

A Subspace-Constrained Tyler’s Estimator and its Applications to Structure from Motion *

Feng Yu [†]
University of Minnesota
fyu@umn.edu

Teng Zhang
University of Central Florida
teng.zhang@ucf.edu

Gilad Lerman [‡]
University of Minnesota
lerman@umn.edu

Abstract

We present the subspace-constrained Tyler’s estimator (STE) designed for recovering a low-dimensional subspace within a dataset that may be highly corrupted with outliers. STE is a fusion of the Tyler’s M-estimator (TME) and a variant of the fast median subspace. Our theoretical analysis suggests that, under a common inlier-outlier model, STE can effectively recover the underlying subspace, even when it contains a smaller fraction of inliers relative to other methods in the field of robust subspace recovery. We apply STE in the context of Structure from Motion (SfM) in two ways: for robust estimation of the fundamental matrix and for the removal of outlying cameras, enhancing the robustness of the SfM pipeline. Numerical experiments confirm the state-of-the-art performance of our method in these applications. This research makes significant contributions to the field of robust subspace recovery, particularly in the context of computer vision and 3D reconstruction.

1. Introduction

In many applications, data has been collected in large quantities and dimensions. It is a common practice to represent such data within a low-dimensional subspace that preserves its essential information. Principal Component Analysis (PCA) is frequently employed to identify this subspace. However, PCA faces challenges when dealing with data contaminated by outliers. Consequently, the field of Robust Subspace Recovery (RSR) aims to develop a framework for outlier-robust PCA. RSR is particularly relevant to problems in computer vision, such as fundamental matrix estimation, which involves recovering a hidden subspace associated with “good correspondence pairs” among highly corrupted measurements.

Various algorithms have been proposed to address RSR, employing methods such as projection pursuit [1, 5, 14, 23, 29, 35, 38], subspace energy minimization (in particular least absolute

deviations and its relaxations) [9, 24, 27, 33, 34, 42, 43, 51], robust covariance estimation [50], filtering-based methods [4, 8, 49] and exhaustive subspace search methods [10, 16]. An in-depth exploration and comprehensive overview of robust subspace recovery and its diverse algorithms can be found in [25].

Methods based on robust covariance estimators, such as the Tyler’s M-estimator (TME), offer additional useful information on the shape of the data within the subspace, similarly to PCA in the non-robust setting. They also offer maximum-likelihood interpretation, which is missing in many other methods. Application of the TME [47] to RSR has been shown to be successful on basic benchmarks [25, 50]. Moreover, under a model of inliers in a general position on a subspace and outliers in general position in the complement of the subspace, TME was shown to recover the subspace within a desirable fraction of inliers [50]. Below this fraction it was proved to be Small Set Expansion (SSE) hard to solve the RSR problem [16].

One may still succeed with solving the RSR problem with a computationally efficient algorithm when the fraction of inliers is lower than the one required by [16], considering a more restricted data model or violating other assumptions made in [16]. For example, some special results in this direction are discussed in [32]. Also, [33] proposes the generalized haystack model of inliers and outliers to demonstrate the possibility of handling lower fractions of inliers by an RSR algorithm. This model extends the limited standard haystack model [27], where basic methods (such as PCA filtering) can easily work with low fractions of outliers. Nevertheless, it is unclear how practical the above theoretical ideas are for applied settings.

One practical setting that requires a fraction of inliers significantly lower than the one stated in [16] arises in the problem of robust fundamental (or essential) matrix estimation. The fundamental matrix encompasses the epipolar geometry of two views in stereo vision systems. It is typically computed using point correspondences between the two projected images. This computation requires finding an 8-dimensional subspace within a 9-dimensional ambient space. In this setting, the theoretical framework of [16] requires that the fraction of inliers be at least $8/9 \approx 88.9\%$, which is clearly unreasonable to require.

To date, the RANDOM Sample Consensus (RANSAC)

*This work was supported by NSF DMS awards 2124913 and 2318926.

[†]Supplementary code: <https://github.com/alexfengg/STE>

[‡]Corresponding author. All authors equally contributed.

method [10] is the only RSR method that has been highly successful in addressing this nontrivial scenario, gaining widespread popularity in computer vision. RANSAC is an iterative method that randomly selects minimal subsets of the data and fits models, in particular subspaces, to identify the best consensus set, that is, the set in most agreement with the hypothesized model. There are numerous approaches proposed to improve RANSAC, especially for this particular application, including locally optimized RANSAC (LO-RANSAC, [6]), maximum likelihood estimator RANSAC (MLE-SAC) [45], degeneracy-check enabled RANSAC (DEGENSAC) [7] and M-estimator guided RANSAC (MAGSAC) [2]). A near recovery theory for a variant of RANSAC under some assumptions on the outliers was suggested in [32]. Nevertheless, in general, RANSAC is rather slow and its application to higher-dimensional problems is intractable.

This work introduces a novel RSR algorithm that is guaranteed to robustly handle a lower fraction of outliers than the theoretical threshold proposed by [16], under special settings. Our basic idea is to adapt Tyler’s M-Estimator to utilize the information of the underlying d -dimensional subspace, while avoiding estimation of the full covariance. By using less degrees of freedom we obtain a more accurate subspace estimator than the one obtained by TME with improved computational complexity. We show that STE is a fusion of the Tyler’s M-estimator (TME) and a variant of the fast median subspace (FMS) [24] that aims to minimize a subspace-based ℓ_0 energy.

Our theory shows that our proposed subspace-constrained Tyler’s estimator (STE) algorithm can effectively recover the underlying subspace, even when it contains a smaller fraction of inliers relative to other methods. We obtain this nontrivial achievement first in a generic setting, where we establish when an initial estimator for STE is sufficiently well-conditioned to guarantee the desired robustness of STE. We then assume the asymptotic generalized haystack model and show that under this model, TME itself is a well-conditioned initial estimator for STE, and that unlike TME, STE with this initialization can deal with a lower fraction of inliers than the theoretical threshold specified in [16].

We demonstrate competitive performance in robust fundamental matrix estimation, relying solely on subspace information without additional methods for handling degenerate scenarios, in contrast to [7, 12, 37]. We also propose a potential application of RSR for removing bad cameras in order to enhance the SfM pipeline and show competitive performance of STE. This is a completely new idea and it may require additional exploration to make it practical. Nevertheless, it offers a very different testbed where $N = D$ is very large and RANSAC is generally intractable.

The rest of the paper is organized as follows: §2 introduces the STE framework, §3 establishes theoretical guarantees of STE, §4 applies STE to two different problems in SfM, demonstrating its competitive performance relative to existing algorithms, and §5 provides conclusions and future directions.

2. The STE Algorithm

We present our proposed STE. We review basic notation in §2.1 and Tyler’s original estimator in §2.2. We describe our method in §2.3, its computational complexity in §2.4, its algorithmic choices in §2.5 and an interpretation for it as a fusion of TME and FMS with $p = 0$ in §2.6.

2.1. Notation

We use bold upper and lower case letters for matrices and column vectors, respectively. Let \mathbf{I}_k denote the identity matrix in $\mathbb{R}^{k \times k}$, where if k is obvious we just write \mathbf{I} . For a matrix \mathbf{A} , we denote by $\text{tr}(\mathbf{A})$ and $\text{Im}(\mathbf{A})$ the trace and image (i.e., column space) of \mathbf{A} . We denote by $S_+(D)$ and $S_{++}(D)$ the sets of positive semidefinite and definite matrices in $\mathbb{R}^{D \times D}$, respectively. We denote by $O(D, d)$ the set of semi-orthogonal $D \times d$ matrices, i.e., $\mathbf{U} \in O(D, d)$ if and only if $\mathbf{U} \in \mathbb{R}^{D \times d}$ and $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_d$. We refer to linear d -dimensional subspaces as d -subspaces. For a d -subspace L , we denote by \mathbf{P}_L the $D \times D$ matrix representing the orthogonal projector onto L . We also arbitrarily fix \mathbf{U}_L in $O(D, d)$ such that $\mathbf{U}_L \mathbf{U}_L^\top = \mathbf{P}_L$ (such \mathbf{U}_L is determined up to right multiplication by an orthogonal matrix in $O(d, d)$). Throughout the paper, $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^D$ is assumed to be a given centered dataset, that is, $\sum_{i=1}^N \mathbf{x}_i = \mathbf{0}$.

2.2. Tyler’s Estimator and its Application to RSR

Tyler’s M-estimator (TME) [47] robustly estimates the covariance Σ^* of the dataset $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^D$ by minimizing

$$\frac{D}{N} \sum_{i=1}^N \log(\mathbf{x}_i^\top \Sigma^{-1} \mathbf{x}_i) + \log \det(\Sigma) \quad (1)$$

over $\Sigma \in S_{++}(D)$ such that $\text{tr}(\Sigma) = 1$. The cost function in (1) can be motivated by writing the maximum likelihood of the multivariate t -distribution and letting its degrees of freedom parameter, ν , approach zero [31]. This cost function is invariant to dilations of Σ , and the constraint on $\text{tr}(\Sigma)$, whose value can be arbitrarily chosen, fixes a scale. TME also applies to scenarios where the covariance matrix does not exist. In such cases, TME estimates the shape (or scatter matrix) of the distribution, which is defined up to an arbitrary scale. More direct interpretations of TME as a maximum likelihood estimator can be found in [11, 46]. When D is fixed and N approaches infinity, TME is the “most robust” estimator of the shape matrix for data i.i.d. sampled from a continuous elliptical distribution [47] in a minimax sense, that is, as a minimizer of the maximal variance.

Tyler [47] proposed the following iterative formula for computing TME:

$$\Sigma^{(k)} = \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\mathbf{x}_i^\top (\Sigma^{(k-1)})^{-1} \mathbf{x}_i} / \text{tr} \left(\sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\mathbf{x}_i^\top (\Sigma^{(k-1)})^{-1} \mathbf{x}_i} \right).$$

Kent and Tyler [22] proved that if any d -subspace of \mathbb{R}^D , where $1 \leq d \leq D - 1$, contains fewer than Nd/D data points, then

the above iterative procedure converges to TME. Linear rate of convergence was proved for the regularized TME in [15] and for TME in [13].

One can apply the TME estimator to solve the RSR problem with a given dimension d by forming the subspace spanned by the top d eigenvectors of TME. Zhang [50] proved that as long as there are more than Nd/D inliers lying on a subspace, and the projected coordinates of these inliers on the d -subspace and the projected coordinates of the outliers on the $(D-d)$ -dimensional orthogonal complement of the subspace are in general position, then TME recovers this subspace. Zhang [50] also showed that in this setting the above iterative formula converges (note that the condition of convergence in [22] does not apply in this case). The above lower bound of Nd/D on the number of inliers coincides with the general bound for the noiseless RSR problem, beyond which the problem becomes Small Set-Expansion (SSE) hard [16].

Numerical experiments in [50] and [25] indicated state-of-the-art accuracy of TME compared to other RSR algorithms in various settings. The computational complexity of TME is of order $O(K(ND^2 + D^3))$, where K is the number of iterations. On the other hand, the cost of faster RSR algorithms is of order $O(KNDd)$ [24, 25, 33].

2.3. Motivation and Formulation of STE

We aim to use more cleverly the d -subspace information within the TME framework to form an RSR algorithm, instead of first estimating the full covariance. By using less degrees of freedom we can obtain a more accurate subspace estimator, especially when the fraction of outliers can be large. Furthermore, our idea allows us to improve the computational cost to become state-of-the-art for high-dimensional settings.

Many RSR algorithms can be formulated as minimizing a best orthogonal projector onto a d -subspace [24, 25, 27, 33, 51]. We are going to do something similar, but unlike using an orthogonal projector, we will still use information from TME to get the shape of the data on the projected subspace. We will make the rest of the eigenvalues (i.e., bottom $D-d$ ones) equal and shrink them by a parameter $0 < \gamma < 1$. We thus use a regularized version of a reduced-dimension covariance matrix. This parameter γ plays a role in our theoretical estimates. Making γ smaller helps with better subspace recovery, whereas making γ bigger enhances the well-conditioning of the estimator.

Following these basic ideas, we formulate our method, STE. For simplicity, we utilize covariance matrices and their inverses. Since these covariance matrices are essentially d -dimensional and include an additional simple regularizing component, our overall computations can be expressed in terms of the computation of the top d singular values of an $N \times D$ matrix (see §2.4).

At iteration k we follow a similar step to that of TME:

$$\mathbf{Z}^{(k)} := \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^\top / (\mathbf{x}_i^\top (\boldsymbol{\Sigma}^{(k-1)})^{-1} \mathbf{x}_i).$$

We compute the eigenvalues $\{\sigma_i\}_{i=1}^D$ of $\mathbf{Z}^{(k)}$ and replace each of the bottom $(D-d)$ of them with $\gamma \cdot \sigma_{d+1,D}$, where

$$\sigma_{d+1,D} := \frac{1}{D-d} \sum_{i=d+1}^D \sigma_i. \quad (2)$$

We also compute the eigenvectors of $\mathbf{Z}^{(k)}$ and form the matrix $\boldsymbol{\Sigma}^{(k)}$ with the same eigenvectors as those of $\mathbf{Z}^{(k)}$ and the modified eigenvalues, scaled to have trace 1. We iteratively repeat this procedure until the two estimators are sufficiently close. Algorithm 1 summarizes this procedure. Note that it is invariant to scaling of the data, similarly to TME.

Algorithm 1 STE: Subspace-Constrained Tyler's Estimator

- 1: **Input:** $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$: centered data matrix, d : subspace dimension, K : maximum number of iterations, τ, γ : parameters.
 - 2: **Output:** L : d -subspace in \mathbb{R}^D
 - 3: $\boldsymbol{\Sigma}^{(0)} = \mathbf{I}_D / D$
 - 4: **for** $k = 1, 2, \dots$ **do**
 - 5: $\mathbf{Z}^{(k)} \leftarrow \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^\top / (\mathbf{x}_i^\top (\boldsymbol{\Sigma}^{(k-1)})^{-1} \mathbf{x}_i)$
 - 6: $[\mathbf{U}^{(k)}, \mathbf{S}^{(k)}, \mathbf{V}^{(k)}] \leftarrow \text{EVD}(\mathbf{Z}^{(k)})$
 - 7: $\sigma_i \leftarrow [\mathbf{S}^{(k)}]_{ii}$ and $\sigma_{d+1,D} \leftarrow \sum_{i=d+1}^D \sigma_i / (D-d)$
 - 8: $\tilde{\mathbf{S}}^{(k)} \leftarrow \text{diag}(\sigma_1, \dots, \sigma_d, \gamma \cdot \sigma_{d+1,D}, \dots, \gamma \cdot \sigma_{d+1,D})$,
 - 9: $\boldsymbol{\Sigma}^{(k)} \leftarrow \mathbf{U}^{(k)} \tilde{\mathbf{S}}^{(k)} (\mathbf{U}^{(k)})^\top / \text{tr}(\mathbf{U}^{(k)} \tilde{\mathbf{S}}^{(k)} (\mathbf{U}^{(k)})^\top)$
 - 10: Stop if $k \geq K$ or $\|\boldsymbol{\Sigma}^{(k)} - \boldsymbol{\Sigma}^{(k-1)}\|_F < \tau$
 - 11: **end for**
 - 12: $L = \text{Span of the first } d \text{ columns of } \mathbf{U}^{(k)}$
-

2.4. Computational Complexity

Setting $w_i^{(k)} = (\mathbf{x}_i^\top (\boldsymbol{\Sigma}^{(k-1)})^{-1} \mathbf{x}_i)^{-1}$, we can express $\mathbf{Z}^{(k)}$ as $\mathbf{Z}^{(k)} = \tilde{\mathbf{X}} \tilde{\mathbf{X}}^\top$, where $\tilde{\mathbf{X}} = [(w_1^{(k)})^{1/2} \mathbf{x}_1, \dots, (w_N^{(k)})^{1/2} \mathbf{x}_N]$. With some abuse of notation we denote by $\sigma_1, \dots, \sigma_D$ the eigenvalues of $\boldsymbol{\Sigma}^{(k-1)}$ (and not $\boldsymbol{\Sigma}^{(k)}$). Since they are scaled to have trace 1, $\sigma_{d+1,D} = (1 - \sum_{j=1}^d \sigma_j) / (D-d)$. We thus only need the top d eigenvectors and top d eigenvalues of $\boldsymbol{\Sigma}^{(k-1)}$ to update $\tilde{w}_i^{(k)}$. Therefore, the complexity of STE can be of order $O(KNDd)$ if a special fast algorithm is utilized for computing only the top d eigenvectors.

2.5. Implementation Details

STE depends on the parameters K , τ and γ and the initialization of $\boldsymbol{\Sigma}^{(0)}$. The first two parameters are rather standard in iterative procedures and do not raise any concern.

Our theory sheds some light on possible choices of γ and in particular it indicates that the algorithm can be more sensitive to choices of γ when the quantity defined later in (3) is relatively small. In this case, it may be beneficial to try several values of γ . We propose here a constructive way of doing it. We first form a sequence of $0 < \gamma \leq 1$, e.g., $\gamma_k = 1/k$, $k = 1, \dots, m$. In order to

determine the best choice of γ , we compute the distance of each data point \mathbf{x} to each subspace L_k , corresponding to the choice of γ_k , where $\text{dist}(\mathbf{x}, L_k) = \|\mathbf{x} - \mathbf{P}_{L_k} \mathbf{x}\|$. We set a threshold ζ , obtained by the median among all points and all subspaces and for each subspace, L_k , we count the number of the inliers with distance below this threshold. The best γ_k is determined according to the subspace yielding the largest number of inliers. We describe this procedure in Algorithm 2.

For simplicity, we initialize with $\Sigma^{(0)} = \mathbf{I}_D/D$ and note that with this choice $\Sigma^{(1)}$ reflects the trimmed covariance matrix and thus reflects the PCA subspace. One can also initialize with TME or other subspaces (see §3 where the theory of STE is discussed). One can further try several initialization (with possible random components) and use a strategy similar to Algorithm 2 to choose the best one.

At last, we remark that when computing $\mathbf{Z}^{(k)}$ we want to ensure that $\mathbf{x}_i^\top (\Sigma^{(k-1)})^{-1} \mathbf{x}_i$ cannot be zero and we thus add the arbitrarily small number 10^{-15} to this value.

Algorithm 2 Estimating best γ for STE

- 1: **Input:** $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$: centered data matrix, d : subspace dimension, $\{\gamma_1, \dots, \gamma_m\}$: a set of pre-selected γ 's.
 - 2: **Output:** γ^* : optimal γ among $\{\gamma_1, \dots, \gamma_m\}$
 - 3: **for** $j = 1, 2, \dots, m$ **do**
 - 4: $L^{(j)} \leftarrow \text{STE}(\mathbf{X}, d, \gamma_j)$
 - 5: $\mathcal{D}^{(j)} \leftarrow \{\text{dist}(\mathbf{x}_i, L^{(j)}) \mid \mathbf{x}_i \in \mathcal{X}\}$.
 - 6: **end for**
 - 7: Set $\zeta = \text{median}(\{\mathcal{D}^{(1)}, \dots, \mathcal{D}^{(m)}\})$
 - 8: $j^* = \arg\max_{1 \leq j \leq m} |\mathcal{D}^{(j)}| < \zeta$
 - 9: $\gamma^* = \gamma_{j^*}$
-

2.6. STE fuses TME and a Variant of FMS

STE is formally similar to both TME and FMS. Indeed, at each iteration these algorithms essentially compute $\Sigma^{(k+1)} = \sum_{i=1}^N w_i \mathbf{x}_i \mathbf{x}_i^\top$, where $w_i \equiv w_i(\Sigma^{(k)})$. We summarize the formal weights for FMS (with any choice of p for minimizing an ℓ_p energy in [24]), TME and STE. We ignore an additional scaling constant for TME and STE, obtained by dividing $w_i \mathbf{x}_i \mathbf{x}_i^\top$ above by its trace, and a regularization parameter δ for FMS. We express these formulas using the eigenvalues $\sigma_1, \dots, \sigma_D$ and eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_D$ of the weighted sample covariance, $\sum_{i=1}^N w_i \mathbf{x}_i \mathbf{x}_i^\top$ for each method and $\beta := \gamma \cdot \sigma_{d+1, D}$ (see (2)) as follows:

$$w_i^{\text{FMS}} = \frac{1}{\left(\sum_{j=d+1}^D (\mathbf{x}_i^\top \mathbf{u}_j)^2 \right)^{(2-p)/2}},$$

$$w_i^{\text{TME}} = \frac{1}{\sum_{j=1}^D \sigma_j^{-1} (\mathbf{x}_i^\top \mathbf{u}_j)^2},$$

$$w_i^{\text{STE}} = \frac{1}{\sum_{j=1}^d \sigma_j^{-1} (\mathbf{x}_i^\top \mathbf{u}_j)^2 + \beta^{-1} \sum_{j=d+1}^D (\mathbf{x}_i^\top \mathbf{u}_j)^2}.$$

These weights aim to mitigate the impact of outliers in different ways. For FMS, $\sum_{j=d+1}^D (\mathbf{x}_i^\top \mathbf{u}_j)^2$ is the squared distance of a data point \mathbf{x}_i to the subspace L . Thus for $p < 2$, w_i^{FMS} is smaller for “subspace-outliers”, where the robustness to such outliers increases when $p \geq 0$ decreases.

The weights of TME are inversely proportional to the squared Mahalanobis distance of \mathbf{x}_i to the empirical distribution. They mitigate the effect of “covariance-outliers”. If the dataset is concentrated on a k -subspace where $k < d$, then TME can provide smaller weights to points lying away from this subspace, unlike FMS that does not distinguish between points within the larger d -subspace.

We note that the weights of STE fuse the above two weights. Within a d -subspace, they use the shape of the data. They can thus avoid outliers within this d -subspace. Within the orthogonal component of this subspace, they use a term proportional to that of FMS with $p=0$. We remark that such ℓ_0 minimization has a clear interpretation for RSR, though is generally hard to guarantee. Indeed, [24] has no guarantees for FMS with $p=0$. It can also yield unwanted spurious stationary points [26].

3. Theory

We review a theoretical guarantee for STE, whose proof is given in [28]. It requires some conditions and we verify they hold with high probability under the asymptotic generalized haystack model. We assume a noiseless inliers-outliers RSR model. Let L_* denote the underlying d -subspace in \mathbb{R}^D , $\mathcal{X}_{in} = \mathcal{X} \cap L_*$ and $\mathcal{X}_{out} = \mathcal{X} \setminus \mathcal{X}_{in}$ be the set of inliers and outliers, respectively, and $n_1 = |\mathcal{X}_{in}|$ and $n_0 = |\mathcal{X}_{out}|$ be the number of inliers and outliers. Our first assumption is a mild one on how well-conditioned the inliers are in L_* (compare e.g., other assumptions in [25, 32]).

Assumption 1: Any k -subspace of L_* , $1 \leq k \leq d$, contains at most $n_1 k/d$ points.

Motivation for Assumption 2: The ratio of inliers per outliers, n_1/n_0 , in RSR is often referred to as the SNR (signal-to-noise ratio) [25, 32, 33]. The smaller it is, the best the subspace recovery is. We define the dimension-scaled SNR (DS-SNR) as the SNR obtained when scaling n_1 and n_0 by their respective dimensions (of L_* and L_*^\perp):

$$\text{DS-SNR} := \frac{n_1/d}{n_0/(D-d)}. \quad (3)$$

Zhang [50] showed that exact recovery by TME is guaranteed whenever $\text{DS-SNR} > 1$ (assuming general position assumptions on the inliers and outliers) and Hardt and Moitra [16] showed that when considering general datasets with general position assumptions on the inliers and outliers, the RSR problem is SSE hard if the DS-SNR is lower than 1. We aim to show that under the following weaker generic condition, STE can obtain exact recovery with DS-SNR, strictly lower than 1.

Assumption 2: $\text{DS-SNR} > \gamma$, where $\gamma < 1$ is the STE parameter.

Our last assumption requires a sufficiently good initialization for STE, but also implicitly involves additional hidden

assumptions on the inliers and outliers. This is expected, since Assumption 1 does not require anything from the outliers and also has a very weak requirement from the inliers. To formulate the new assumption we define below some basic condition numbers for good initialization (which are more complicated than the one for initialization by PCA suggested by [33] and [32]) and also quantities similar to the ones used to guarantee landscape stability in the theory of RSR [25, 27, 33, 51].

Definitions required for Assumption 3: Recall that $\Sigma^{(0)}$ denotes the initial value in Algorithm 1, and denote

$$\Sigma_{L_1, L_2}^{(0)} = \mathbf{U}_{L_1}^\top \Sigma^{(0)} \mathbf{U}_{L_2}.$$

We define the following condition number

$$\kappa_1 = \frac{\sigma_d \left(\Sigma_{L_*, L_*}^{(0)} - \Sigma_{L_*, L_*^\perp}^{(0)} \Sigma_{L_*^\perp, L_*^\perp}^{(0)-1} \Sigma_{L_*^\perp, L_*}^{(0)} \right)}{\sigma_1 \left(\Sigma_{L_*^\perp, L_*^\perp}^{(0)} \right)}.$$

To get a better intuition to this primary quantity of Assumption 3, we first express the initial estimator $\Sigma^{(0)}$, using basis vectors for L_* and L_*^\perp , as a 2×2 block matrix

$$\begin{pmatrix} \Sigma_{L_*, L_*}^{(0)} & \Sigma_{L_*, L_*^\perp}^{(0)} \\ \Sigma_{L_*^\perp, L_*}^{(0)} & \Sigma_{L_*^\perp, L_*^\perp}^{(0)} \end{pmatrix}.$$

Defining $\Sigma' = \Sigma_{L_*, L_*^\perp}^{(0)} \Sigma_{L_*^\perp, L_*^\perp}^{(0)-1} \Sigma_{L_*^\perp, L_*}^{(0)}$, we decompose this block matrix as

$$\begin{pmatrix} \Sigma' & \Sigma_{L_*, L_*^\perp}^{(0)} \\ \Sigma_{L_*^\perp, L_*}^{(0)} & \Sigma_{L_*^\perp, L_*^\perp}^{(0)} \end{pmatrix} + \begin{pmatrix} \Sigma_{L_*, L_*}^{(0)} - \Sigma' & 0 \\ 0 & 0 \end{pmatrix}.$$

We note that the numerator of κ_1 is the d -th eigenvalue of the second matrix in the above sum. We show in [28] that this eigenvalue is positive if $\Sigma^{(0)}$ is positive definite, which can be easily enforced. The condition number is thus the ratio between the smallest positive eigenvalue of the second matrix of the sum and the largest eigenvalue of the component of the first matrix associated with L_*^\perp . Therefore, κ_1 expresses a ratio between a quantifier of a d -dimensional component of $\Sigma^{(0)}$, associated with L_* , and a quantifier of the projection onto L_*^\perp of a full rank component of $\Sigma^{(0)}$.

We also define $\Sigma_{in,*}$ as the TME solution to the set of the projected inliers $\{\mathbf{U}_{L_*} \mathbf{x} \mid \mathbf{x} \in \mathcal{X}_{in}\} \subset \mathbb{R}^d$ and the following two condition numbers

$$\kappa_2 = \frac{\sigma_1 \left(\Sigma_{L_*^\perp, L_*^\perp}^{(0)} \right)}{\sigma_D \left(\Sigma^{(0)} \right)} \quad \text{and} \quad \kappa_{in} = \frac{\sigma_1 \left(\Sigma_{in,*} \right)}{\sigma_d \left(\Sigma_{in,*} \right)}.$$

We note that κ_{in} is analogous to the condition number in (25) of [32], where we replace the sample covariance by the TME estimator. An analog to the alignment of outliers statistic [27, 33] for STE is

$$\mathcal{A} = \left\| \sum_{\mathbf{x} \in \mathcal{X}_{out}} \frac{\mathbf{x} \mathbf{x}^\top}{\|\mathbf{U}_{L_*^\perp} \mathbf{x}\|^2} \right\|.$$

An analog to the stability statistic [27, 33] for STE is

$$\mathcal{S} = \sigma_{d+1,D} \left(\sum_{\mathbf{x} \in \mathcal{X}} \frac{\mathbf{x} \mathbf{x}^\top}{\|\mathbf{x}\|^2} \right),$$

where $\sigma_{d+1,D}(\mathbf{X})$ was defined in (2).

Assumption 3: There exists $C = C(\gamma, \text{DS-SNR}) > 0$ such that

$$\kappa_1 \geq C \frac{d \kappa_{in} \mathcal{A}}{n_1} \left(\kappa_{in} + \frac{\mathcal{A}}{\frac{n_1}{d} - \gamma \frac{n_0}{D-d}} + \frac{\kappa_2 \mathcal{A}}{\gamma \mathcal{S}} (1 + \kappa_{in}) \right). \quad (4)$$

The exact technical requirement on C is specified in [28]. In general, the larger the RHS of (4), the more restricted the choice of $\Sigma^{(0)}$ is. In particular, when $\kappa_1 = \infty$, the definition of κ_1 implies that $\text{Im}(\Sigma^{(0)}) = L_*$, so the subspace is already recovered by the initial estimate. Therefore, reducing the lower bound of κ_1 may allow some flexibility, so a marginally suboptimal initialization could still work out. In [28], we show that under the asymptotic generalized haystack model, Assumption 3 can be interpreted as an upper bound on the largest principal angle between the initial and ground truth subspaces.

Generic Theory: The next theorem suggests that under assumptions 1-3, STE nicely converges to an estimator that recovers L_* . The main significance of this theory is that its assumptions can allow DS-SNR lower than 1 for special instances of datasets (for which the assumptions hold), unlike the general recovery theories of [16] and [50].

Theorem 1. *Under assumptions 1-3, the sequence $\Sigma^{(k)}$ generated by STE converges to $\mathbf{U}_{L_*} \Sigma_{in,*} \mathbf{U}_{L_*}^\top$, the TME solution for the set of inliers \mathcal{X}_{in} . In addition, let $L^{(k)}$ be the subspace spanned by the top d eigenvectors of $\Sigma^{(k)}$, then the angle between $L^{(k)}$ and L_* , $\angle(L^{(k)}, L_*) = \cos^{-1}(\|\mathbf{U}_{L^{(k)}}^\top \mathbf{U}_{L_*}\|)$, converges r -linearly to zero.*

We discuss insights of this theory on choices of the algorithms and further verify the above stated advantage of STE over TME assuming a common probabilistic model.

Choice of γ for subspace recovery: In order to avoid too large lower bound for κ_1 in (4), which we motivated above, it is good to find ϵ_1 and $\epsilon_2 > 0$, such that γ lies in $(\epsilon_1, \text{DS-SNR} - \epsilon_2)$ (to notice this, observe the terms involving γ in the denominators of the last two additive terms in (4)). We thus note that if the DS-SNR is expected to be sufficiently larger than 1, we can use, e.g., $\gamma = 0.5$, but when the DS-SNR can be close to 1 or lower (e.g., in fundamental matrix estimation), it is advisable to choose small values of γ according to Algorithm 2 and their sizes may depend on the expected value of the DS-SNR.

Possible ways of Initialization: If one expects an initial estimated subspace \hat{L} to have a sufficiently small angle θ with L_* , where $\theta = \angle(\hat{L}, L_*)$, then for $\Sigma^{(0)} := \Pi_{\hat{L}} + \epsilon \mathbf{I}$ it can be shown that $\kappa_1 > O(1/(\epsilon + \theta))$ and $\kappa_2 < O(1 + \frac{\theta}{\epsilon})$. Thus one may use a trusted RSR method, e.g., FMS. As discussed in §2.5, the choice $\Sigma^{(0)} = \mathbf{I}$ (or a scaled version of it) corresponds to \hat{L} being the

PCA subspace (obtained at iteration 1). Also, using the TME solution for $\Sigma^{(0)}$ corresponds to using the TME subspace as \hat{L} .

Theory under a probabilistic model: We show that under a common probabilistic model, the assumptions of Theorem 1, where $\Sigma^{(0)}$ is obtained by TME, hold. Moreover, we show that STE (initialized by TME) can recover the correct subspace in situations with DS-SNR < 1 , whereas TME cannot recover the underlying subspace in such cases. We follow [33] and study the Generalized Haystack Model, though for simplicity, we assume Gaussian instead of sub-Gaussian distributions and an asymptotic setting. We assume n_1 inliers i.i.d. sampled from a Gaussian distribution $N(0, \Sigma^{(in)}/d)$, where $\Sigma^{(in)} \in S_+(D)$ and $L_* = \text{Im}(\Sigma^{(in)})$ (so $\Sigma^{(in)}$ has d nonzero eigenvalues), and n_0 outliers are i.i.d. sampled from a Gaussian distribution $N(0, \Sigma^{(out)}/D)$, where $\Sigma^{(out)}/D \in S_{++}(D)$. We define the following condition numbers of inliers (in L_*) and outliers:

$$\kappa_{in} = \frac{\sigma_1(\Sigma^{(in)})}{\sigma_d(\Sigma^{(in)})} \quad \text{and} \quad \kappa_{out} = \frac{\sigma_1(\Sigma^{(out)})}{\sigma_D(\Sigma^{(out)})}.$$

Clearly, Assumption 1 holds under this model, and Assumption 2 constrains some of its parameters. Our next theorem shows that Assumption 3 holds under this model when the initial estimate $\Sigma^{(out)}$ for STE is obtained by TME. It also shows that in this case STE can solve the RSR problem even when DS-SNR < 1 , unlike TME. For simplicity, we formulate the theory for the asymptotic case, where $N \rightarrow \infty$ and the theorem holds almost surely. It is possible to formulate it for a very large N with high probability, but it requires stating complicated constants depending on various parameters.

Theorem 2. *Assume data generated from the above generalized haystack model. Assume further that for $0 < \mu < 1$, which can be arbitrarily small, $d < (1 - \mu)D - 2$. Then, for any chosen $0 < c_0 < 1$, which is a lower bound for γ , there exists $\eta := \eta(\kappa_{in}, \kappa_{out}, c_0, \mu) < 1$ such that if DS-SNR $\geq \eta$ and $\Sigma^{(0)}$ is obtained by TME, then Assumption 3 for $\Sigma^{(0)}$ is satisfied with $c_0 < \gamma < \eta - c_0$ almost surely as $N \rightarrow \infty$. Consequently, the output of the STE algorithm, initialized by TME and with the choice of $c_0 < \gamma < \eta - c_0$, recovers L_* . On the other hand, if $\Sigma_{L_*, L_*^\perp}^{(out)} \neq 0$ and DS-SNR < 1 , then the top d eigenvectors of TME do not recover L_* .*

There are three different regimes that the theorem covers. When DS-SNR ≥ 1 , both TME+STE (i.e., STE initialized by TME) and TME solve the RSR problem. When $\eta \leq \text{DS-SNR} < 1$, TME+STE solves the RSR problem and TME generally fails. When $\gamma \leq \text{DS-SNR} < \eta$, TME+STE might also fail, but STE with extremely good initialization (that satisfies Assumption 3) can still solve the problem.

To get a basic idea of the dependence of η on its parameters, we remark that $\eta \rightarrow 1$ if either $c_0 \rightarrow 0$, $\kappa_{in} \rightarrow \infty$, $\kappa_{out} \rightarrow \infty$ or $\mu \rightarrow 0$, where the parameter μ is somewhat artificial and might be removed with a tighter proof. Therefore, successful

performance of TME+STE requires a DS-SNR that is close to 1 when γ is close to either 0 or η (so that c_0 is very small) or when either the inlier or outlier distribution is highly non-symmetric, that is, either κ_{in} or κ_{out} is large.

4. Applications to Structure from Motion

We apply STE to problems relevant to SfM: robust estimation of fundamental matrices (see §4.1), and initial screening of undesirable cameras (see §4.2).

4.1. Robust Fundamental Matrix Estimation

Fundamental matrix estimation from noisy and inexact keypoint matches is a core computer vision problem. It provides a challenging setting for applying RSR methods.

We review this setting as follows. Let $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^3 \times \mathbb{R}^3$ be a correspondence pair of two points in different images that are projections of the same 3D point in the scene, where \mathbf{x} and \mathbf{x}' are expressed by homogeneous coordinates of planar points. The fundamental matrix $\mathbf{F} \in \mathbb{R}^{3 \times 3}$ relates these corresponding points and the epipolar lines they lie on as follows: $\mathbf{x}'^\top \mathbf{F} \mathbf{x} = 0$ [17], or equivalently,

$$\text{vec}(\mathbf{F}) \cdot \text{vec}(\mathbf{x}\mathbf{x}'^\top) = 0. \quad (5)$$

where $\text{vec}(\cdot)$ denotes the vectorized form of a matrix. Therefore, ideally, the set of all vectors in \mathbb{R}^9 of the form $\text{vec}(\mathbf{x}\mathbf{x}'^\top)$, where $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^3 \times \mathbb{R}^3$ is a correspondence pair, lies on an 8-subspace in \mathbb{R}^9 and its orthogonal complement yields the fundamental matrix. In practice, the measurements of correspondence pairs can be highly corrupted due to poor matching. Moreover, some choices of correspondence pairs and the corruption mechanism may lead to concentration on low-dimensional subspaces within the desired 8-subspace. Furthermore, the corruption mechanism can lead to nontrivial settings of outliers. Lastly, since $d = 8$ and $D = 9$, the theoretical threshold of [16] translates to having the fraction of inliers among all data points be at least $8/9 \approx 88.9\%$.

Therefore, this application is often a very challenging setting for direct RSR methods. The best performing RSR methods to date for fundamental matrix estimation are variants of RANSAC [10]. RANSAC avoids any subspace-modeling assumptions, but estimates the subspace based on testing myriads of samples, each having 7 or 8 point correspondences [17].

We test the performance of STE in estimating the fundamental matrix on the Photo Tourism database [41], where the image correspondences are obtained by SIFT feature similarities [30]. We compare STE with the following 3 top RSR performers according to [25]: FMS [24], spherical FMS (SFMS) [24] and TME [47, 50]. We also compared with vanilla RANSAC [10] and two of its specialized extensions, which are state-of-the-art performers for estimating fundamental matrices: locally optimized RANSAC (LO-RANSAC) [6] and degeneracy-check enabled RANSAC (DEGENSAC) [7]. For the RSR methods we used codes from the supplementary material of [25]

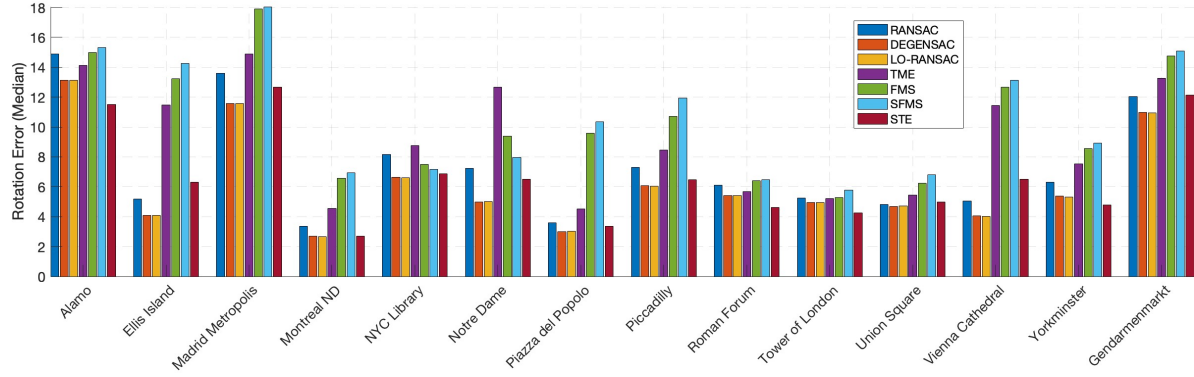


Figure 1. Median (relative) rotation errors obtained by seven algorithms for the 14 datasets of Photo Tourism.

with their default options. We further used the Python package `pydegensac` for implementing LO-RANSAC and DEGENSAC with the inlier threshold $\eta=0.75$. For STE, we used Algorithm 2 to estimate the best γ from $\{(2i)^{-1}\}_{i=1}^5$.

We measure the accuracy of the results according to the median and mean errors of relative rotation and direction vectors directly obtained by the fundamental matrices for each method. For computing these errors, we compared with ground-truth values provided by [41, 48]. Figure 1 describes the result of the mean errors for relative rotation per dataset of Photo Tourism, where the other three errors and $\text{mAA}(10^\circ)$ are in the supplemental material. STE is significantly better than top RSR performers (TME, FMS and SFMS). Overall, it appears that STE performs better than vanilla RANSAC, except for the Ellis Island and Vienna Cathedral datasets, where RANSAC outperforms STE. STE is still competitive when compared with LO-RANSAC and DEGENSAC, except for Notre Dame and the latter two datasets.

4.2. Initial Camera Removal for SfM

We propose a novel application of RSR for SfM and test STE for this application. Even though our framework is not sufficiently practical at this point, it allows testing STE in a different setting where $N=D$ is very large and $d=6$. Our idea is to use RSR within the SfM pipeline right after estimating the fundamental matrices, in order to remove some cameras that result in inaccurate estimated fundamental matrices. The hope is that eventually such methods may reduce corruption and speed up the costly later computationally intensive stages of the global SfM pipeline.

There are two main reasons to question such a process. One may first question the gain in improving accuracy. Indeed, since the rest of the pipeline already identifies corrupted pairwise measurements, this process may not improve accuracy and may even harm it as it removes whole cameras and not pairs of cameras. That is, it is possible that a camera, which results in bad pairwise measurement, also contributes to some other accurate pairwise estimates that can improve the overall accuracy. The second concern is in terms of speed. In general,

the removal of cameras may result in higher or comparable speed. Indeed, the LUD global pipeline [36], which we follow, examines the parallel rigidity of the viewing graph and extracts the maximal parallel rigid subgraph. Thus earlier removal of cameras may worsen the parallel rigidity of the graph and increase the computation due to the need of finding a maximal parallel rigid subgraph. For example, [40] removes cameras in an earlier stage of the LUD pipeline, but results in higher computational cost than the LUD pipeline. Therefore, improvement of speed for the LUD pipeline by removing cameras is generally non-trivial. Moreover, currently we use scale factors obtained by first running LUD, so we do not get a real speed improvement. Nevertheless, the proposed method is insightful whenever it may indicate clear improvement in accuracy for a dataset, since one can then infer that the current pipeline is not effective enough in handling corrupted measurements, which can be easily recognized by a simple method. Furthermore, improvement in “speed” can be indicative of maintaining parallel rigidity.

Our RSR formulation is based on a fundamental observation by Sengupta et al. [39] on the low-rank of the n -view essential (or fundamental) matrix. The n -view essential matrix \mathbf{E} of size $3n \times 3n$ is formed by stacking all $\binom{n}{2}$ essential matrices, while being appropriately scaled. That is, the ij -th block of \mathbf{E} is the essential matrix for the i -th and j -th cameras, where each \mathbf{E}_{ij} is scaled by a factor λ_{ij} in accordance with the global coordinate system (see [20, 21, 39]). It was noticed in [39] that \mathbf{E} has rank 6. Moreover, [39] characterized the set of n -view essential matrices whose camera centers are not all collinear by the satisfaction of a few algebraic conditions, where the major one is $\text{rank}(\mathbf{E})=6$. Further explanation appears in [20].

We propose a straightforward application of RSR, utilizing these ideas to initially eliminate cameras that introduce significant corruption to the essential matrices. For this purpose, we compute the essential matrices (by computing first the fundamental matrices and then using the known camera calibration) and scale each matrix according to the factor obtained by the LUD pipeline [36] (note that this is the initial

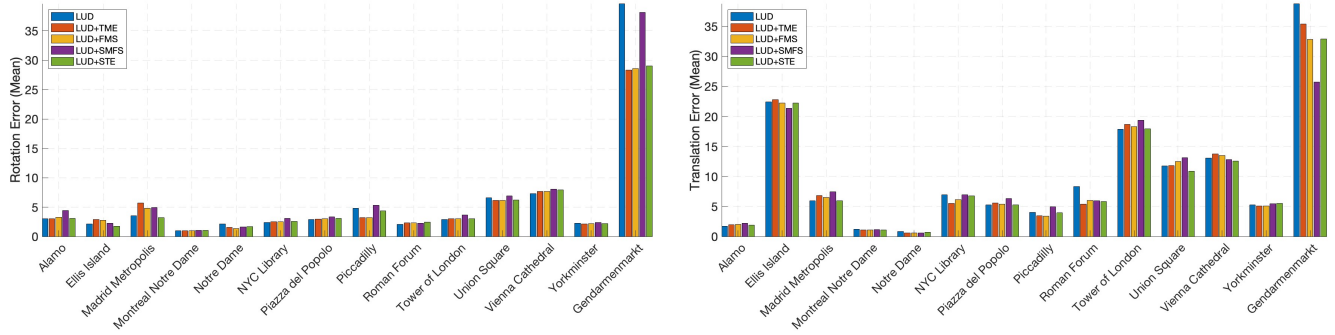


Figure 2. Mean (absolute) rotation errors (in degrees, left) and mean translation errors (in degrees, right) of LUD and four RSR methods used to initially screen bad cameras within LUD applied to the 14 datasets of Photo Tourism.

scaling applied in [20, 21, 39] before applying a non-convex and nontrivial optimization procedure that refines such scales). Using these appropriately scaled essential matrices, we form the n -view essential matrix \mathbf{E} of size $3n \times 3n$. We denote the $3n \times 3$ column blocks of \mathbf{E} by $\mathbf{E}_{:,1}, \dots, \mathbf{E}_{:,n}$ (since \mathbf{E} is symmetric they are the same as the row blocks transposed). We treat \mathbf{E} as a data matrix with $D = N = 3n$, where the columns of \mathbf{E} are the data points. We apply RSR with $d=6$, recover a d -dimensional robust subspace and identify the outlying columns whose distance is largest from this subspace. To avoid heuristic methods for the cutoff of outliers we assume a fixed percentage of 20% outlying columns. If a column block, $\mathbf{E}_{:,i}$ contains an outlying column, we remove its corresponding camera i . Consequently, a smaller percentage of cameras (about 10 – 15%) will be eliminated.

We use the Photo Tourism database [41] with precomputed pairwise image correspondences provided by [39] (they were obtained by thresholding SIFT feature similarities). To compute scale factors for the essential matrices we use the output of the LUD pipeline [36] as follows (following an idea proposed in [39] for initializing these values): Given the essential matrix for cameras i and j computed at an early stage of our pipeline, \mathbf{E}_{ij} , and the one obtained by the full LUD pipeline, $\mathbf{E}_{ij}^{\text{LUD}}$, the scaling factor is $\lambda_{ij} = \langle \mathbf{E}_{ij}, [\mathbf{E}_{ij}^{\text{LUD}}] \rangle / \|\mathbf{E}_{ij}^{\text{LUD}}\|_F^2$. Since many values of \mathbf{E}_{ij} are missing, we also apply matrix completion.

We compare the LUD pipeline with the LUD pipeline combined with the filtering processes achieved by STE, FMS, SFMS, and TME. For STE we fix $\gamma = 1/3$, though any other value we tried yielded the same result. We report both mean and median errors of rotations and translations and runtime of the standard LUD and the RSR+LUD methods with initial screening of cameras. Figure 2 shows the mean rotation and translation errors, where the rest of the figures and a summarizing table are in the supplementary material. In general, STE demonstrates slightly higher accuracy compared to other RSR methods. Improved accuracy is particularly notable when matrix completion is not utilized, as demonstrated in the supplementary material. We observe that LUD+STE generally improves the estimation of camera parameters (both rotations and translations) over LUD.

The improvement of LUD+STE is noticeable in Roman Forum and Gendarmenmarkt. In the supplementary material we show further improvement for Gendarmenmarkt with the removal of 45% outlying columns. While the resulting errors are still large, their improvement shows some potential in dealing with difficult SfM structure by initially removing cameras in a way that may help eliminate some scene ambiguities, which are prevalent in Gendarmenmarkt. In terms of runtime, both LUD+STE and LUD+SFMS demonstrate significant improvements, where LUD+SFMS is even faster than LUD+STE. While this does not yet imply faster handling of the datasets (as we use initial scaling factors obtained by LUD), it indicates the efficiency of the removal of outliers in maintaining parallel rigidity.

5. Conclusions

We introduce STE, a meticulously crafted adaptation of TME designed to address challenges within RSR. Theoretical guarantees demonstrate its ability to recover the true underlying subspace reliably, even with a smaller fraction of inliers compared to the well-known theoretical threshold. Under the generalized haystack model, we show that this initialization can be chosen as TME itself, leading to improved handling of a smaller fraction of inliers compared to TME. Our exploration extends to practical applications, where STE proves effective in two 3D vision tasks: robust fundamental matrix estimation and screening of bad cameras for improved SfM.

Several avenues for future research include: • Exploring adaptations of other robust covariance estimation methods to RSR. • Studying effective initialization for STE both in theory and in practice. • In-depth theoretical exploration of the optimal choice of the parameter γ . • Study of alternative ways of adapting TME to RSR problems. • Improving STE for fundamental matrix estimation following ideas similar to those in [7, 12, 37] for addressing challenging degeneracies. • Enhancing our initial idea of initial removal of bad cameras, specifically attempting to use it to rectify challenging scene ambiguities. • Testing our methods for SfM using more recent feature matching algorithms.

References

- [1] Larry P. Ammann. Robust singular value decompositions: A new approach to projection pursuit. *Journal of the American Statistical Association*, 88(422):pp. 505–514, 1993. [1](#)
- [2] Daniel Barath, Jiri Matas, and Jana Noskova. MAGSAC: marginalizing sample consensus. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10197–10205, 2019. [2](#)
- [3] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010. [17](#)
- [4] Yeshwanth Cherapanamjeri, Prateek Jain, and Praneeth Netrapalli. Thresholding based outlier robust PCA. In *Conference on Learning Theory*, pages 593–628. PMLR, 2017. [1](#)
- [5] Vartan Choulakian. L_1 -norm projection pursuit principal component analysis. *Computational Statistics & Data Analysis*, 50(6):1441–1451, 2006. [1](#)
- [6] Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized RANSAC. In *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25*, pages 236–243. Springer, 2003. [2](#), [6](#), [13](#)
- [7] Ondrej Chum, Tomas Werner, and Jiri Matas. Two-view geometry estimation unaffected by a dominant plane. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 772–779. IEEE, 2005. [2](#), [6](#), [8](#), [13](#)
- [8] Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, page 2683–2702, USA, 2018. Society for Industrial and Applied Mathematics. [1](#)
- [9] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha. R1-PCA: rotational invariant L_1 -norm principal component analysis for robust subspace factorization. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, pages 281–288, New York, NY, USA, 2006. ACM. [1](#)
- [10] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. [1](#), [2](#), [6](#), [13](#)
- [11] Gabriel Frahm and Uwe Jaekel. A generalization of Tyler’s M-estimators to the case of incomplete data. *Computational Statistics & Data Analysis*, 54(2):374–393, 2010. [2](#)
- [12] J-M Frahm and Marc Pollefeys. RANSAC for (quasi-) degenerate data (QDEGSAC). In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pages 453–460. IEEE, 2006. [2](#), [8](#)
- [13] William Cole Franks and Ankur Moitra. Rigorous guarantees for Tyler’s M-estimator via quantum expansion. In *Conference on Learning Theory*, pages 1601–1632. PMLR, 2020. [3](#)
- [14] Jerome H. Friedman and John W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23(9):881–890, 1974. [1](#)
- [15] John Goes, Gilad Lerman, and Boaz Nadler. Robust sparse covariance estimation by thresholding Tyler’s M-estimator. *The Annals of Statistics*, 48(1):86 – 110, 2020. [3](#)
- [16] Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In *Conference on Learning Theory*, pages 354–375. PMLR, 2013. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#)
- [17] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. [6](#), [11](#), [12](#), [13](#)
- [18] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2):517–547, 2021. [13](#)
- [19] Arman Karimian and Roberto Tron. Essential matrix estimation using convex relaxations in orthogonal space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17142–17152, 2023. [15](#)
- [20] Yoni Kasten, Amnon Geifman, Meirav Galun, and Ronen Basri. Algebraic characterization of essential matrices and their averaging in multiview settings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5895–5903, 2019. [7](#), [8](#)
- [21] Yoni Kasten, Amnon Geifman, Meirav Galun, and Ronen Basri. Gpsfm: Global projective sfm using algebraic constraints on multi-view fundamental matrices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3264–3272, 2019. [7](#), [8](#)
- [22] John T. Kent and David E. Tyler. Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*, 15(2):247–254, 1988. [2](#), [3](#)
- [23] Nojun Kwak. Principal component analysis based on L_1 -norm maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(9):1672–1680, 2008. [1](#)
- [24] Gilad Lerman and Tyler Maunu. Fast, robust and non-convex subspace recovery. *Information and Inference: A Journal of the IMA*, 7(2):277–336, 2018. [1](#), [2](#), [3](#), [4](#), [6](#), [13](#)
- [25] Gilad Lerman and Tyler Maunu. An overview of robust subspace recovery. *Proceedings of the IEEE*, 106(8):1380–1410, 2018. [1](#), [3](#), [4](#), [5](#), [6](#)
- [26] Gilad Lerman and Teng Zhang. l_p -recovery of the most significant subspace among multiple subspaces with outliers. *Constr. Approx.*, 40(3):329–385, 2014. [4](#)
- [27] Gilad Lerman, Michael B. McCoy, Joel A. Tropp, and Teng Zhang. Robust computation of linear models by convex relaxation. *Found. Comput. Math.*, 15(2):363–410, 2015. [1](#), [3](#), [5](#)
- [28] Gilad Lerman, Feng Yu, and Teng Zhang. Theoretical guarantees for the subspace-constrained Tyler’s estimator, 2024. [4](#), [5](#), [11](#)
- [29] Guoying Li and Zhonglian Chen. Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and monte carlo. *Journal of the American Statistical Association*, 80(391):759–766, 1985. [1](#)
- [30] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. [6](#)
- [31] Ricardo A. Maronna and Víctor J. Yohai. Robust estimation of multivariate location and scatter. *Wiley StatsRef: Statistics Reference Online*, pages 1–12, 2014. [2](#)
- [32] Tyler Maunu and Gilad Lerman. Robust subspace recovery with adversarial outliers, 2019. [1](#), [2](#), [4](#), [5](#)

- [33] Tyler Maunu, Teng Zhang, and Gilad Lerman. A well-tempered landscape for non-convex robust subspace recovery. *J. Mach. Learn. Res.*, 20(1):1348–1406, 2019. 1, 3, 4, 5, 6
- [34] Tyler Maunu, Chenyu Yu, and Gilad Lerman. Stochastic and private nonconvex outlier-robust PCAs. In *Proceedings of Mathematical and Scientific Machine Learning*, pages 173–188. PMLR, 2022. 1
- [35] Michael McCoy and Joel A. Tropp. Two proposals for robust PCA using semidefinite programming. *Electronic Journal of Statistics*, 5(none):1123 – 1160, 2011. 1
- [36] Onur Ozyesil and Amit Singer. Robust camera location estimation by convex programming. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2674–2683, 2015. 7, 8
- [37] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. USAC: A universal framework for random sample consensus. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):2022–2038, 2012. 2, 8
- [38] Elvezio M. Ronchetti and Peter J. Huber. *Robust statistics*. John Wiley & Sons Hoboken, NJ, USA, 2009. 1
- [39] Soumyadip Sengupta, Tal Amir, Meirav Galun, Tom Goldstein, David W Jacobs, Amit Singer, and Ronen Basri. A new rank constraint on multi-view fundamental matrices, and its application to camera location recovery. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4798–4806, 2017. 7, 8, 16
- [40] Yunpeng Shi, Shaohan Li, Tyler Maunu, and Gilad Lerman. Scalable cluster-consistency statistics for robust multi-object matching. In *International Conference on 3D Vision, 3DV 2021, London, United Kingdom, December 1-3, 2021*, pages 352–360. IEEE, 2021. 7
- [41] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM siggraph 2006 papers*, pages 835–846. Association for Computing Machinery, 2006. 6, 7, 8
- [42] Nathan Srebro and Tommi Jaakkola. Weighted low-rank approximations. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 720–727, 2003. 1
- [43] Jacob Steinhardt, Moses Charikar, and Gregory Valiant. Resilience: A criterion for learning in the presence of arbitrary outliers. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, pages 45:1–45:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. 1
- [44] Ben Tordoff and David W. Murray. Guided sampling and consensus for motion estimation. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part I 7*, pages 82–96. Springer, 2002. 15
- [45] Philip H.S. Torr and Andrew Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer vision and image understanding*, 78(1):138–156, 2000. 2
- [46] David E. Tyler. Statistical analysis for the angular central gaussian distribution on the sphere. *Biometrika*, 74(3):579–589, 1987. 2
- [47] David E. Tyler. A distribution-free m-estimator of multivariate scatter. *The Annals of Statistics*, pages 234–251, 1987. 1, 2, 6, 13
- [48] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13*, pages 61–75. Springer, 2014. 7
- [49] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust PCA via outlier pursuit. *Advances in neural information processing systems*, 23, 2010. 1
- [50] Teng Zhang. Robust subspace recovery by Tyler’s M-estimator. *Information and Inference: A Journal of the IMA*, 5(1):1–21, 2016. 1, 3, 4, 5, 6, 13
- [51] Teng Zhang and Gilad Lerman. A novel M-estimator for robust PCA. *The Journal of Machine Learning Research*, 15(1): 749–808, 2014. 1, 3, 5

Supplementary Material

Our theoretical results are proved and further explained in [28]. Additional numerical results and discussion are provided in §A. We refer to references, equation numbers and sections from the main text. We also follow the same numbering of figures, tables and equations, and thus start, for example, with Figure 3.

A. Supplementary Details for the Numerical Experiments

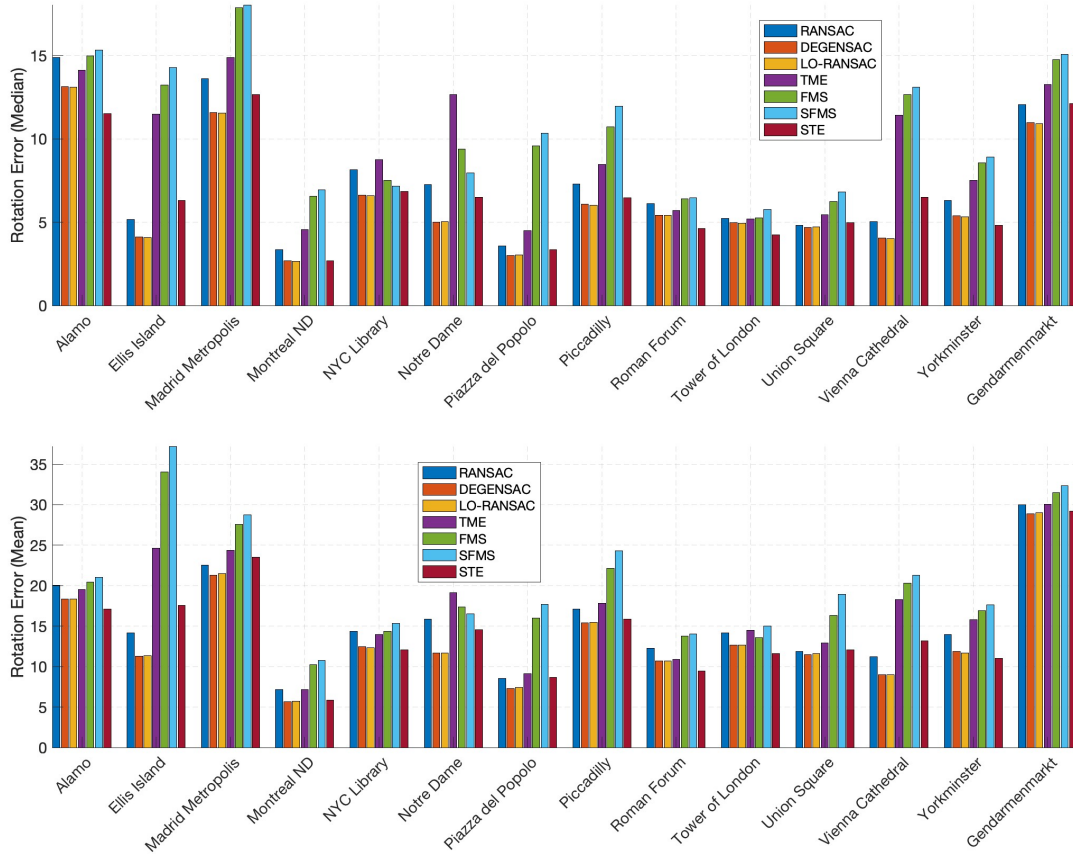


Figure 3. Median (top) and mean (bottom) relative rotation errors (in degrees) obtained by seven algorithms for the 14 datasets of Photo Tourism. The numerical results are reported in Table 1

We provide additional numerical results and details for our two different types of applications in §A.1 and §A.2.

A.1. Supplementary Details for Fundamental Matrix Estimation

We review some details about data normalization applied for RSR methods in §A.1.1 and the implementation of the RSR methods in §A.1.2. We provide additional numerical results in §A.1.3, and runtime comparison of only the RSR algorithms (including RANSAC) in §A.1.4.

A.1.1 Details of Data Normalization

Fundamental matrix estimation requires proper normalization of the input data [17]. We chose to use a normalization that is very similar to the commonly used one in the 8-point algorithm; nevertheless, we apply the normalization to the full dataset as our method is applied to the full dataset. We notice that there are even slightly better normalization techniques, however, to make sure that our competitive performance is due to the proposed method and not the normalization itself, we use the direct generalization of the standard method.

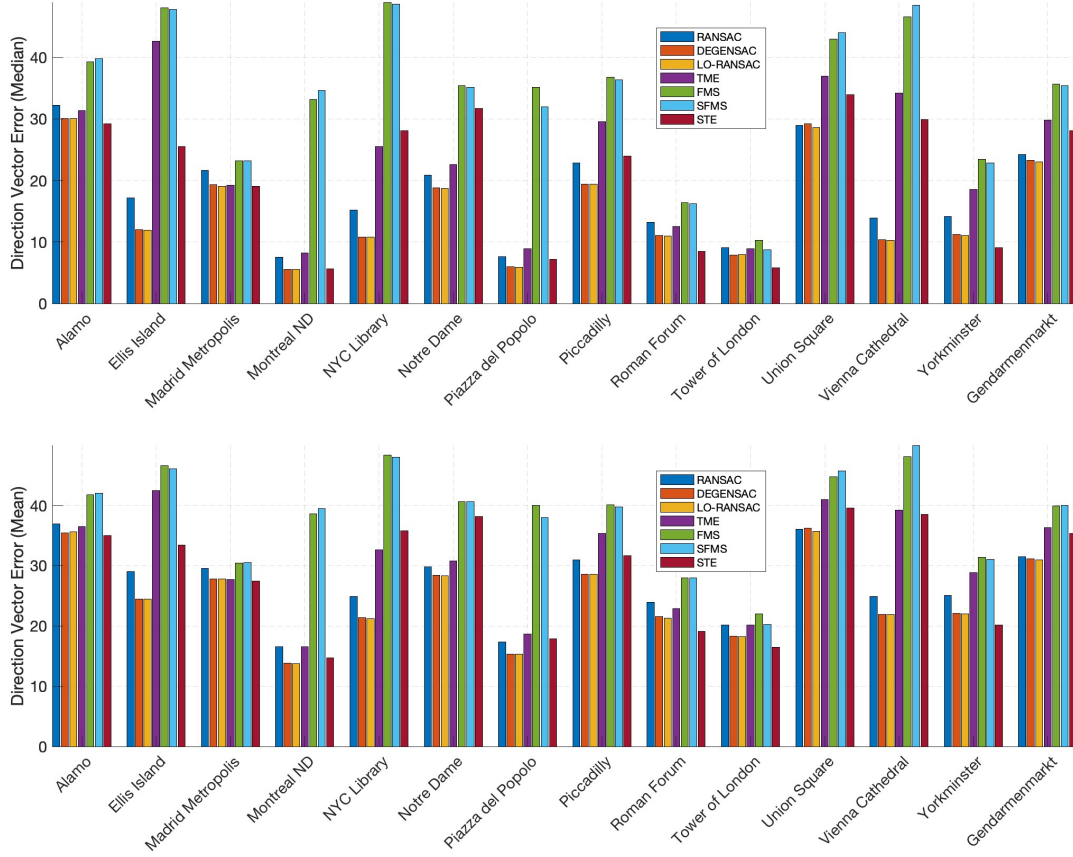


Figure 4. Median (top) and mean (bottom) errors of direction vectors (in degrees) obtained by seven algorithms for the 14 datasets of Photo Tourism. The numerical results are reported in Table 2

Each data point is represented by a column vector of homogeneous coordinates, $\mathbf{x} \in \mathbb{R}^3 = (x_1, x_2, 1)^\top$. For the N data points, $\mathbf{x}_1, \dots, \mathbf{x}_N$, we form the data matrices $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{3 \times N}$. Let (μ_1, σ_1) and (μ_2, σ_2) be the means and standard deviations of the first and the second rows of \mathbf{X} , respectively. We normalize the points as follows:

$$\hat{\mathbf{x}}_i = \mathbf{T} \mathbf{x}_i, 1 \leq i \leq N,$$

where the normalizing transformation $\mathbf{T} \in \mathbb{R}^{3 \times 3}$ is given by

$$\mathbf{T} = \begin{bmatrix} \sigma_1^{-1} & 0 & 0 \\ 0 & \sigma_2^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\mu_1 \\ 0 & 1 & -\mu_2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_1^{-1} & 0 & -\sigma_1^{-1}\mu_1 \\ 0 & \sigma_2^{-1} & -\sigma_2^{-1}\mu_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

This normalization is applied to any RSR method that uses all data points at once to form the subspace. In our experiments, these methods include FMS, SFMS, TME and STE. For RANSAC and its variants (which compute subspaces for 8 points at a time) we use the common normalization [17], which is exactly the normalization above, but applied to the 8 points chosen each time.

A.1.2 Review of the Implementation of the RSR Algorithms

We apply each RSR method (TME, FMS, SFMS and STE) as follows. Given two normalized data matrices (see Section A.1.1) $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N] \in \mathbb{R}^{3 \times N}$ and $\hat{\mathbf{X}}' = [\hat{\mathbf{x}}'_1, \dots, \hat{\mathbf{x}}'_N] \in \mathbb{R}^{3 \times N}$ whose columns are normalized feature points for the two given images, we form the data matrix $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N] \in \mathbb{R}^{9 \times N}$ with the following columns: $\tilde{\mathbf{x}}_i = \text{vec}(\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i'^\top)$, $i = 1, \dots, N$. We apply each RSR method with $d = 8$ and $D = 9$ (that is, we aim to recover an 8-subspace in \mathbb{R}^9) to $\tilde{\mathbf{X}}$. We find the orthogonal vector to this 8-dimensional subspace in \mathbb{R}^9 and reshape it in into a 3 by 3 matrix $\tilde{\mathbf{F}}$. In order to report a proper fundamental matrix with rank

2, we replace its lowest singular value with 0. Finally, we obtain the estimated fundamental matrix for the original data (before normalization) as follows: $\mathbf{T}^\top \tilde{\mathbf{F}} \mathbf{T}'$, where \mathbf{T} and \mathbf{T}' are the normalization transformations defined in Appendix A.1.1.

A.1.3 Additional Numerical Results for Fundamental Matrix Estimation

We assess the quality of the fundamental matrix estimation by the median and mean errors of relative rotation and direction vectors directly obtained by the fundamental matrices for the various methods. We recall that these methods include STE (our proposed algorithm), TME [47, 50], FMS [24], SFMS [24], vanilla RANSAC [10], DEGENSAC [7] and LO-RANSAC [6]. We first explain how we compute the above errors for any of these methods.

For any two cameras, i and j , We estimate the fundamental matrix $\tilde{\mathbf{F}}_{ij}$ from the correspondence pairs of the images of these cameras and then extract from $\tilde{\mathbf{F}}_{ij}$ the relative rotation, $\tilde{\mathbf{R}}_{ij}$, and the direction vector (that is, relative translation normalized to have norm 1) in the coordinates of camera i , $\tilde{\mathbf{t}}_{ij}$ [17]. We remark that we need to normalize the relative translation to obtain direction vectors due to the scale ambiguity of the fundamental matrix. For cameras i and j , the relative rotation error compares the estimated relative rotation, $\tilde{\mathbf{R}}_{ij}$, and the ground-truth one (provided by the Photo Tourism database), $\{\mathbf{R}_{ij}^*\}$, as follows:

$$e_R = \cos^{-1} \left(\frac{\text{tr}(\mathbf{R}_{ij}^{*\top} \tilde{\mathbf{R}}_{ij}) - 1}{2} \right).$$

Note that the estimated direction vectors, $\{\tilde{\mathbf{t}}_{ij}\}$, are defined up to a global orientation and in order to align them with the ground-truth information (given in the frame of camera i) obtained from the Photo Tourism data set, $\{\mathbf{t}_{ij}^*\}$, we find a rotation matrix $\mathbf{R}_{\text{align}}$, which minimizes $\sum_{1 \leq i, j \leq N} \|\mathbf{t}_{ij}^* - \mathbf{R}_{\text{align}} \tilde{\mathbf{t}}_{ij}\|_F^2$. For cameras i and j , the error of direction vectors is

$$e_T = \cos^{-1}(|\mathbf{t}_{ij}^* \cdot \mathbf{R}_{\text{align}} \tilde{\mathbf{t}}_{ij}|).$$

We remark that the estimated direction vector, $\tilde{\mathbf{t}}_{ij}$ in the frame of camera i implicitly requires estimating both the direction vector in world's coordinates and the absolute rotation matrix (up to global scale) that maps from world's coordinates to the frame of camera i , and thus seems to be susceptible to large errors.

Location	RANSAC		DEGENSAC		LO-RANSAC		TME		SFMS		FMS		STE	
	\tilde{e}_R	\hat{e}_R	\tilde{e}_R	\hat{e}_R	\tilde{e}_R	\hat{e}_R	\tilde{e}_R	\hat{e}_R	\tilde{e}_R	\hat{e}_R	\tilde{e}_R	\hat{e}_R	\tilde{e}_R	\hat{e}_R
Alamo	14.90	20.02	13.13	18.38	13.12	18.35	14.13	19.50	14.98	20.44	15.33	21.02	11.51	18.65
Ellis Island	5.17	14.17	4.10	11.26	4.09	11.38	11.48	24.64	13.22	34.06	14.27	37.16	6.71	17.57
Madrid Metropolis	13.61	22.52	11.57	21.28	11.56	21.48	14.89	24.38	17.89	27.58	18.05	28.78	12.11	21.63
Montreal N.D.	3.36	7.14	2.69	5.69	2.66	5.72	4.55	7.18	6.56	10.22	6.94	10.74	2.72	5.80
NYC Library	8.16	14.34	6.64	12.49	6.60	12.33	8.76	13.97	7.50	14.35	7.18	15.35	6.99	12.25
Notre Dame	7.25	15.84	4.99	11.67	5.03	11.68	12.67	19.16	9.40	17.36	7.96	16.54	6.86	15.11
Piazza del Popolo	3.59	8.54	3.00	7.32	3.02	7.45	4.51	9.10	9.59	15.98	10.35	17.69	3.28	8.21
Piccadilly	7.30	17.11	6.08	15.42	6.03	15.47	8.47	17.85	10.73	22.12	11.95	24.29	6.81	16.38
Roman Forum	6.11	12.29	5.42	10.69	5.41	10.68	5.69	10.90	6.41	13.75	6.46	14.05	4.55	9.60
Tower of London	5.24	14.16	4.96	12.69	4.94	12.64	5.20	14.47	5.27	13.55	5.77	15.02	4.18	12.03
Union Square	4.82	11.89	4.70	11.46	4.73	11.60	5.44	12.95	6.24	16.34	6.82	18.91	4.81	11.94
Vienna Cathedral	5.04	11.23	4.05	8.98	4.03	9.01	11.44	18.29	12.67	20.31	13.12	21.27	7.05	13.75
Yorkminster	6.31	13.97	5.39	11.85	5.33	11.70	7.52	15.83	8.55	16.93	8.92	17.63	4.87	11.75
Gendarmenmarkt	12.05	29.99	10.99	28.89	10.93	28.99	13.28	30.05	14.77	31.48	15.09	32.38	11.83	29.10

Table 1. Median and mean relative rotation errors (in degrees) obtained by seven algorithms for the 14 datasets of Poto Tourism. \tilde{e}_R is the median relative rotation error, \hat{e}_R is the mean relative rotation error.

Figure 3 is a bar plot presenting the errors of relative rotations (in angles) for the different methods and all 14 datasets and Figure 4 showcases the errors of direction vectors. For completeness, we record the presented numerical values in Table 1 (for relative rotations) and Table 2 (for direction vectors). Figure 5 presents the mean Average Accuracy (mAA) [18] for rotation with threshold 10° , where as opposed to the former errors which need to be small, higher mAA is better.

We note that STE significantly outperforms FMS and SFMS. STE also outperforms TME, where the only cases where TME outperforms STE is for direction vector estimation for the NYC Library and Notre Dame datasets. In terms of relative rotation

Location	RANSAC		DEGENSAC		LO-RANSAC		TME		SFMS		FMS		STE	
	\tilde{e}_T	\hat{e}_T	\tilde{e}_T	\hat{e}_T	\tilde{e}_T	\hat{e}_T	\tilde{e}_T	\hat{e}_T	\tilde{e}_T	\hat{e}_T	\tilde{e}_T	\hat{e}_T	\tilde{e}_T	\hat{e}_T
Alamo	32.20	36.97	30.09	35.46	30.11	35.57	31.36	36.49	39.31	41.74	39.85	41.98	29.25	35.00
Ellis Island	17.16	29.04	11.98	24.43	11.90	24.45	42.62	42.42	48.11	46.62	47.79	46.02	25.48	33.44
Madrid Metropolis	21.67	29.56	19.36	27.83	19.09	27.82	19.23	27.71	23.22	30.47	23.18	30.52	19.04	27.47
Montreal N.D.	7.49	16.51	5.57	13.80	5.53	13.75	8.25	16.55	33.20	38.61	34.65	39.51	5.62	14.71
NYC Library	15.17	24.87	10.77	21.35	10.80	21.25	25.48	32.62	48.89	48.34	48.69	48.02	28.09	35.79
Notre Dame	20.91	29.83	18.82	28.40	18.75	28.30	22.61	30.74	35.38	40.63	35.16	40.60	31.68	38.12
Piazza del Popolo	7.58	17.35	5.95	15.31	5.93	15.34	8.88	18.70	35.17	39.99	31.98	37.96	7.16	17.83
Piccadilly	22.89	30.99	19.40	28.58	19.39	28.58	29.58	35.36	36.83	40.06	36.33	39.76	23.95	31.64
Roman Forum	13.24	23.96	11.07	21.52	10.97	21.30	12.51	22.86	16.42	27.96	16.24	27.95	8.52	19.13
Tower of London	9.10	20.18	7.90	18.30	7.95	18.20	8.88	20.15	10.27	22.00	8.77	20.21	5.80	16.46
Union Square	28.97	36.03	29.23	36.21	28.60	35.73	36.98	40.98	42.97	44.76	44.02	45.71	33.93	39.60
Vienna Cathedral	13.88	24.94	10.33	21.89	10.26	21.89	34.24	39.24	46.63	48.11	48.46	49.88	29.88	38.53
Yorkminster	14.17	25.11	11.20	22.05	11.05	21.99	18.54	28.87	23.48	31.42	22.82	31.03	9.09	20.17
Gendarmenmarkt	24.25	31.51	23.32	31.11	23.03	30.96	29.82	36.33	35.70	39.88	35.44	40.00	28.09	35.37

Table 2. Median and mean errors of direction vectors (in degrees) obtained by seven algorithms for the 14 datasets of Photo Tourism. \tilde{e}_T is the corresponding median error, \hat{e}_T is the corresponding mean error.

estimation, STE outperforms vanilla RANSAC, where Ellis Island and Vienna Cathedral are the only two datasets where RANSAC outperforms STE. In terms of direction vector estimation, STE outperforms vanilla RANSAC in 7 out of the 14 datasets and in the rest of them vanilla RANSAC outperforms STE. Anyway, the errors of direction vectors are higher than relative rotation errors because of an issue discussed above and many references do not present these errors. As shown in Figure 5, STE has a higher $mAA(10^\circ)$ than other RSR methods. Overall, STE is slightly better than RANSAC. Indeed, the averaged $mAA(10^\circ)$ for STE is 0.44 for STE and 0.41 for RANSAC. However, RANSAC has a noticeable higher $mAA(10^\circ)$ for the following 3 datasets: Ellis Island, Notre Dame and Vienna Cathedral. On the other hand, DEGENSAC and LO-RANSAC are overall slightly better than STE (with averaged $mAA(10^\circ)$ 0.45 for both methods), but STE has a noticeably higher $mAA(10^\circ)$ for the following 5 datasets: Alamo, Montreal Notre Dame, Roman Forum, Tower of London and York Minster.

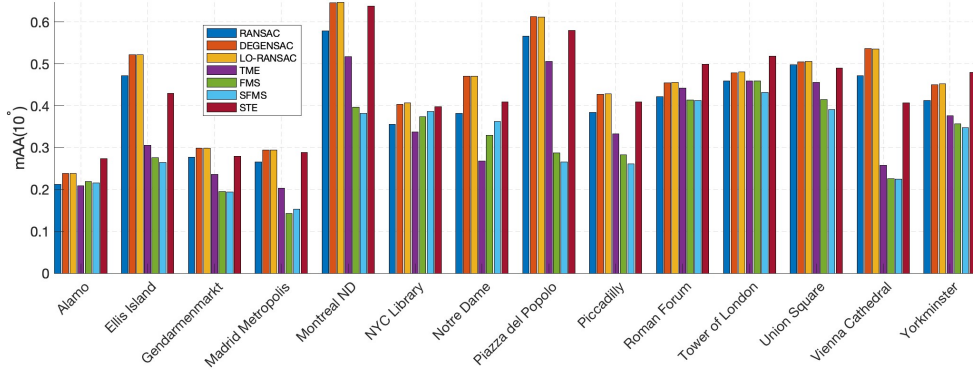


Figure 5. $mAA(10^\circ)$ obtained by seven algorithms for the 14 datasets of Photo Tourism.

A.1.4 Runtime Comparison of the RSR components

We compare the elapsed time of the different RSR algorithms, including RANSAC and its variants. Since our current implementation of STE uses Algorithm 2 with $m=5$ values of γ ($\gamma \in \{(2i)^{-1}\}_{i=1}^5$), it is slower than the basic implementation of Algorithm 1 by a factor 5. This choice keeps the similar accuracy of STE with more values of γ but reduces the overall runtime.

We compared the runtime of STE (with $m=5$) with RANSAC, TME, FMS and SFMS for the 14 Photo Tourism datasets (using Matlab implementations). These experiments were carried out on an Apple Silicon M2. For each dataset, the average runtime among all image pairs is reported in Table 3. Overall, the runtimes of STE and RANSAC are somewhat comparable, where the average

over all datasets of the runtime of RANSAC is 22.91 msec and of STE is 21.91 msec. However, the variance of the runtime of RANSAC is rather large, unlike STE. We further note that TME was the fastest method, however running STE with $m = 1$, instead of $m = 5$, is slightly faster than TME. FMS and SFMS are slower than TME and overall faster than STE (with $m = 5$) and RANSAC.

Locations	Algorithm Runtimes (ms)				
	STE	TME	SFMS	FMS	RANSAC
Alamo	19.2	4.0	13.7	13.2	11.8
Ellis Island	19.8	4.8	9.8	9.3	44.6
Madrid Metropolis	18.6	3.9	9.3	8.8	11.3
Montreal N.D.	18.0	3.6	18.5	17.8	31.6
NYC Library	21.1	4.5	22.5	22.0	51.7
Notre Dame	19.4	4.7	18.5	17.9	15.5
Piazza del Popolo	19.1	4.2	9.3	8.8	22.7
Piccadilly	19.0	4.4	8.8	8.2	24.0
Roman Forum	18.1	3.7	16.0	15.5	7.3
Tower of London	17.6	3.7	15.4	14.8	4.5
Union Square	19.0	4.2	9.6	9.1	10.2
Vienna Cathedral	20.5	4.5	21.3	20.8	37.5
Yorkminster	17.4	3.6	15.1	14.5	18.2
Gendarmenmarkt	18.7	3.7	8.5	8.0	29.9

Table 3. Elapsed runtimes (in milliseconds) of STE, TME, SFMS, FMS and RANSAC for the 14 datasets of Photo Tourism. The runtimes are averaged over all image pairs in the dataset.

The runtime of RANSAC mainly depends on the fraction of the outliers. For higher fractions of outliers, RANSAC significantly slowed down compared to other methods, showcasing exponential dependence of time on the fraction of outliers. Indeed, if ε denotes the fraction of inliers and d the subspace dimension and one requires 99% confidence, the expected number of iterations of RANSAC is given as $\log(1 - .99)/\log(1 - \varepsilon^d)$ [44].

To further investigate the expected correlation of the runtime and the fraction of outliers, we also considered artificial instances, which we repeatedly generated 100 times. We fixed $N = 400$ and generated two data matrices $\mathbf{X} = [x_1, \dots, x_N] \in \mathbb{R}^{3 \times N}$ and $\mathbf{X}' = [x'_1, \dots, x'_N] \in \mathbb{R}^{3 \times N}$, whose columns contain inlier and outlier feature correspondence pairs in homogenous coordinates as follows. We varied the fraction of designated outliers, where a fixed fraction of columns of \mathbf{X} and \mathbf{X}' (with the same index) were designated as outliers and the rest as inliers. For generating an outlier correspondence pair, the first two coordinates of x_i and x'_i were uniformly sampled in $[0, 1000]$ and the third coordinates was set to be 1. For an inlier correspondence pair, we followed the procedure of [19] to sample (x_i, x'_i) satisfying the epipolar constraint.

Given the generated matrices \mathbf{X} and \mathbf{X}' , we first normalized them according to the procedure described in §A.1.1 and then used the procedure described in §A.1.2 to estimate the fundamental matrix from these two matrices using any of the RSR methods with $d = 8$ and $D = 9$. The application of RANSAC is also similar to before. The relative rotation was extracted from the estimated fundamental matrix and its error was computed based on the given ground truth rotation for the inliers, which are clarified when following the details of generating them in [19]. Using these 100 samples we computed $\text{mAA}(10^\circ)$.

Figure 6 reports $\text{mAA}(10^\circ)$ and runtimes of STE (with $m = 5$), TME, SFMS, FMS and RANSAC for different fractions of outliers. The maximum number of iteration of RANSAC was set to be 1000. We note that STE (with $m = 5$) achieved the highest $\text{mAA}(10^\circ)$ when the fraction of outliers was at most 50% and in this case the $\text{mAA}(10^\circ)$ values of FMS are slightly below those of STE. On the other hand, when the fraction of outliers is at least 60% the $\text{mAA}(10^\circ)$ values of STE are higher than STE, but they indicate low accuracy. The values of RANSAC, SFMS and TME are below the ones of STE. Moreover, the $\text{mAA}(10^\circ)$ values of TME are rather low and indicate low accuracy.

Regarding runtime, we first note that FMS, SFMS and TME have fixed times, since the number of data points, N , is fixed. Moreover, TME is faster than FMS and SFMS whose time is comparable. On the other hand, STE has a smaller run time for low fraction of outliers, since it converges faster then. Furthermore, the runtime of STE stabilizes when the fraction of outliers is large and it is slightly above FMS and SFMS. RANSAC is relatively fast for a small fraction of outliers, but it is much slower than all other method when the fraction of outliers is at least 50%.

The above mentioned formula of expected number of iterations implies that RANSAC can easily fail with large d . In the setting of 3-

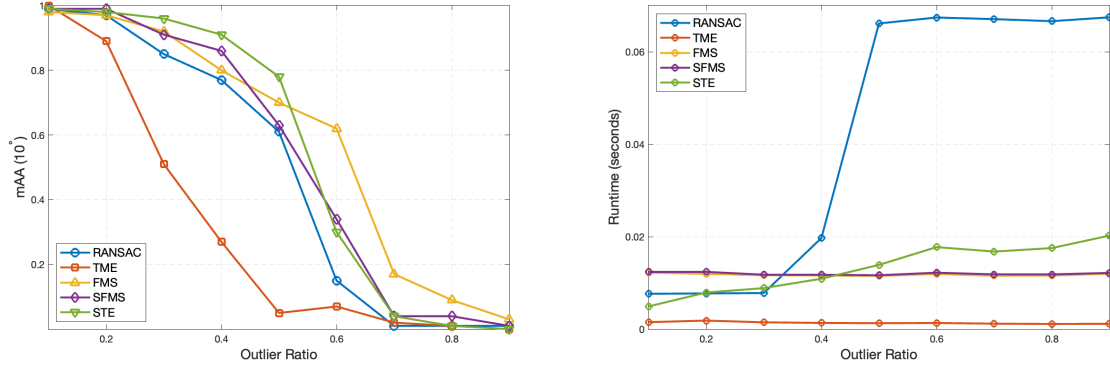


Figure 6. mAA(10°) (left) and runtime (right, in seconds) of five algorithms for the synthetic feature points data. The x -axis indicates the fraction of the outliers, where the total number of points is fixed as 400.

view homography estimation, $D=27$ and $d=26$. We thus tried the standard haystack model with $D=27$, $d=26$, $N=400$ and varied fractions of outliers. To measure error, we use the angle between L_* , the inlier subspace of the haystack model, and L , the estimated subspace. Figure 7 compares the errors and runtimes of the five algorithms for different fractions of outliers. We note that STE has the lowest error, but they notably increase for fraction of outliers larger than 50%. The errors of SFMS and TME are only slightly above those of STE. RANSAC performed poorly, even for very small fractions of outliers and FMS is even worse than RANSAC.

We note that the runtimes of the RSR methods behave similarly to the ones in Figure 6. In each one of the experiments RANSAC needed the maximal number of iterations, which was set as 1000, and obtained the same runtime, which is significantly larger than the other runtimes.

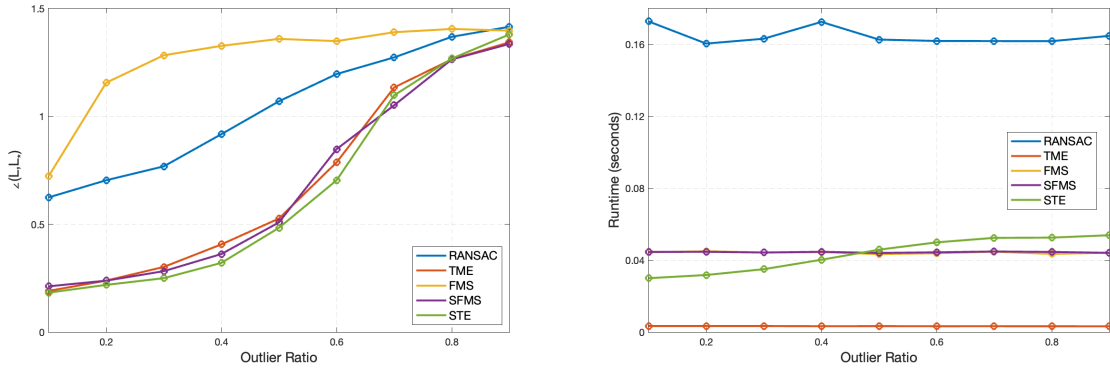


Figure 7. Angular error (left, in radius) and runtime (right, in seconds) of five algorithms for the generalized haystack synthetic data. The error is angle between L_* and the estimated subspace \hat{L} recovered by the algorithm. The x -axis is the fraction of outliers.

A.2. Additional Numerical Results for Initial Camera Removal for SfM

We use all 14 datasets of the Photo Tourism database. As mentioned in the main text, scale factors are first obtained by the LUD algorithm. Many of the blocks \mathbf{E}_{ij} are absent when there are no matching features between images i and j . On the other hand, the geometric relationship between cameras i and j still exists and the ground-truth essential matrix is thus well defined. In such scenarios, a common strategy is to assign to \mathbf{E}_{ij} a zero matrix [39]. For completeness, we apply our proposed procedure with this strategy in Section A.2.1. However, the results should be more accurate when using matrix completion to fill the missing blocks of \mathbf{E} . We follow such a strategy here and provide its details as follows. Let Ω denote the set of indices, ij , of the blocks \mathbf{E}_{ij} that can be directly obtained by the images. We denote by $\mathbf{M} \in \mathbb{R}^{3n \times 3n}$ the desired matrix with blocks $\{\lambda_{ij}\mathbf{E}_{ij} \mid (i,j) \in \Omega\}$ and additional

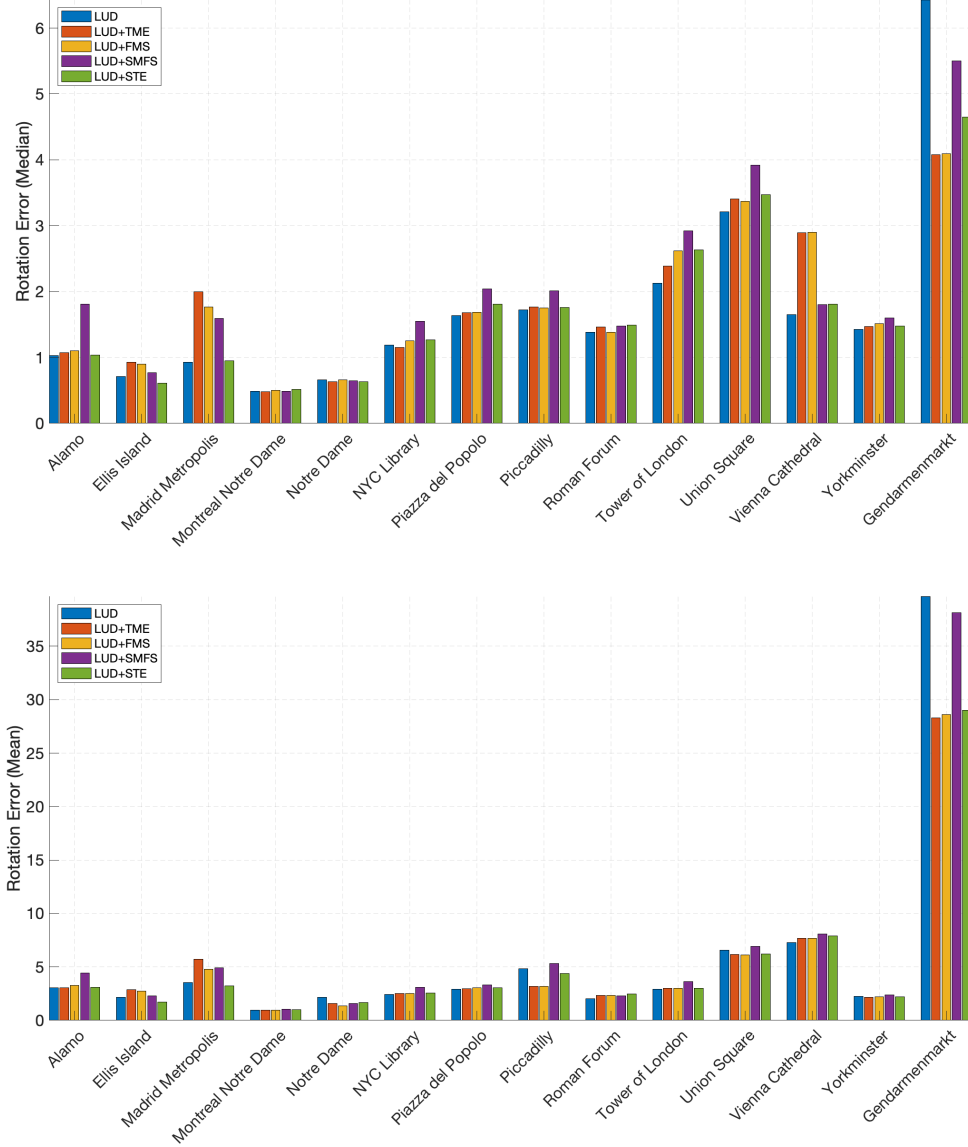


Figure 8. Median (top) and mean (bottom) absolute rotation errors (in degrees) of LUD and four RSR methods used to initially screen bad cameras within LUD applied to the 14 datasets of Photo Tourism. For Gendarmenmarkt, the median rotation error of SFMS was 31.04 degrees and mean error of SFMS was 62.89, which are above the figure range.

completed ones. We approximate \mathbf{M} by solving the following relaxed optimization problem (see e.g., [3]):

$$\widehat{\mathbf{M}} = \underset{\mathbf{M} \in \mathbb{R}^{3n \times 3n}}{\operatorname{argmin}} \|\mathbf{M}\|_*, \quad \text{subject to } \mathbf{M}_{ij} = \lambda_{ij} \mathbf{E}_{ij}, \quad (i, j) \in \Omega. \quad (6)$$

The Singular Value Thresholding (SVT) algorithm [3] is applied to solve Equation (6). It is worth noting that for some datasets, the sampling ratio $|\Omega|/n^2$ is smaller than 10%. To prevent the SVT algorithm from diverging, we set a step size $\delta = n^2/(10|\Omega|)$, which is smaller than the suggested value in [3, Section 5.1.2]. Since the block \mathbf{M}_{ij} is defined as an essential matrix, we further

project it on the essential matrix manifold. Specifically, we define $\tilde{\mathbf{E}} \in \mathbb{R}^{3n \times 3n}$ as follows:

$$\tilde{\mathbf{E}}_{ij} = \begin{cases} \lambda_{ij} \mathbf{E}_{ij}, & \text{if } (i,j) \in \Omega \\ \mathbf{P}_E(\widehat{\mathbf{M}}_{ij}), & \text{if } (i,j) \notin \Omega \\ \mathbf{0} & \text{if } i=j, \end{cases}$$

where for $\widehat{\mathbf{M}}_{ij} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, $\mathbf{P}_E(\widehat{\mathbf{M}}_{ij}) = \mathbf{U} \text{diag}([1, 1, 0]) \mathbf{V}^\top$ is the projector to the essential manifold. Following the procedures described in Section 4.2, we apply RSR with $d=6$ by treating $\tilde{\mathbf{E}}$ as a data matrix of $D = N = 3n$, recover a d -dimensional robust subspace and identify the outlying columns whose distance is largest from this subspace.

We compare mean and median errors of rotations and translations and runtime for the LUD pipeline and the LUD+RSR pipeline for the following four RSR algorithms: STE, FMS, SFMS, and TME. By LUD+RSR we refer to our proposed camera filtering process by the chosen RSR algorithm followed by LUD. Figure 8 is a bar plot of rotation errors (in degrees) for all methods and all 14 datasets, Figure 9 is a bar plot of translation errors (in meters), and Figure 10 is a bar plot of run times. As explained in the main text, we only aim to compare the effect of the camera screening on the overall time (and this screening may potentially increase time due to issues with parallel rigidity), but it is not the actual time as LUD+RSR uses information for scaling factors obtained by LUD. For completeness, Table 4 summarizes the results of LUD, LUD+STE, and LUD+SFMS, whereas Table 5 presents the results of LUD+TME and LUD+FMS.

Overall, STE yields better accuracy than other RSR methods. We observe that LUD+STE generally improves the estimation of camera parameters (both rotations and translations) over LUD, though the median rotation errors are comparable. The improvement of LUD+STE is significant for both the Roman Forum and Gendarmenmarkt. In terms of runtime, both LUD+STE and LUD+SFMS demonstrate improvements, where LUD+SFMS is even faster than LUD+STE. While this does not yet imply faster handling of the datasets (as we use initial scaling factors obtained by LUD), it indicates some efficiency in the removal of outliers.

Table 6 reports results for Gendarmenmarkt when removing 45% of outlying columns. We note that STE significantly reduces the pipeline’s runtime and improves the accuracy for both rotations and translations. TME also exhibits improvement, but it is less accurate and is also slower than STE. On the other hand, FMS and SFMS do not improve the accuracy. While the resulting errors are still large, their improvement shows some potential in dealing with difficult SfM structure by initially removing cameras in a way that may help eliminate some scene ambiguities, which are prevalent in Gendarmenmarkt.

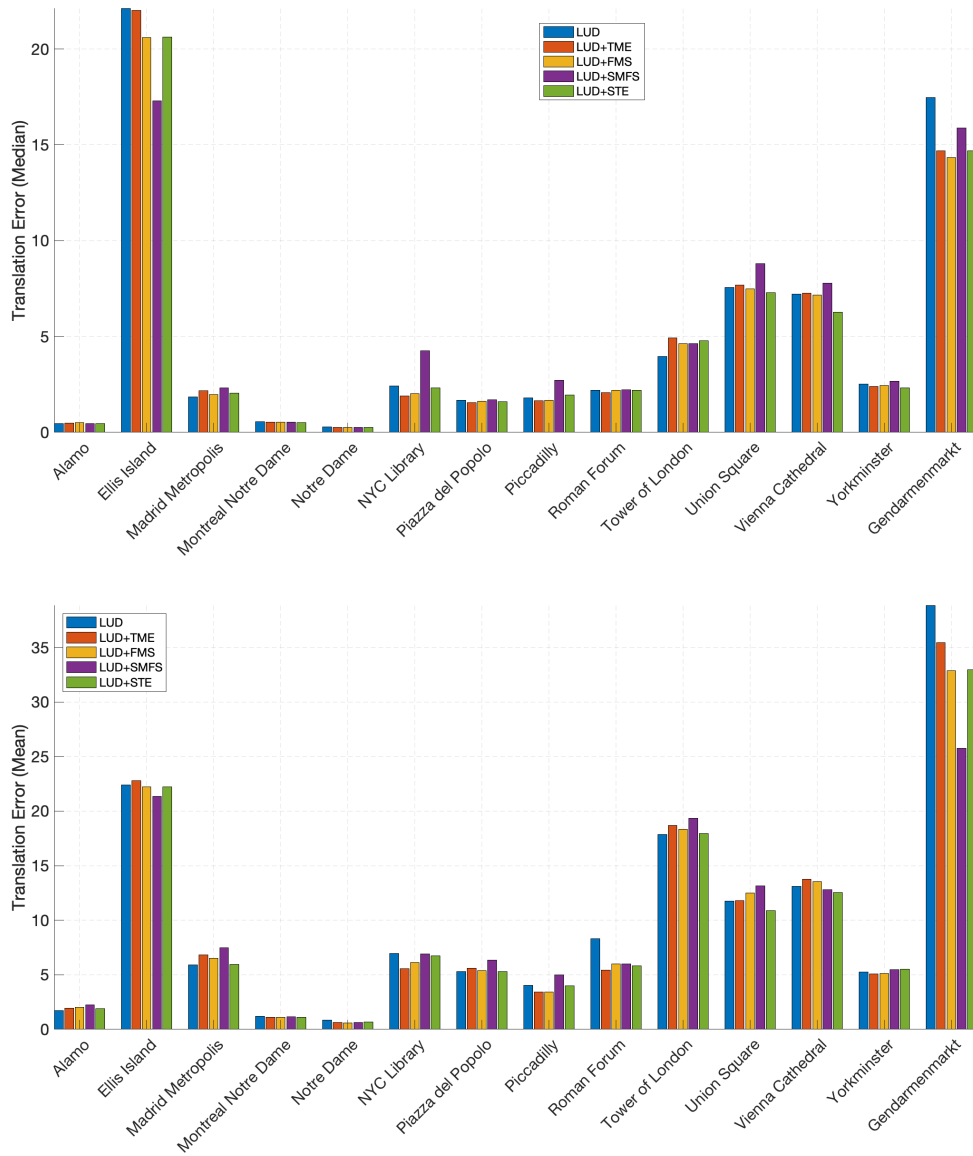


Figure 9. Median (top) and mean (bottom) absolute translation errors (in meters) of LUD and four RSR methods used to initially screen bad cameras within LUD applied to the 14 datasets of Photo Tourism.

Finally, in order to get an idea about the behavior of the outliers, Figure 11 plots the distances of all data points to the STE subspace, where it uses a logarithmic scale. Distinguished outliers are noticed in all datasets, but Piazza del Popolo. Moreover, such significant outliers constitute only a small portion of the columns. As mentioned in the main text, we preferred to avoid heuristic methods for the cutoff of outliers, and thus assumed a fixed percentage of 20% outlying column vectors.

Location	n N		LUD						LUD+STE						LUD+SFMS					
			n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}
Alamo	570	606,963	563	1.74	0.47	3.04	1.03	328.8	497	1.90	0.45	3.08	1.04	242.6	466	2.24	0.46	4.44	1.81	183.3
Ellis Island	230	178,324	227	22.41	22.10	2.14	0.71	44.6	203	22.22	20.62	1.73	0.61	34.9	194	21.35	17.30	2.28	0.77	34.7
Madrid Metropolis	330	187,790	319	5.93	1.84	3.56	0.93	45.0	281	5.97	2.05	3.21	0.95	27.3	288	7.47	2.33	4.91	1.59	28.5
Montreal N.D.	445	633,938	437	1.22	0.56	0.96	0.49	237.8	388	1.12	0.52	1.02	0.52	159.7	381	1.15	0.54	1.05	0.49	149.5
Notre Dame	547	1,345,766	543	0.85	0.29	2.15	0.66	894.0	473	0.69	0.25	1.68	0.63	763.7	480	0.62	0.25	1.60	0.65	785.6
NYC Library	313	259,302	309	6.95	2.42	2.41	1.19	53.7	276	6.76	2.31	2.58	1.27	39.5	237	6.93	4.27	3.09	1.55	87.7
Piazza del Popolo	307	15,791	298	5.29	1.67	2.90	1.64	65.2	258	5.30	1.60	3.06	1.81	36.6	245	6.34	1.70	3.33	2.04	36.2
Piccadilly	2,226	1,278,612	2,171	4.04	1.81	4.82	1.72	1,649.0	1,937	3.98	1.95	4.37	1.76	857.4	1,827	4.98	2.72	5.33	2.01	654.4
Roman Forum	995	890,945	967	8.32	2.20	2.04	1.38	272.2	810	5.84	2.19	2.47	1.49	170.4	811	5.98	2.23	2.28	1.48	182.9
Tower of London	440	474,171	431	17.86	3.96	2.92	2.13	65.2	355	17.95	4.79	3.01	2.63	44.9	338	19.35	4.62	3.64	2.92	43.1
Union Square	733	323,933	715	11.75	7.57	6.59	3.21	53.7	651	10.90	7.29	6.21	3.47	53.8	571	13.14	8.81	6.93	3.92	36.5
Vienna Cathedral	789	1,361,659	774	13.10	7.21	7.28	1.65	762.6	685	12.56	6.28	7.91	1.81	547.1	666	12.82	7.78	8.08	1.80	365.6
Yorkminster	412	525,592	405	5.25	2.51	2.23	1.43	91.9	359	5.51	2.32	2.21	1.48	62.9	335	5.49	2.66	2.40	1.60	52.2
Gendarmenmarkt	671	338,800	654	38.82	17.46	39.63	6.43	124.2	596	32.94	14.70	29.01	4.65	79.2	564	25.77	15.88	38.16	5.50	77.8

Table 4. Performance of the LUD, LUD+STE and LUD+SFMS pipelines on the Photo Tourism datasets: n and N are the number of cameras and key points, respectively (when applying screening, n also denotes the number of the remaining cameras); \hat{e}_R, \tilde{e}_R indicate mean and median errors of absolute camera rotations in degrees, respectively; \hat{e}_T, \tilde{e}_T indicate mean and median errors of absolute camera translations in meters, respectively; T_{total} is the runtime of the pipeline (in seconds).

Location	n N		LUD+TME						LUD+FMS					
			n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}
Alamo	570	606,963	463	1.96	0.48	3.04	1.07	216.6	480	2.05	0.51	3.26	1.10	203.0
Ellis Island	230	178,324	205	22.80	22.01	2.87	0.93	39.6	207	22.22	20.60	2.75	0.90	38.4
Madrid Metropolis	330	187,790	288	6.83	2.17	5.73	2.00	34.0	289	6.52	1.97	4.80	1.77	40.0
Montreal N.D.	445	633,938	381	1.10	0.53	0.96	0.48	155.8	389	1.12	0.54	0.97	0.50	167.0
Notre Dame	547	1,345,766	463	0.62	0.25	1.59	0.63	747.4	479	0.61	0.26	1.37	0.66	765.9
NYC Library	313	259,302	267	5.55	1.91	2.53	1.15	41.6	263	6.14	2.03	2.52	1.25	42.1
Piazza del Popolo	307	15,791	251	5.61	1.56	2.96	1.68	41.8	257	5.40	1.62	3.05	1.69	41.3
Piccadilly	2,226	1,278,612	1,980	3.44	1.65	3.19	1.77	1,211.7	1,888	3.43	1.68	3.20	1.75	1,172.9
Roman Forum	995	890,945	828	5.42	2.08	2.34	1.46	157.0	814	6.01	2.19	2.35	1.38	185.8
Tower of London	440	474,171	364	18.67	4.93	3.02	2.39	48.8	359	18.34	4.62	3.02	2.62	40.9
Union Square	733	323,933	612	11.80	7.68	6.18	3.41	49.4	610	12.49	7.48	6.13	3.37	44.1
Vienna Cathedral	789	1,361,659	694	13.77	7.26	7.69	2.89	468.7	692	13.54	7.17	7.70	2.90	472.0
Yorkminster	412	525,592	360	5.07	2.40	2.15	1.47	70.8	358	5.11	2.44	2.19	1.51	73.3
Gendarmenmarkt	671	338,800	602	35.43	14.69	28.30	4.08	96.7	600	32.87	14.34	28.59	4.09	85.0

Table 5. Performance of the LUD+TME and LUD+FMS pipelines on the Photo Tourism datasets: n and N are the number of cameras and key points, respectively (when applying screening, n also denotes the number of the remaining cameras); \hat{e}_R, \tilde{e}_R indicate mean and median errors of absolute camera rotations in degrees, respectively; \hat{e}_T, \tilde{e}_T indicate mean and median errors of absolute camera translations in meters, respectively; T_{total} is the runtime of the pipeline (in seconds).

Method	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}
LUD	654	38.82	17.46	39.63	6.43	124.2
LUD+STE	497	22.82	12.53	26.76	3.60	51.24
LUD+TME	516	25.07	14.72	24.93	4.18	64.35
LUD+FMS	559	40.74	18.6	40.28	6.54	102.14
LUD+SFMS	549	41.31	19.33	40.65	6.47	100.40

Table 6. Comparison of the four pipelines when 45% outlying columns are eliminated in Gendarmenmarkt. Recall that \hat{e}_R, \tilde{e}_R indicate mean and median errors of absolute camera rotations in degrees, respectively; \hat{e}_T, \tilde{e}_T indicate mean and median errors of absolute camera translations in meters, respectively; and T_{total} is the runtime of the pipeline (in seconds).

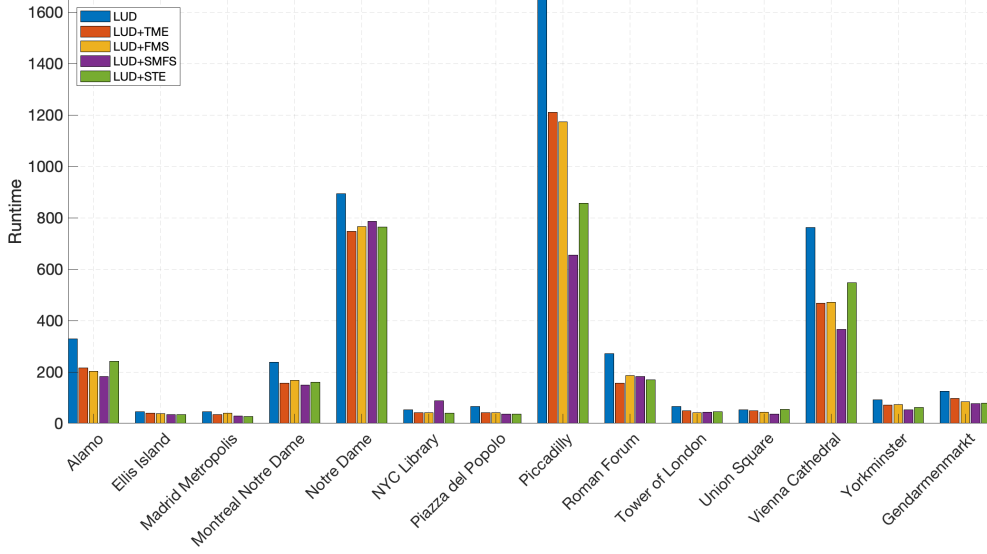


Figure 10. Runtime (in seconds) of the LUD pipeline and four LUD-type pipelines including an RSR method to initially screen bad cameras, applied to the 14 datasets of Photo Tourism.

A.2.1 Initial Camera Removal for SfM without Matrix Completion

For completeness, we also provide the numerical results for the initial camera removal experiment when the absent blocks \mathbf{E}_{ij} are assigned zero matrices. Table 7 and Table 8 showcase the performance of the LUD+RSR pipelines without the matrix completion step. Overall, integrating matrix completion into the LUD+RSR pipelines enhances their performance. The most notable enhancement for STE is for the mean translation error of the Roman Forum and Gendarmenmarkt. For other RSR algorithm we notice even more significant improvement with matrix completion. In other words, without matrix completion, the higher accuracy of STE in comparison to other RSR methods is more noticeable.

Location	n N		LUD						LUD+STE						LUD+SFMS					
			n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}
Alamo	570	606,963	563	1.74	0.47	3.04	1.03	328.8	518	1.80	0.45	3.02	1.03	283.2	466	2.24	0.47	4.44	1.82	219.0
Ellis Island	230	178,324	227	22.41	22.10	2.14	0.71	44.6	209	22.51	20.78	2.81	0.90	41.0	195	21.51	18.50	2.27	0.77	33.3
Madrid Metropolis	330	187,790	319	5.93	1.84	3.56	0.93	45.0	290	5.69	1.81	3.24	0.99	37.6	278	5.82	2.25	3.54	1.12	32.1
Montreal N.D.	445	633,938	437	1.22	0.56	0.96	0.49	237.8	400	1.09	0.55	1.01	0.50	186.5	383	1.14	0.54	1.03	0.53	146.5
Notre Dame	547	1,345,766	543	0.85	0.29	2.15	0.66	894.0	495	0.79	0.27	2.15	0.64	798.4	476	0.72	0.26	1.87	0.66	815.6
NYC Library	313	259,302	309	6.95	2.42	2.41	1.19	53.7	278	6.78	2.51	2.57	1.20	42.6	238	6.82	4.61	3.13	1.49	42.2
Piazza del Popolo	307	15,791	298	5.29	1.67	2.90	1.64	65.2	247	5.60	1.69	2.22	1.24	35.7	242	6.62	1.59	3.51	2.15	45.9
Piccadilly	2,226	1,278,612	2171	4.04	1.81	4.82	1.72	1,649.0	1976	4.29	1.98	4.76	1.71	950.7	1827	4.98	2.72	5.33	2.01	661.6
Roman Forum	995	890,945	967	8.32	2.20	2.04	1.38	272.2	849	10.04	2.33	2.18	1.45	175.6	814	9.46	2.68	2.15	1.50	150.5
Tower of London	440	474,171	431	17.86	3.96	2.92	2.13	65.2	373	18.83	3.21	2.80	2.01	53.2	351	22.31	4.10	3.18	2.46	27.0
Union Square	733	323,933	715	11.75	7.57	6.59	3.21	53.7	624	11.92	7.55	6.24	3.59	49.6	558	13.42	9.04	6.87	3.48	38.5
Vienna Cathedral	789	1,361,659	774	13.10	7.21	7.28	1.65	762.6	715	12.78	6.26	7.66	1.66	590.2	666	12.82	7.78	8.08	1.80	444.0
Yorkminster	412	525,592	405	5.25	2.51	2.23	1.43	91.9	368	5.26	2.50	2.22	1.50	73.2	338	5.67	3.13	2.41	1.56	48.0
Gendarmenmarkt	671	338,800	654	38.82	17.46	39.63	6.43	124.2	616	37.60	15.49	31.29	4.75	108.3	543	35.34	27.53	62.89	31.04	83.7

Table 7. Performance of the LUD, LUD+STE and LUD+SFMS pipelines **without matrix completion** on the Photo Tourism datasets: n and N are the number of cameras and key points, respectively (when applying screening, n also denotes the number of the remaining cameras); \hat{e}_R, \tilde{e}_R indicate mean and median errors of absolute camera rotations in degrees, respectively; \hat{e}_T, \tilde{e}_T indicate mean and median errors of absolute camera translations in meters, respectively; T_{total} is the runtime of the pipeline (in seconds).

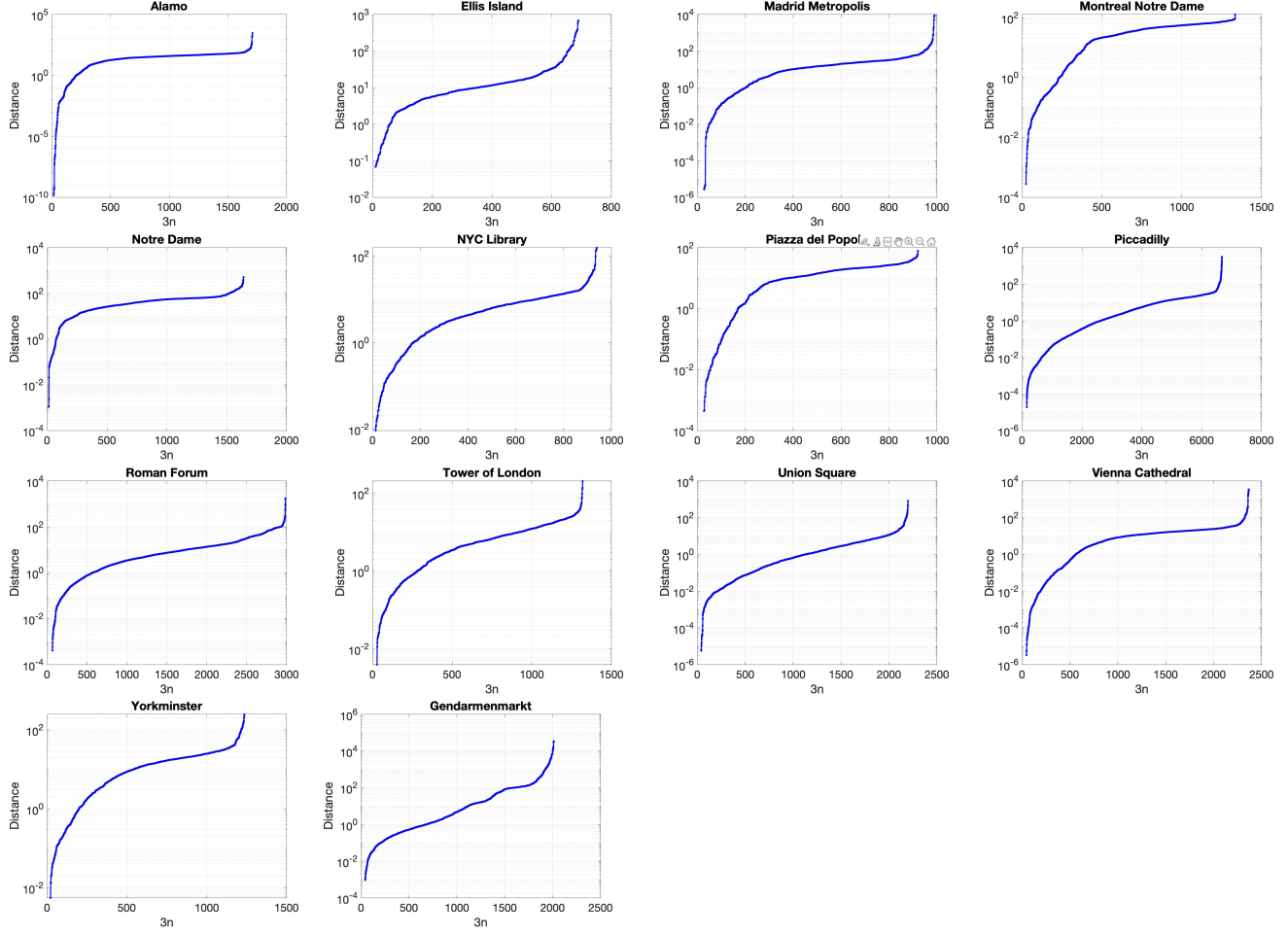


Figure 11. Distance of columns of \mathbf{E} to the subspace recovered by STE (in log scale). The distances are sorted and plotted in descending order. All datasets exhibit significant outlying columns.

Location	n	N	LUD+TME						LUD+FMS					
			n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}	n	\hat{e}_T	\tilde{e}_T	\hat{e}_R	\tilde{e}_R	T_{total}
Alamo	570	606,963	463	1.96	0.48	3.04	1.07	216.6	480	2.05	0.51	3.26	1.10	203.0
Ellis Island	230	178,324	205	22.80	22.01	2.87	0.93	39.6	207	22.22	20.60	2.75	0.90	38.4
Madrid Metropolis	330	187,790	288	6.83	2.17	5.73	2.00	34.0	289	6.52	1.97	4.80	1.77	40.0
Montreal Notre Dame	445	633,938	381	1.10	0.53	0.96	0.48	155.8	389	1.12	0.54	0.97	0.50	167.0
Notre Dame	547	1,345,766	463	0.62	0.25	1.59	0.63	747.4	479	0.61	0.26	1.37	0.66	765.9
NYC Library	313	259,302	267	5.55	1.91	2.53	1.15	41.6	263	6.14	2.03	2.52	1.25	42.1
Piazza del Popolo	307	15,791	251	5.61	1.56	2.96	1.68	41.8	257	5.40	1.62	3.05	1.69	41.3
Piccadilly	2,226	1,278,612	1980	3.44	1.65	3.19	1.77	1,211.7	1888	3.43	1.68	3.20	1.75	1,172.9
Roman Forum	995	890,945	828	5.42	2.08	2.34	1.46	157.0	814	6.01	2.19	2.35	1.38	185.8
Tower of London	440	474,171	364	18.67	4.93	3.02	2.39	48.8	359	18.34	4.62	3.02	2.62	40.9
Union Square	733	323,933	612	11.80	7.68	6.18	3.41	49.4	610	12.49	7.48	6.13	3.37	44.1
Vienna Cathedral	789	1,361,659	694	13.77	7.26	7.69	2.89	468.7	692	13.54	7.17	7.70	2.90	472.0
Yorkminster	412	525,592	360	5.07	2.40	2.15	1.47	70.8	358	5.11	2.44	2.19	1.51	73.3
Gendarmenmarkt	671	338,800	602	35.43	14.69	28.30	4.08	96.7	600	32.87	14.34	28.59	4.09	85.0

Table 8. Performance of the LUD+TME and LUD+FMS pipelines **without matrix completion** on the Photo Tourism datasets: n and N are the number of cameras and key points, respectively (when applying screening, n also denotes the number of the remaining cameras); \hat{e}_R, \tilde{e}_R indicate mean and median errors of absolute camera rotations in degrees, respectively; \hat{e}_T, \tilde{e}_T indicate mean and median errors of absolute camera translations in meters, respectively; T_{total} is the runtime of the pipeline (in seconds).