On the Recovery of Cameras from Fundamental Matrices

Supplementary Material

In this document we provide additional information that was not included in the main paper due to space constraints.

A. Convergence Analysis

We now provide some considerations about convergence properties of our approach. Let us first recall our global minimization loss from Eq. (5) and (6):

$$\min_{\mathbf{p_1},\dots,\mathbf{p}_n} \Psi(\mathbf{p}_1,\dots,\mathbf{p}_n) = \min_{\mathbf{p_1},\dots,\mathbf{p}_n} \sum_{(i,j)\in\mathcal{E}} d(A_j\mathbf{p_i},\mathbf{0}). \tag{14}$$

As explained in Sec. 3, we handle this challenging problem by alternating minimizations restricted over entries of individual cameras $\mathbf{p_i}$, a technique known in literature as *Block Relaxation*. There are several theoretical studies on convergence properties of this scheme. Among them, here we follow [11] since it provides general results that require mild assumptions. Specifically, [11] relies on the *existence* assumption, i.e., that the global objective restricted over the current block of variables has a global minimizer over its feasible domain (although it does not need to be unique).

In our case, this translates into the fact that the cost restricted over a single camera (i.e., $\Psi(\bar{\mathbf{p_1}},..,\mathbf{p_i},..,\bar{\mathbf{p_n}})$) has a global minimizer. For the least squares null-space problem in Sec. 3.1 (LS-SVD), actually there exists a minimizer from the SVD decomposition of the matrix obtained by stacking the A_j matrices for each neighbor, but this is not unique[11]. Consequently, using the same reasoning as [11], we can conclude the following properties:

1. The local sub-problem – Eq. (8) – is solved optimally (via SVD) with respect to the current camera $\mathbf{p}_i$, hence the global objective Ψ is *non-increasing*. In other terms, if we denote the sequence of objective values over the iterations by $\Psi^{(k)}$ with $k = 1, 2, \dots$, then we have $\Psi^{(k)} \leq \Psi^{(k-1)}$. Since Ψ is *bounded* by 0, then the sequence of objective values is convergent (to some Ψ^{∞}), by the Monotone Convergence Theorem.
2. Concerning the sequence of variables, denoted by $\{\mathbf{p_1},\dots,\mathbf{p_n}\}^{(k)}$, note that the set of feasible cameras is *bounded* (due to the unit norm constraint), hence the Bolzano-Weierstrass Theorem guarantees the existence of a convergent sub-sequence: $\{\mathbf{p_1},\dots,\mathbf{p_n}\}^{\infty}$.
3. If we have a sub-sequence converging to some $\{\mathbf{p_1},\dots,\mathbf{p_n}\}^{\infty}$, then the corresponding sequence of objective values converges to $\Psi(\{\mathbf{p}_1\dots\mathbf{p}_n\}^{\infty})$ since Ψ is *continuous*. But all sub-sequences of a convergent sequence converge to the same point, hence we have $\Psi(\{\mathbf{p}_1\dots\mathbf{p}_n\}^{\infty}) = \Psi^{\infty}$.

To summarize, thanks to the above analysis, we have convergence of the objective values, but about the sequence of variables (cameras) we only know that it has one or more accumulation points, and that all accumulation points have the same function value. However, we cannot conclude that $\{\mathbf{p}_1\dots\mathbf{p}_n\}^{\infty}$ is a local/global minimum. The lack of local/global convergence results also involves other problems in Computer Vision based on a block relaxation framework (e.g., [12, 18]). However, this is an interesting area that we wish to explore in future work.

[11] If a vector $\mathbf{p_i^*}$ is the minimizer, then -$\mathbf{p_i^*}$ will also be a minimizer.

B. Initialization Approach

We now provide further details on the initialization of our approach (Step 1 of the iterative scheme in Sec. 3). We attempted different ideas: initializing all cameras to the Canonical Camera, to random 3×4 matrices, and to the output of other camera recovery methods (such as GPSFM [15], SINHA [28] and COLOMBO [9]). Among the variants of our approach, we select ANGLE-IT. We consider the same synthetic scenario as described in Sec. 4.2 and we analyze the effect of increasing noise on the different initialization methods, setting constant the other parameters: $n = 25, \rho = 0.4, \gamma = 0.1$. In this experiment, the viewing graph is covered by triplets. The result of this test is shown in Fig. 6 (Left): initializing all nodes with Canonical cameras, or random 3×4 matrices results in very large errors, showing that our approach necessitates initialization from other methods in practice. While initializing with SINHA and COLOMBO results in relatively lower errors than canonical/random cameras, their sensitivity to outliers make

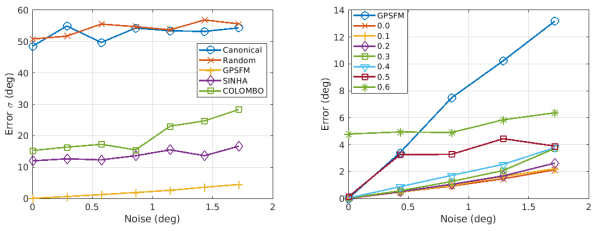

Figure 6. Left: error [degrees] in camera recovery achieved by our approach with different initialization methods, with increasing noise σ and other parameters fixed ($n = 25$, $\rho = 0.4$, $\gamma = 0.1$). Right: error [degrees] in camera recovery achieved by GPSFM and by our approach with increasing fraction of "missing" node initializations, for increasing noise and other parameters fixed ($n = 25$, $\rho = 0.4$, $\gamma = 0.0$).

them unusable in practice (they get nonzero errors for 0 noise due to outliers, see Fig. 4). We achieve the best performance from initialization with the output of GPSFM, which we choose as initialization for the experiments in Sec. 4.

Note that, by initializing with the output of GPSFM, it is possible that some cameras are not given an initial value: indeed, as already observed, GPSFM requires a graph covered by triplets and it is unable to estimate nodes not appearing in any triplet. Such nodes are initialized to the canonical cameras, as explained in Sec. 4.1. We analyze the effect of missing initialization for a subset of nodes in a synthetic environment, which simulates a real scenario when there are nodes in a viewing graph not covered by triplets. We test the effect of increasing fraction of nodes not initialized by GPSFM, with varying noise and other parameters constant ($n = 25, \rho = 0.4, \gamma = 0.0$). Results are plotted in Fig. 6 (Right), showing that our method outperforms GPSFM even when a significant amount of nodes are initialized to the canonical cameras. Results get slightly worse, still reasonable, when such amount comprises 60% of the cameras. This clearly shows that, even in the absence of a valid initialization for some nodes, our method is still effective in camera recovery, showing the potential in situations where the viewing graph is not entirely covered by triplets.

C. Direct Optimization

In addition to the methods presented in the main paper, we also attempt to estimate all the cameras $P_1 \dots P_n$ by directly solving the optimization problem from (14), given an initial estimate of the cameras. More precisely, we solve:

$$\min_{\mathbf{p_1},\dots,\mathbf{p}_n} \sum_{(i,j)\in\mathcal{E}} ||A_j \mathbf{p_i}||_2^2 \tag{15}$$

for A_j as defined in (4). We solve (15) using the interior-point method based Matlab solver *fmincon*, with $\mathbf{p_1^{(0)}} \dots \mathbf{p_n^{(0)}}$ initialized, as our methods are, from the output of GPSFM . We tested this method, referred to as *Fmincon*, in the same synthetic set up as in Fig. 3 (left), and Fig. 4 (right), for its sensitivity to noise and execution times respectively, with the results shown in Fig. 7. While *Fmincon* performs reasonably regarding sensitivity to noise, it is still worse than our method (ANGLE-IT) in terms of accuracy, and requires significantly higher execution times. We also tested the case where a fraction of the cameras are not initialized, but *Fmincon* failed to provide reasonable results, which renders it incapable of handling general graphs.

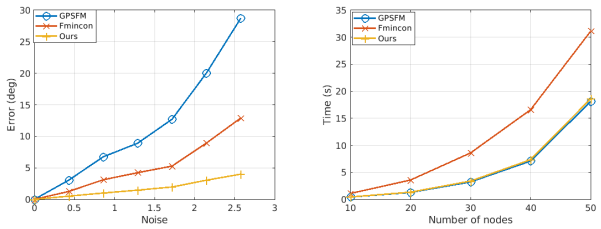

Figure 7. Left: Error [degrees] in camera recovery with increasing noise σ, with other parameters fixed ($n = 25$, $\rho = 0.4$, $\gamma = 0$). Right: execution times [seconds] for increasing number of cameras n, with other parameters fixed ($\sigma = 0.015$ rad, $\rho = 0.4$, $\gamma = 0.0$). The viewing graph is covered by triplets in these experiments. Results for our method, GPSFM and *Fmincon* are reported.