21} \mathbf{e}_{13}$. The solution of the associated homogeneous system is the rank 1 matrix $\lambda \Lambda_1$, where $\Lambda_1 \doteq \mathbf{l}_1 \mathbf{l}_1^\top$ and λ is a free parameter. A particular solution of the non-homogeneous system above is $\lambda_2 \Lambda_1 [\mathbf{e}_{12}]_\times + \lambda_3 [\mathbf{e}_{13}]_\times \Lambda_1$ with $\lambda_2 \doteq k / \|[\mathbf{e}_{12}]_\times \mathbf{e}_{13}\|$ and $\lambda_3 \doteq h / \|[\mathbf{e}_{12}]_\times \mathbf{e}_{13}\|$, as one can verify by direct substitution. The general solution [19] of the system is therefore

$$\mathtt{M}(\lambda) = \lambda \Lambda_1 + \lambda_2 \Lambda_1 [\mathbf{e}_{12}]_\times + \lambda_3 [\mathbf{e}_{13}]_\times \Lambda_1 \quad , \tag{9}$$

from which the scalar λ_1 such that $\mathtt{M}(\lambda_1) = \mathtt{Q}$ can be determined as $\lambda_1 = \mathrm{trace}(\mathtt{Q})$, since $\Lambda_1 [\mathbf{e}_{12}]_\times$ and $[\mathbf{e}_{13}]_\times \Lambda_1$ are both traceless and $\mathrm{trace}(\Lambda_1) = 1$.

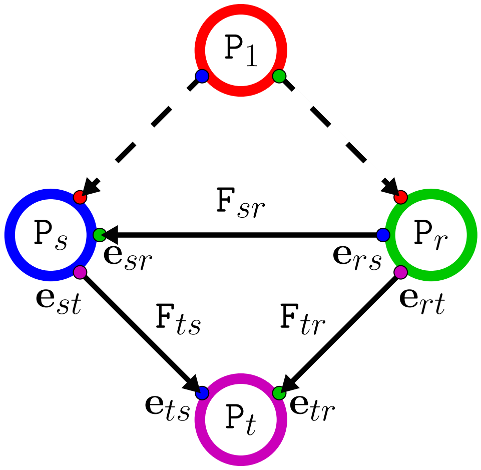


Figure 2. A view triplet from a larger viewing graph built using view 1 as global reference.

Now that a convenient representation has been established for $\mathtt{Q}$, $\mathbf{q}$ and $\mathbf{r}$, it is only a matter of algebraic computation to derive the closed form expression for the unique set of four parameters that force the matrix in eq. 4 to be skew-symmetric and ensure that $(\mathtt{P}_2, \mathtt{P}_3)$ be consistent with $\mathtt{F}_{32}$. This evaluates as

$$[{}^1\boldsymbol{\rho}_3^\top \; {}^1\sigma_3] = -\xi^{-1}\left(\mathbf{l}_1^\top \mathtt{B}\mathtt{P}_1 + \chi[{}^1\boldsymbol{\rho}_2^\top \; {}^1\sigma_2]\right) \quad , \tag{10}$$

where $\mathtt{B} \doteq \lambda_1 \mathtt{I} + \lambda_2[\mathbf{e}_{12}]_\times - \lambda_3[\mathbf{e}_{13}]_\times$.

2.2. Extension to larger viewing graphs

Once a camera, say $\mathtt{P}_1$, has been chosen as reference, the formula of eq. 10 allows one to obtain the unique parameter vector $[{}^1\boldsymbol{\rho}_s^\top \; {}^1\sigma_s]$ for camera $\mathtt{P}_s$ from a triplet of compatible fundamental matrices $\mathtt{F}_{r1}$, $\mathtt{F}_{s1}$, $\mathtt{F}_{sr}$ and the parameter vector $[{}^1\boldsymbol{\rho}_r^\top \; {}^1\sigma_r]$ for camera $\mathtt{P}_r$ (without loss of generality, the formula uses the indexes $r = 2$ and $s = 3$).

Suppose now to have a new view, say t, for which to compute $\mathtt{P}_t$ from the already known cameras $\mathtt{P}_r$ and $\mathtt{P}_s$, and the compatible triplet $\mathtt{F}_{sr}$, $\mathtt{F}_{tr}$, $\mathtt{F}_{ts}$. This situation, depicted in Fig. 2, is common when constructing large viewing graphs, in which a triplet of cameras may have no points in common with the *global* reference view (hereafter identified with view 1). Let us arbitrarily choose $\mathtt{P}_r \doteq [\mathtt{A}_r \; \mathbf{a}_r]$ as *local* reference for the new triplet. In order to exploit the formula of eq. 10 for obtaining $[{}^r\boldsymbol{\rho}_t^\top \; {}^r\sigma_t]$ and eventually $\mathtt{P}_t$, we need the parameter vector $[{}^r\boldsymbol{\rho}_s^\top \; {}^r\sigma_s]$ relating view s with view r. However, this is not immediately available, since $\mathtt{P}_s \doteq [\mathtt{A}_s \; \mathbf{a}_s] \doteq {}^1\mathtt{P}_s$ is expressed in terms of the 4-vector $[{}^1\boldsymbol{\rho}_s^\top \; {}^1\sigma_s]$, that relates view s with view 1. In other words, what is missing here is the local representation (relative to view r) of $\mathtt{P}_s$, for which only the global representation (relative to view 1) is currently known. (Notice, in passing, that a given camera can admit multiple, equivalent 4-vector representations according to the reference camera at hand. This does not contradict, of course, the fact that a camera requires exactly four parameters to be uniquely specified within a given triplet.) Recovering the local camera representation 4-vector can be done as follows. Let us define ${}^r\mathtt{P}_s \doteq {}^r\eta_s\left[[\mathbf{e}_{sr}]_\times \mathtt{F}_{sr}\mathtt{A}_r + \mathbf{e}_{sr}{}^r\boldsymbol{\rho}_s^\top \;\; [\mathbf{e}_{sr}]_\times \mathtt{F}_{sr}\mathbf{a}_r + {}^r\sigma_s\mathbf{e}_{sr}\right]$: This expression explicitly includes the overall scale factor ${}^r\eta_s$ of the camera matrix, which is also unknown, and must be computed together with ${}^r\boldsymbol{\rho}_s$ and ${}^r\sigma_s$. This goal can be reached by constraining the two matrices to be identical, i.e.

$${}^r\mathtt{P}_s = {}^1\mathtt{P}_s \quad . \tag{11}$$

After some algebraic passages, we get

$$[{}^r\boldsymbol{\rho}_s^\top \; {}^r\sigma_s] = \zeta^{-1}\mathbf{e}_{sr}^\top \, {}^1\mathtt{P}_s \quad , \tag{12}$$

where $\zeta \doteq {}^r\eta_s\|\mathbf{e}_{sr}\|^2 = \left(\mathbf{a}_s^\top[\mathbf{e}_{sr}]_\times\mathtt{F}_{sr}\mathbf{a}_r\right)^{-1}\|[\mathbf{e}_{sr}]_\times\mathbf{a}_s\|^2$. Notice that, thanks to the special definition of consistency given in eq. 2, the camera matrix ${}^r\mathtt{P}_t$, although computed using the formula of eq. 10 by means of the local reference $\mathtt{P}_r$, is automatically coherent with the projective frame chosen for the global reference $\mathtt{P}_1$ (usually $[\mathtt{I} \; \mathbf{0}]$), without needing any further adjustment. Hence, all the cameras of the viewing graph computed as above are expressed, as it should be, in a unique, global projective frame. If desired, the formula of eq. 12 can be used to recover $[{}^1\boldsymbol{\rho}_t^\top \; {}^1\sigma_t]$ from ${}^r\mathtt{P}_t$ and $\mathtt{P}_1$. Similarly, if missing and required, the fundamental matrix between the views t and 1, $\mathtt{F}_{t1}$, can also be obtained with the formula of eq. 1, with epipole $\mathtt{P}_t\mathtt{C}_1$, the camera center of view 1 being the null vector of $\mathtt{P}_1$. Reworking the formula in eq. 12 leads to an expression which is formally similar to that of eq. 10:

$$[{}^r\boldsymbol{\rho}_s^\top \; {}^r\sigma_s] = \zeta^{-1}\left(\mathbf{l}_s^\top \mathtt{C}\mathtt{P}_1 + x[{}^1\boldsymbol{\rho}_s^\top \; {}^1\sigma_s]\right) \quad , \tag{13}$$

with $\mathtt{C} \doteq -\|[\mathbf{e}_{s1}]_\times\mathbf{e}_{sr}\|\mathtt{F}_{s1}$ and $x \doteq \mathbf{e}_{s1}^\top\mathbf{e}_{sr}$. The difference is that while eq. 10 connects the representations of two distinct cameras relative to the same reference, eq. 13 connects two distinct representations of the same camera.

3. Evaluation

The closed form solution described in the previous Section is based on the key hypothesis that the compatibility constraints of eq. 3 are perfectly met for each camera triplet at hand. However, in real situations, noisy point measurements are to be expected, which would result in an inaccurate set of fundamental matrices, and determine a loss in trifocal compatibility. Such loss would affect the estimates of all the third cameras of the triplets, and alter the consistency of camera pairs with respect to the associated fundamental matrix. Eventually, the partially incorrect camera estimates would result in erroneous 3D estimates.

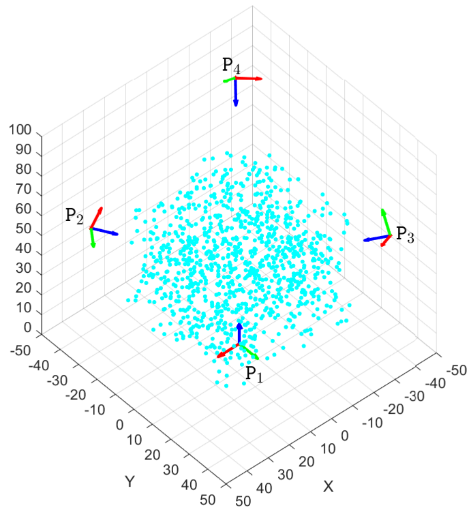


Figure 3. Camera configuration for setup #1. The 3D point cloud is shown in cyan dots. The four cameras $\mathtt{P}_1$, $\mathtt{P}_2$, $\mathtt{P}_3$, and $\mathtt{P}_4$ surround the point cloud. Camera axes are shown in red, green, and blue, respectively for the x-axis, y-axis, and z-axis (optical ray).

In order to experimentally validate the proposed method in the presence of noise, an essential simulation setup using four cameras was devised, and several tests were done, as described in the following.

3.1. Experimental setup

A 3D point cloud $\mathcal{M}_{GT}$ was created by uniformly sampling random points $\mathbf{X}_{GT} = [X, Y, Z]^\top$ from a three-dimensional volume with dimensions $\Delta_X = 60$, $\Delta_Y = 60$, $\Delta_Z = 50$. Then four calibrated cameras $\tilde{\mathtt{P}}_i$, with $i = \{1, \ldots, 4\}$, were defined, using a common calibration matrix $\mathtt{K}$. $\tilde{\mathtt{P}}_1$ was set in the origin of the coordinate system, i.e. $\tilde{\mathtt{P}}_1 \doteq \mathtt{K}[\mathtt{I} \quad \mathbf{0}]$. Camera centers for $\tilde{\mathtt{P}}_2$, $\tilde{\mathtt{P}}_3$, and $\tilde{\mathtt{P}}_4$ were then sampled randomly so as to have the four cameras arranged around the 3D point cloud with similar distances to each other. Once the camera centers were obtained, each rotation matrix was defined to have the optical ray pointing toward the center of the point cloud, and the other two axes arranged in a random configuration so as to form an orthonormal basis (see Fig. 3). In order to run the tests on multiple random camera configurations, 25 setups were created using the same procedure.

For each setup, the 3D points were projected onto the four cameras (neglecting occlusions), and matches among corresponding 2D points were obtained. Using the correspondences, the fundamental matrices $\mathtt{F}_{ji}$ were computed for each pair of cameras $\mathtt{P}_i$, $\mathtt{P}_j$. In all the experiments, we set the first projective camera as $\mathtt{P}_1 \doteq [\mathtt{I} \quad \mathbf{0}]$, and $\mathtt{P}_2 \doteq [[\mathbf{e}_{21}]_\times \mathtt{F}_{21} \quad \mathbf{e}_{21}]$. $\mathtt{P}_3^{(c)}$ and $\mathtt{P}_4^{(c)}$ were then obtained from the computed fundamental matrices using the closed form expressions presented in Sect. 2.

To compare the proposed approach with a classical solution, $\mathtt{P}_3^{(r)}$ and $\mathtt{P}_4^{(r)}$ were also estimated by minimizing the reprojection error between a set of 3D points and the respective 2D projections [21]. In particular, to compute $\mathtt{P}_3^{(r)}$, 3D points triangulated from $\mathtt{P}_1$ and $\mathtt{P}_2$ were used, while for $\mathtt{P}_4^{(r)}$ the point cloud generated from $\mathtt{P}_2$ and $\mathtt{P}_3^{(r)}$ was employed. Notice that, while our closed form approach requires at least 7 matches on two cameras to compute the fundamental matrices, the method based on the reprojection of 3D points needs at least 6 matches on three cameras (i.e., the two producing the 3D and the one to be estimated), which limits its applicability in a more general and realistic scenario—similar considerations apply for trifocal tensor based solutions [6].

Tests for each of the 25 different setups were repeated injecting Gaussian noise with $\mu = 0$ and $\sigma \in [0, 5]$, with steps of 0.1 (i.e., 51 noise sets were generated) to observe the behaviour of the approaches in presence of not perfect correspondences. The noise was independently sampled for each of the four images, and added to the 2D points. For the sake of uniformity, the same noise sets were used for all the setups. In total 25x51=1275 configurations were evaluated.

3.2. Error metrics

In order to evaluate the performance of the four projective cameras, three types of error were considered.

Firstly the consistency of all camera pairs w.r.t. the associated fundamental matrix were evaluated using the *consistency error* $\alpha_{ij} \doteq ||\mathtt{D}_{ij}||_F$, i.e., the Frobenius norm of the matrix

$$\mathtt{D}_{ij} \doteq (\mathtt{P}_j^\top \mathtt{F}_{ji} \mathtt{P}_i) + (\mathtt{P}_j^\top \mathtt{F}_{ji} \mathtt{P}_i)^\top \quad . \qquad (14)$$

In the case of perfect consistency, all entries of $\mathtt{D}_{ij}$ should be 0, and so the error α_{ij}. The higher the value of α_{ij}, the worse is the consistency of the matrices.

A second metric was the mean *2D reprojection error* among the pairs of cameras $\{\mathtt{P}_1, \mathtt{P}_3\}$, $\{\mathtt{P}_1, \mathtt{P}_4\}$, $\{\mathtt{P}_2, \mathtt{P}_3\}$, $\{\mathtt{P}_2, \mathtt{P}_4\}$, $\{\mathtt{P}_3, \mathtt{P}_4\}$. Let $\mathcal{S}_{ij} = \{\tilde{\mathbf{X}}_{ij}\}$ be the set of 3D points estimated by triangulation from $\mathtt{P}_i$ and $\mathtt{P}_j$ using the matched 2D points $\mathbf{x}_i \in \mathtt{P}_i$, and $\mathbf{x}_j \in \mathtt{P}_j$. The error is computed as

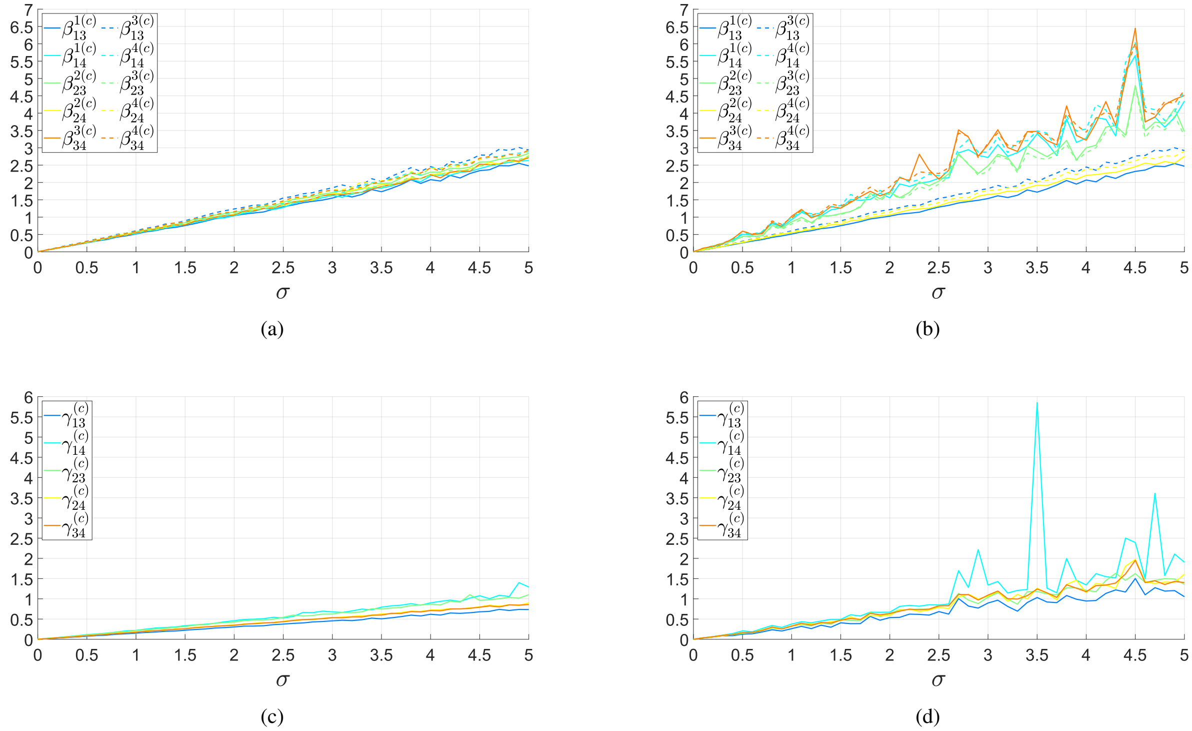
$$\beta_{ij}^i = \frac{1}{|\mathcal{S}_{ij}|} \sum_{\substack{\tilde{\mathbf{X}}_{ij} \in \mathcal{S}_{ij} \\ \mathbf{x}_i \in \mathtt{P}_i}} \sqrt{(\mathbf{x}_i - \mathtt{P}_i \tilde{\mathbf{X}}_{ij})^2} \quad , \qquad (15)$$

where $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. A similar error β_{ij}^j is obtained for the j-th camera.

Finally, a *3D reconstruction error* was also measured. Exploiting the knowledge of the metric ground-truth 3D point cloud $\mathcal{M}_{GT}$, each 3D point set $\mathcal{S}_{ij}$ was promoted to a metric reconstruction estimating a 4x4 transformation $\mathtt{H}_{ij}$, as done in [17]. The promoted point cloud $\mathcal{M}_{ij} = \mathtt{H}_{ij} \mathcal{S}_{ij}$

$\beta_{13}^{1(c)}$	$\beta_{13}^{3(c)}$	$\beta_{14}^{1(c)}$	$\beta_{14}^{4(c)}$	$\beta_{23}^{2(c)}$	$\beta_{23}^{3(c)}$	$\beta_{24}^{2(c)}$	$\beta_{24}^{4(c)}$	$\beta_{34}^{3(c)}$	$\beta_{34}^{4(c)}$	$\gamma_{13}^{(c)}$	$\gamma_{14}^{(c)}$	$\gamma_{23}^{(c)}$	$\gamma_{24}^{(c)}$	$\gamma_{34}^{(c)}$
1.29	1.52	1.33	1.46	1.42	1.37	1.37	1.46	1.35	1.46	0.38	0.57	0.54	0.44	0.44

Table 1. Errors obtained in the tests using ground-truth $\mathtt{F}$'s, averaged on all the setups and all the noise σ's.



(a)

(b)

(c)

(d)

Figure 4. Average errors on all setups for increasing values of σ. (a) and (c) show the errors obtained using ground-truth $\mathtt{F}$'s, while (b) and (d) report results using noisy fundamental matrices. The scale of the y-axes of the plots in the same row is the same, so as to ease the comparison. Best viewed in colour. The reader is invited to zoom in the electronic version of the paper in order to appreciate finer details.

was then compared with the ground-truth and its error was evaluated as

$$\gamma_{ij} = \frac{1}{|\mathcal{M}_{ij}|} \sum_{\substack{\mathbf{X}_{ij} \in \mathcal{M}_{ij}, \\ \mathbf{X}_{GT} \in \mathcal{M}_{GT}}} \sqrt{(\mathbf{X}_{ij} - \mathbf{X}_{GT})^2} \quad . \tag{16}$$

Since $\mathtt{H}_{ij}$ was obtained by solving an over-constrained linear system that minimizes a similar error, γ_{ij} would partially hide the effect of the introduced noise. To be more fair in the comparison, we decided to estimate a unique and perfect $\hat{\mathtt{H}}_{ij}$ from each noiseless setup (i.e., $\sigma = 0$), and apply it to the noisy estimates of the same camera pair $\mathtt{P}_i$ and $\mathtt{P}_j$.

For the purpose of comparison, the metrics above were evaluated using both the projective cameras obtained with the proposed method (i.e., $\mathtt{P}_3^{(c)}$ and $\mathtt{P}_4^{(c)}$), and those estimated with the reprojection based solution (i.e., $\mathtt{P}_3^{(r)}$ and $\mathtt{P}_4^{(r)}$).

Finally, in order to observe how trifocal compatibility degrades with noise, and better explain the inferior performances in the presence of noise for our solution, we also measured the overall *trifocal compatibility error*

$$\delta_{ijk} = \frac{|\epsilon_{ijk}| + |\epsilon_{jki}| + |\epsilon_{kij}|}{3} \quad , \tag{17}$$

where ϵ_{ijk} is defined in eq. 3 .

3.3. Results using ground-truth F matrices

Before presenting the comparative results, since the proposed approach performances are based on the estimation accuracy of the $\mathtt{F}$ matrices, a test using perfect $\mathtt{F}$'s was conducted and reported hereafter. Using perfect fundamental matrices, obtained from noiseless matches, the projective $\mathtt{P}_3^{(c)}$ and $\mathtt{P}_4^{(c)}$ matrices are perfectly recovered, and the noise only affects the triangulation process, which uses inexact matches.

In Table 1 the $\beta_{ij}^{(c)}$ and $\gamma_{ij}^{(c)}$ errors were reported, averaged on all the setups and all the noise σ's. (In this case

$\alpha_{13}^{(c)}$	$\alpha_{14}^{(c)}$	$\alpha_{23}^{(c)}$	$\alpha_{24}^{(c)}$	$\alpha_{34}^{(c)}$	$\alpha_{13}^{(r)}$	$\alpha_{14}^{(r)}$	$\alpha_{23}^{(r)}$	$\alpha_{24}^{(r)}$	$\alpha_{34}^{(r)}$
0.00000	0.16657	0.10525	0.00000	0.42145	0.09039	0.01988	0.06028	0.06206	0.00542

Table 2. Mean camera consistency errors, obtained by averaging on all the setups and on all the noise σ's.

we did not report the results obtained by the reprojection method, since it will be an unfair comparison. Note also that we did not present tables for the $\alpha_{ij}^{(c)}$ and δ_{ijk} errors, since using perfect $\mathtt{F}$ matrices these errors are all zero.) Using ground-truth $\mathtt{F}$'s, both the errors are quite stable for each camera pair. This is due to the fact that the consistency of the camera matrices is guaranteed by construction, and no particular pair suffers from inconsistency issues.

Nevertheless, the noise has an appreciable effect on the triangulation and, as a consequence, on both the errors. As can be seen in Figs. 4a and 4c, representing respectively the $\beta_{ij}^{(c)}$ and $\gamma_{ij}^{(c)}$ errors averaged on all the setups at different values of σ, both the errors grow almost linearly with the noise. In the absence of noise ($\sigma = 0$), all error metrics give a zero value, thus experimentally confirming the correctness of the formulas provided in Sect. 2. As noises increases, the errors increase in a similar way for each camera pair.

3.4. Comparative results

In this Section, comparative results between the proposed approach and the reprojection based method are presented. In this case, also the $\mathtt{F}$ matrices were estimated from noisy matches.

Firstly, in Tab. 2 camera consistency errors for all the pairs (i.e., $\alpha_{ij}^{(c)}$ and $\alpha_{ij}^{(r)}$) are reported. As it can be noticed, the consistency error for cameras $\mathtt{P}_3^{(c)}$ and $\mathtt{P}_4^{(c)}$ is unchanged from the previous case only for the pairs $\{1,3\}$, and $\{2,4\}$: This is due to the fact that $\mathtt{P}_3^{(c)}$ was built to be perfectly consistent with $\mathtt{P}_1$, and, similarly, $\mathtt{P}_4^{(c)}$ was built with perfect consistency w.r.t. $\mathtt{P}_2$. For the other pairs, consistency is not guaranteed by construction, and they suffer from the introduction of noise in the $\mathtt{F}$ estimation, particularly the pair $\{3,4\}$. Indeed, the loss of consistency was due to the decreased trifocal compatibility among the estimated $\mathtt{F}$ matrices. Looking to Fig. 5—that reports the mean δ_{ijk} for the triplets $\{1,2,3\}$, $\{2,3,4\}$, and $\{1,2,4\}$ averaged on all setups for different values of noise σ's—the compatibility decreases as the noise increases. On the other hand, using the reprojection method, camera consistencies reached similar error values for all the pairs, since in estimating the cameras their consistency was implicitly optimized by minimizing the reprojection error.

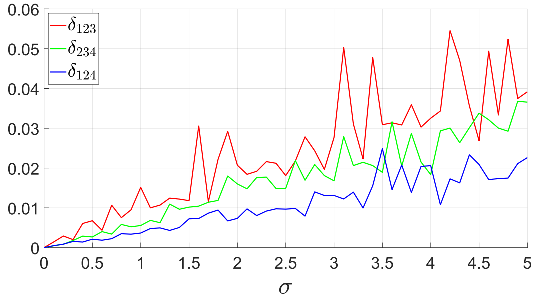


Figure 5. Average trifocal compatibility errors on all setups for increasing values of σ, using $\mathtt{F}$'s estimated on noisy matches. Best viewed in colour. The reader is invited to zoom in the electronic version of the paper in order to appreciate finer details.

The worst consistencies $\alpha_{14}^{(c)}$, $\alpha_{23}^{(c)}$, and $\alpha_{34}^{(c)}$ produced the worst results on the reprojection errors $\beta_{ij}^{(c)}$ for the relative camera pairs. As can be seen in Fig. 4b (where a plot similar to that of Fig. 4a was reported, by averaging the errors on all the setups for different values of σ) $\beta_{14}^{(c)}$, $\beta_{23}^{(c)}$, and $\beta_{34}^{(c)}$ were higher than $\beta_{13}^{(c)}$ and $\beta_{24}^{(c)}$. However, this effect is not generally reflected on the $\gamma_{ij}^{(c)}$ errors—excluding some peaks for $\gamma_{14}^{(c)}$ (see Fig. 4d). In our opinion, this is due to the fact that transformation matrices $\hat{\mathtt{H}}_{ij}$, being estimated independently for each pair, reduced the effect of low consistency for pairs $\{1,4\}$, $\{2,3\}$, and $\{3,4\}$.

The overall results averaged on all the setups and all σ's, are finally reported in Tables 3a and 3b, giving respectively the scores for our solution and for the reprojection approach. From a comparison of Tab. 3a and Tab. 1 (where ground-truth $\mathtt{F}$'s were used) the most relevant differences are appreciable for the pairs $\{1,4\}$, $\{2,3\}$, and $\{3,4\}$, which are those whose consistency degrades more with noise. Comparing instead the scores for the closed form and the reprojection based approaches, both methods obtained similar error values, indicating that our approach does not suffer of particular drawbacks w.r.t. the more classical reprojection solution which is based on error minimization. However, some particular results can be pointed out. Considering the average scores, our solution obtained lower β_{ij} errors for the camera pairs $\{1,3\}$ and $\{2,4\}$: Indeed, while $\beta_{13}^{1(c)}$ and $\beta_{13}^{3(c)}$ reached respectively the values of 1.28 and 1.51, $\beta_{13}^{1(r)}$ and $\beta_{13}^{3(r)}$ obtained 1.59 and 1.85. More pronounced is the difference for the pair $\{2,4\}$, with $\beta_{24}^{2(c)} = 1.36$ and $\beta_{24}^{4(c)} = 1.45$, while $\beta_{24}^{2(r)} = 2.45$ and $\beta_{24}^{4(r)} = 2.50$. For the remaining pairs ($\{1,4\}$, $\{2,3\}$, and $\{3,4\}$) the β_{ij} errors were quite similar, only slightly lower for the reprojection based approach. This behaviour indicates again that the proposed solution is particularly reliable for camera pairs with good consistency, and suggests

$\beta^{1(c)}_{13}$	$\beta^{3(c)}_{13}$	$\beta^{1(c)}_{14}$	$\beta^{4(c)}_{14}$	$\beta^{2(c)}_{23}$	$\beta^{3(c)}_{23}$	$\beta^{2(c)}_{24}$	$\beta^{4(c)}_{24}$	$\beta^{3(c)}_{34}$	$\beta^{4(c)}_{34}$	$\gamma^{(c)}_{13}$	$\gamma^{(c)}_{14}$	$\gamma^{(c)}_{23}$	$\gamma^{(c)}_{24}$	$\gamma^{(c)}_{34}$
1.28	1.51	2.24	2.39	2.04	1.97	1.36	1.45	2.41	2.44	0.66	1.13	0.80	0.82	0.82

(a) Closed form method

$\beta^{1(r)}_{13}$	$\beta^{3(r)}_{13}$	$\beta^{1(r)}_{14}$	$\beta^{4(r)}_{14}$	$\beta^{2(r)}_{23}$	$\beta^{3(r)}_{23}$	$\beta^{2(r)}_{24}$	$\beta^{4(r)}_{24}$	$\beta^{3(r)}_{34}$	$\beta^{4(r)}_{34}$	$\gamma^{(r)}_{13}$	$\gamma^{(r)}_{14}$	$\gamma^{(r)}_{23}$	$\gamma^{(r)}_{24}$	$\gamma^{(r)}_{34}$
1.59	1.85	2.24	2.50	1.74	1.68	2.45	2.50	2.10	2.54	0.55	1.07	0.81	1.30	1.07

(b) Reprojection based method

Table 3. Average errors obtained using $\mathtt{F}$'s estimated from noisy matches.

that in a practical application the $\alpha^{(c)}_{ij}$ score could be used to choose from which camera pairs to compute the 3D reconstruction. Concerning the γ_{ij} errors, the main difference can be observed for $\gamma^{(c)}_{24} = 0.82$ and $\gamma^{(r)}_{24} = 1.30$, favourable toward the proposed approach. The other pairs obtained mostly similar scores.

Tables 4a and 4b report the errors obtained for the setup #1, for all values of σ. As can be seen, the errors grow linearly with the noise, and both methods obtain comparable results.

3.5. Computational times

In this section, some indication on the average computational times for the compared approaches are reported. Times were measured using a non optimized Matlab code on a PC with an Intel Core i7-10510U CPU with 16GB of RAM. On average, the proposed method used about 3×10^{-5} s to compute a camera, while the reprojection approach took about 3×10^{-2} s, hence was about 1000 times slower. Note additionally that, for the reprojection method, most of the time was required for point triangulation, which has an increasing complexity related to the number of 3D points estimated. Indeed, this is another favourable aspect of the closed form solution, that not only has fast computational times, but also does not require to triangulate points in order to estimate the camera matrices.

4. Conclusions and future work

In this paper, a closed form solution for the estimation of projective camera matrices on a viewing graph was given. Exploiting only matches on pairs of cameras to compute the fundamental matrices, using our formulas is it possible to obtain all the camera matrices in an unique coordinate system, without the need of any a posteriori synchronization. The use of closed form expressions limits to a minimum the amount of run time calculations. This is particularly desirable in applications requiring large viewing graphs.

Results on a simulated environment confirmed the correctness and demonstrated the noise resilience of the formulas, thanks to which the proposed approach obtains similar, if not superior, performances w.r.t. a classical solution based on the minimization of the reprojection error. Moreover, as it does not require any point triangulation, our solution is faster than the standard approach by three orders of magnitude.

Future work will encompass extending the theory to the degenerate case of collinear camera centers, addressing the problem of camera set optimization, and performing further tests on real images and larger viewing graphs, in order to better evaluate the performances in a practical, realistic application scenario.

References

[1] Shai Avidan and Amnon Shashua. Threading Fundamental Matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(1):73–77, Jan. 2001. 1

[2] Antonio Canclini, Francesco Malapelle, Marco Marcon, Stefano Tubaro, and Andrea Fusiello. View-synthesis from uncalibrated cameras and parallel planes. *Signal Processing: Image Communication*, 79:40–53, 2019. 1

[3] Jérôme Courchay, Arnak Dalalyan, Renaud Keriven, and Peter Sturm. A global camera network calibration method with linear programming. In *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 05 2010. 1

[4] Olivier Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *European Conference on Computer Vision (ECCV)*, pages 563–578. Springer, 1992. 1

[5] Olivier Faugeras, Quang-Tuan Luong, and T. Papadopoulou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, Cambridge, MA, USA, 2001. 2

[6] Jacob Goldberger. Reconstructing camera projection matrices from multiple pairwise overlapping views. *Computer Vision and Image Understanding*, 97(3):283–296, 2005. 1, 4

[7] Richard I. Hartley, Rajiv Gupta, and Tom Chang. Stereo from uncalibrated cameras. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 761–764, 1992. 1

[8] Richard. I. Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 2

σ	$\beta_{13}^{1(c)}$	$\beta_{13}^{3(c)}$	$\beta_{14}^{1(c)}$	$\beta_{14}^{4(c)}$	$\beta_{23}^{2(c)}$	$\beta_{23}^{3(c)}$	$\beta_{24}^{2(c)}$	$\beta_{24}^{4(c)}$	$\beta_{34}^{3(c)}$	$\beta_{34}^{4(c)}$	$\gamma_{13}^{(c)}$	$\gamma_{14}^{(c)}$	$\gamma_{23}^{(c)}$	$\gamma_{24}^{(c)}$	$\gamma_{34}^{(c)}$
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.05	0.07	0.05	0.07	0.05	0.06	0.05	0.06	0.07	0.05	0.02	0.03	0.03	0.02	0.02
0.2	0.10	0.13	0.10	0.13	0.10	0.11	0.10	0.13	0.14	0.09	0.04	0.06	0.05	0.05	0.05
0.3	0.15	0.20	0.15	0.19	0.15	0.17	0.16	0.19	0.20	0.14	0.05	0.08	0.07	0.06	0.06
0.4	0.19	0.25	0.22	0.30	0.20	0.22	0.22	0.26	0.25	0.18	0.07	0.10	0.10	0.09	0.09
0.5	0.24	0.32	0.24	0.32	0.27	0.30	0.25	0.31	0.35	0.25	0.08	0.13	0.12	0.11	0.10
0.6	0.28	0.38	0.31	0.41	0.33	0.35	0.32	0.39	0.42	0.30	0.12	0.16	0.15	0.14	0.14
0.7	0.34	0.44	0.33	0.44	0.38	0.41	0.34	0.41	0.48	0.33	0.17	0.22	0.21	0.19	0.19
0.8	0.41	0.54	0.43	0.58	0.43	0.47	0.44	0.54	0.59	0.42	0.15	0.21	0.22	0.19	0.18
0.9	0.44	0.57	0.50	0.65	0.50	0.55	0.46	0.56	0.64	0.45	0.25	0.32	0.30	0.30	0.28
1.0	0.46	0.61	0.49	0.64	0.57	0.61	0.51	0.61	0.70	0.49	0.28	0.38	0.34	0.32	0.31
1.1	0.55	0.73	0.54	0.73	0.60	0.69	0.57	0.69	0.79	0.54	0.19	0.31	0.27	0.23	0.24
1.2	0.61	0.80	0.59	0.78	0.63	0.68	0.58	0.71	0.83	0.58	0.24	0.37	0.31	0.30	0.29
1.3	0.60	0.80	0.66	0.89	0.70	0.78	0.68	0.82	0.91	0.63	0.26	0.36	0.33	0.31	0.31
1.4	0.67	0.87	0.73	0.98	0.75	0.83	0.71	0.87	0.98	0.68	0.23	0.38	0.32	0.31	0.31
1.5	0.71	0.93	0.75	0.99	0.84	0.89	0.76	0.91	1.02	0.70	0.26	0.39	0.36	0.32	0.32
1.6	0.75	0.98	0.75	0.99	0.94	0.99	0.82	1.00	1.14	0.81	0.35	0.46	0.44	0.40	0.39
1.7	0.85	1.14	0.87	1.10	0.99	1.07	0.92	1.10	1.18	0.84	0.30	0.45	0.39	0.37	0.36
1.8	0.88	1.19	1.00	1.32	1.02	1.10	0.95	1.15	1.31	0.92	0.51	0.66	0.61	0.60	0.57
1.9	0.94	1.26	0.91	1.21	1.10	1.16	0.99	1.21	1.24	0.88	0.34	0.51	0.49	0.43	0.45
2.0	0.94	1.24	1.32	1.79	1.15	1.21	1.00	1.22	1.56	1.10	0.33	0.62	0.48	0.57	0.55
2.1	0.99	1.30	1.03	1.33	1.13	1.22	1.14	1.40	1.39	0.98	0.56	0.67	0.63	0.64	0.63
2.2	1.03	1.36	1.09	1.46	1.16	1.26	1.11	1.34	1.53	1.09	0.53	0.70	0.71	0.68	0.63
2.3	1.03	1.35	1.20	1.56	1.24	1.36	1.22	1.45	1.55	1.11	0.39	0.64	0.56	0.52	0.53
2.4	1.12	1.50	1.17	1.57	1.23	1.32	1.21	1.45	1.64	1.14	0.47	0.73	0.61	0.57	0.55
2.5	1.24	1.68	1.32	1.75	1.48	1.58	1.32	1.57	1.84	1.28	0.42	0.68	0.63	0.61	0.58
2.6	1.27	1.68	1.70	2.22	1.35	1.50	1.38	1.64	1.90	1.33	0.54	0.86	0.70	0.71	0.74
2.7	1.38	1.81	1.49	1.97	1.47	1.60	1.39	1.65	1.81	1.29	0.85	1.08	0.98	0.96	0.92
2.8	1.35	1.76	1.84	2.35	1.51	1.66	1.44	1.68	2.09	1.46	0.48	0.92	0.70	0.76	0.73
2.9	1.36	1.84	1.44	1.88	1.61	1.74	1.45	1.72	1.93	1.37	0.49	0.81	0.71	0.60	0.60
3.0	1.44	1.91	1.58	2.09	1.52	1.68	1.61	1.95	2.09	1.45	0.59	0.89	0.78	0.73	0.72
3.1	1.41	1.88	1.44	1.92	1.67	1.85	1.57	1.87	2.07	1.48	0.85	1.00	0.97	0.92	0.87
3.2	1.50	1.97	1.60	2.09	1.55	1.67	1.61	1.95	2.20	1.59	0.50	0.79	0.73	0.67	0.64
3.3	1.49	1.94	1.79	2.28	1.78	1.97	1.64	1.95	2.32	1.61	0.55	0.89	0.81	0.74	0.69
3.4	1.70	2.21	1.90	2.46	1.86	2.05	1.68	2.04	2.44	1.69	0.93	1.36	1.12	1.08	1.07
3.5	1.70	2.18	1.90	2.47	1.79	1.89	1.87	2.25	2.37	1.72	0.89	1.35	1.09	1.06	1.00
3.6	1.68	2.23	1.84	2.43	1.88	2.10	1.77	2.09	2.56	1.82	0.65	1.01	0.90	0.77	0.75
3.7	1.81	2.37	1.96	2.56	1.97	2.13	1.95	2.33	2.57	1.82	0.64	1.07	0.93	0.83	0.79
3.8	1.73	2.24	1.93	2.56	2.10	2.26	1.97	2.39	2.65	1.89	0.75	1.04	0.99	0.90	0.89
3.9	1.89	2.48	1.95	2.57	2.12	2.25	1.94	2.35	2.67	1.86	0.88	1.14	1.18	1.05	1.06
4.0	1.90	2.51	2.18	2.90	2.23	2.47	2.04	2.42	2.77	1.93	0.86	1.22	1.10	1.03	1.01
4.1	1.93	2.50	2.34	2.96	2.16	2.33	2.19	2.61	2.79	1.98	0.76	1.30	1.07	1.01	0.99
4.2	2.11	2.80	2.80	3.72	2.40	2.66	2.27	2.67	3.02	2.17	0.90	1.30	1.16	1.31	1.31
4.3	1.99	2.69	2.20	2.86	2.25	2.58	2.19	2.57	2.89	2.08	1.48	1.81	1.69	1.61	1.60
4.4	2.11	2.78	2.21	2.91	2.35	2.62	2.35	2.81	3.02	2.13	0.84	1.38	1.10	1.07	1.03
4.5	2.21	2.92	2.26	3.04	2.27	2.43	2.32	2.76	3.17	2.21	1.31	1.61	1.57	1.47	1.45
4.6	2.09	2.83	2.28	2.99	2.52	2.65	2.56	3.08	2.99	2.17	0.93	1.34	1.24	1.17	1.15
4.7	2.25	3.03	2.58	3.41	2.47	2.61	2.33	2.75	3.18	2.26	1.83	2.08	2.02	2.03	1.92
4.8	2.31	3.10	2.32	3.04	2.51	2.81	2.56	3.11	3.49	2.44	0.93	1.32	1.25	1.09	1.11
4.9	2.37	3.07	2.56	3.44	2.68	2.85	2.59	3.15	3.23	2.27	0.98	1.41	1.35	1.27	1.21
5.0	2.22	2.96	2.39	3.14	2.67	2.81	2.56	3.05	3.29	2.37	0.81	1.25	1.17	1.06	1.02
Avgs	1.19	1.57	1.32	1.73	1.35	1.46	1.30	1.55	1.73	1.22	0.55	0.78	0.71	0.67	0.66

(a) Closed form method

σ	$\beta_{13}^{1(r)}$	$\beta_{13}^{3(r)}$	$\beta_{14}^{1(r)}$	$\beta_{14}^{4(r)}$	$\beta_{23}^{2(r)}$	$\beta_{23}^{3(r)}$	$\beta_{24}^{2(r)}$	$\beta_{24}^{4(r)}$	$\beta_{34}^{3(r)}$	$\beta_{34}^{4(r)}$	$\gamma_{13}^{(r)}$	$\gamma_{14}^{(r)}$	$\gamma_{23}^{(r)}$	$\gamma_{24}^{(r)}$	$\gamma_{34}^{(r)}$
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.05	0.07	0.05	0.07	0.05	0.06	0.06	0.07	0.07	0.05	0.02	0.03	0.03	0.02	0.02
0.2	0.11	0.14	0.10	0.14	0.10	0.11	0.11	0.13	0.15	0.11	0.04	0.05	0.05	0.05	0.05
0.3	0.16	0.21	0.18	0.22	0.16	0.18	0.16	0.19	0.21	0.15	0.05	0.08	0.07	0.07	0.07
0.4	0.20	0.26	0.21	0.27	0.20	0.22	0.23	0.28	0.27	0.20	0.08	0.11	0.10	0.09	0.09
0.5	0.25	0.32	0.24	0.32	0.28	0.30	0.26	0.32	0.35	0.24	0.09	0.14	0.12	0.10	0.11
0.6	0.29	0.39	0.33	0.43	0.34	0.36	0.33	0.39	0.46	0.32	0.11	0.16	0.15	0.14	0.14
0.7	0.34	0.45	0.35	0.46	0.41	0.45	0.35	0.43	0.51	0.36	0.18	0.24	0.23	0.21	0.21
0.8	0.42	0.55	0.41	0.56	0.44	0.49	0.52	0.62	0.65	0.48	0.16	0.27	0.22	0.22	0.22
0.9	0.46	0.60	0.46	0.60	0.52	0.57	0.47	0.56	0.63	0.46	0.24	0.31	0.29	0.29	0.26
1.0	0.50	0.66	0.52	0.67	0.58	0.62	0.53	0.64	0.72	0.51	0.28	0.34	0.35	0.31	0.29
1.1	0.57	0.75	0.58	0.78	0.62	0.70	0.66	0.78	0.87	0.63	0.19	0.31	0.28	0.23	0.24
1.2	0.63	0.82	0.58	0.75	0.66	0.72	0.62	0.74	0.82	0.59	0.26	0.37	0.35	0.32	0.34
1.3	0.64	0.85	0.78	1.02	0.73	0.79	0.84	0.95	1.01	0.74	0.27	0.44	0.36	0.40	0.38
1.4	0.70	0.92	0.75	0.99	0.77	0.86	0.74	0.89	0.98	0.68	0.23	0.36	0.33	0.31	0.30
1.5	0.76	0.99	0.83	1.06	0.87	0.94	0.84	0.99	1.05	0.74	0.28	0.42	0.39	0.34	0.37
1.6	0.75	0.99	0.80	1.06	0.95	0.99	0.86	1.04	1.21	0.90	0.35	0.44	0.44	0.38	0.38
1.7	0.87	1.16	0.88	1.10	0.99	1.06	0.94	1.12	1.20	0.88	0.29	0.48	0.39	0.38	0.37
1.8	0.89	1.20	0.99	1.31	1.01	1.08	1.02	1.22	1.33	0.95	0.48	0.63	0.61	0.57	0.56
1.9	0.97	1.30	0.92	1.23	1.11	1.19	1.06	1.30	1.28	0.91	0.35	0.51	0.50	0.43	0.47
2.0	0.99	1.31	1.09	1.39	1.18	1.23	1.14	1.34	1.38	1.01	0.35	0.54	0.47	0.47	0.45
2.1	1.02	1.32	1.09	1.43	1.13	1.23	1.23	1.47	1.57	1.15	0.56	0.60	0.64	0.61	0.63
2.2	1.08	1.41	1.14	1.52	1.17	1.26	1.38	1.60	1.78	1.29	0.59	0.71	0.78	0.75	0.74
2.3	1.03	1.35	1.13	1.44	1.32	1.43	1.28	1.51	1.50	1.08	0.39	0.60	0.56	0.47	0.49
2.4	1.14	1.52	1.27	1.71	1.25	1.33	1.52	1.74	1.94	1.43	0.47	0.69	0.62	0.60	0.60
2.5	1.25	1.69	1.30	1.65	1.49	1.63	1.47	1.74	1.92	1.41	0.42	0.67	0.65	0.63	0.56
2.6	1.28	1.67	1.36	1.71	1.39	1.53	1.56	1.80	1.91	1.36	0.49	0.69	0.67	0.62	0.62
2.7	1.42	1.84	1.57	2.00	1.52	1.63	1.43	1.69	1.96	1.42	0.93	1.10	1.04	1.03	1.01
2.8	1.45	1.89	1.46	1.87	1.50	1.66	1.59	1.84	2.20	1.62	0.48	0.80	0.73	0.64	0.64
2.9	1.38	1.86	1.47	1.91	1.64	1.75	1.51	1.79	1.98	1.40	0.52	0.82	0.73	0.61	0.62
3.0	1.52	2.03	1.59	2.02	1.70	1.83	1.71	2.05	2.13	1.48	0.64	0.92	0.81	0.73	0.73
3.1	1.45	1.96	1.61	2.16	1.87	2.08	2.05	2.31	2.41	1.75	0.89	1.03	1.05	0.98	1.06
3.2	1.52	2.00	1.65	2.12	1.56	1.68	1.68	2.02	2.20	1.58	0.53	0.83	0.76	0.70	0.68
3.3	1.53	1.97	1.64	2.15	1.80	1.99	2.11	2.41	2.77	2.02	0.55	1.05	0.81	0.86	0.82
3.4	1.74	2.29	1.93	2.47	1.94	2.13	1.69	2.05	2.32	1.66	0.92	1.26	1.12	1.08	1.01
3.5	1.79	2.27	1.95	2.48	1.83	1.94	1.94	2.32	2.37	1.69	0.99	1.32	1.22	1.20	1.20
3.6	1.82	2.33	1.92	2.56	1.96	2.16	1.88	2.19	2.70	1.92	0.65	1.07	0.91	0.82	0.82
3.7	1.82	2.38	1.90	2.51	2.05	2.17	2.20	2.54	2.81	2.06	0.71	1.18	0.96	0.87	0.83
3.8	1.80	2.33	2.01	2.63	2.09	2.25	2.21	2.63	2.67	1.87	0.78	1.12	1.04	0.98	1.00
3.9	2.09	2.68	2.23	2.92	2.20	2.34	2.30	2.74	3.10	2.21	0.89	1.18	1.28	1.21	1.23
4.0	1.93	2.54	2.18	2.93	2.27	2.48	2.28	2.69	3.02	2.08	0.84	1.27	1.10	1.03	1.05
4.1	2.01	2.61	2.28	2.92	2.24	2.35	2.55	3.00	2.95	2.16	0.79	1.33	1.12	1.11	1.13
4.2	2.13	2.80	2.15	2.72	2.40	2.65	2.33	2.74	2.82	2.05	0.91	1.22	1.16	1.09	1.10
4.3	2.00	2.69	2.88	3.42	2.26	2.57	2.51	2.93	3.58	2.75	1.47	1.94	1.66	1.78	1.73
4.4	2.21	2.86	2.30	2.99	2.32	2.57	2.50	2.99	3.07	2.26	0.91	1.33	1.15	1.08	1.05
4.5	2.29	3.03	2.66	3.38	2.29	2.44	2.90	3.37	3.96	3.13	1.38	2.02	1.60	1.63	1.59
4.6	2.18	2.91	2.64	3.23	2.59	2.74	2.83	3.39	3.19	2.40	0.92	1.33	1.20	1.18	1.18
4.7	2.39	3.12	2.21	2.97	2.64	2.77	2.52	3.07	3.17	2.30	1.96	2.21	2.05	2.00	2.01
4.8	2.34	3.14	2.53	3.25	2.58	2.92	2.87	3.42	3.87	2.90	0.93	1.63	1.25	1.27	1.23
4.9	2.47	3.20	2.64	3.48	2.81	2.97	2.87	3.46	3.49	2.53	1.03	1.53	1.49	1.37	1.42
5.0	2.30	3.04	2.50	3.28	2.73	2.85	2.89	3.36	3.56	2.64	0.87	1.41	1.23	1.19	1.19
Avgs	1.23	1.62	1.34	1.73	1.38	1.50	1.44	1.70	1.84	1.34	0.56	0.81	0.73	0.70	0.69

(b) Reprojection based method

Table 4. Results for all the σ's values for setup #1.

[9] Noam Levi and Michael Werman. The viewing graph. In *International Conference on Computer Vision and Pattern Recognition (CVPR).*, volume 1, pages I–I, 2003. 1

[10] Quang-Tuan Luong and Thierry Viéville. Canonical Representations for the Geometries of Multiple Projective Views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996. 1

[11] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003. 1

[12] Ludovic Magerand and Alessio Del Bue. Revisiting Projective Structure from Motion: A Robust and Efficient Incremental Solution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):430–443, 2020. 1

[13] Daniel Martinec and Tomás Pajdla. 3D reconstruction by fitting low-rank matrices with missing data. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 198–205 vol. 1, 2005. 1

[14] Antonella Nardi, Dario Comanducci, and Carlo Colombo. Augmented Vision: Seeing beyond Field of View and Occlusions via Uncalibrated Visual Transfer from Multiple Viewpoints. In *2011 Irish Machine Vision and Image Processing Conference*, pages 38–44, 2011 (Best paper award). 1

[15] David Nistér. Untwisting a Projective Reconstruction. *International Journal of Computer Vision*, 60(2):165–183, 2004. 1

[16] John Oliensis and Venu Govindu. An experimental study of projective structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):665–671, 1999. 1

[17] Charlie Rothwell, Olivier Faugeras, and Gabriella Csurka. A Comparison of Projective Reconstruction Methods for Pairs of Views. *Computer Vision and Image Understanding*, 68(1):37–58, 1997. 4

[18] Alessandro Rudi, Matia Pizzoli, and Fiora Pirri. Linear solvability in the viewing graph. In *Asian Conference on Computer Vision (ACCV)*, 2010. 1

[19] Gilbert Strang. *Introduction to linear algebra*. Wellesley-Cambridge, fifth edition, 2016. 2

[20] Chris Sweeney, Torsten Sattler, Tobias Hollerer, Matthew Turk, and Marc Pollefeys. Optimizing the viewing graph for structure-from-motion. In *International Conference on Computer Vision (ICCV)*, December 2015. 1

[21] Roberto Toldo, Riccardo Gherardi, Michela Farenzena, and Andrea Fusiello. Hierarchical structure-and-motion recovery from uncalibrated images. *Computer Vision and Image Understanding*, 140:127–143, 2015. 1, 4

[22] Matthew Trager, Brian Osserman, and Jean Ponce. On the Solvability of Viewing Graphs. In *European Conference on Computer Vision (ECCV)*, pages 335–350. Springer, 2018. 1