Projective Multi-view Structure and Motion from Element-wise Factorization

Yuchao Dai, Student Member, IEEE, Hongdong Li, Member, IEEE, and Mingyi He, Member, IEEE

Abstract—The Sturm-Triggs type iteration is a classic approach for solving the projective structure-from-motion (SfM) factorization problem, which iteratively solves the projective depths, scene structure and camera motions in an alternated fashion. Like many other iterative algorithms, the Sturm-Triggs iteration suffers from common drawbacks such as requiring a good initialization, the iteration may not converge or only converge to a local minimum, etc. In this paper, we formulate the projective SfM problem as a novel and original element-wise factorization (i.e., Hadamard factorization) problem, as opposed to the conventional matrix factorization. Thanks to this formulation, we are able to solve the projective depths, structure and camera motions simultaneously by convex optimization. To address the scalability issue, we adopt a continuation based algorithm. Our method is a global method, in the sense that it is guaranteed to obtain a globally-optimal solution up to relaxation gap. Another advantage is that, our method can handle challenging real-world situations such as missing data and outliers quite easily, and all in a natural and unified manner. Extensive experiments on both synthetic and real images show comparable results compared with the state-of-the-art methods.

Index Terms—Element-wise factorization, projective structure and motion, semi-definite programming, missing data, outlier.

1 INTRODUCTION

Structure and motion recovery from multi-view image feature correspondences is a fundamental problem in computer vision. The Tomasi-Kanade’s factorization framework [1] is one of the most influential methods in multi-view structure-from-motion (SfM). This approach is not only theoretically important, which reveals the deep and elegant mathematical structure of the SfM problem, but also practically useful, thanks to its conceptual and computational simplicity. Given a multi-view measurement matrix \( H \), the factorization approach simultaneously solves the (stacked) camera projection matrix \( P \) (the camera motions) and the 3D structure \( X \), via a simple matrix factorization \( H = PX \) through a single Singular Value Decomposition (SVD). Its nice property also comes from the fact that it treats all points and all camera frames uniformly, no point or frame is deemed more “privileged” or “preferred” than the others.

Originally, the Tomasi-Kanade’s factorization framework was developed for the orthographic camera with complete feature correspondences. In this paper, we revisit the projective extension of the factorization method in order to deal with perspective camera model. In particular, we are motivated by the most popular algorithm-of-choice for projective factorization—the iterative Sturm-Triggs method \(^1\) [2], [3]. The basic idea of the Sturm-Triggs iteration is simple: compared with the affine camera case, now the basic projective imaging process can be compactly written as \( \Lambda \odot H = PX \), where \( \Lambda \) is a properly stacked but unknown projective depth matrix, \( \odot \) denotes the element-wise (Hadamard) multiplication. Since the depth matrix \( \Lambda \), the right-hand factorization \( P \) and \( X \) are unknowns, a natural approach to solve this problem is through iterative alternation till converge. Often, the initial point to start with is to assign all depths to be 1. In other words, this is equivalent to starting from affine camera model. Many other projective generalizations, such as [4], [5], [6], [7], share a similar computational pattern in terms of alternation.

Like any other iterative algorithms, the Sturm-Triggs iteration and its extensions have some common drawbacks. For example, the iterative algorithms all require a good initial point to start with; the iteration procedure may not converge; even if it does converge (theoretically or empirically), it may only converge to a local minimum; global optimality can hardly be guaranteed.

However, for the particular method of the Sturm-Triggs iteration and the alikes, all the aforementioned drawbacks did have been observed in all occasions. Indeed, Mahamud and Hebert [8] pointed out that the iterative Sturm-Triggs method with column-wise and row-wise normalization is not guaranteed to converge in theory. To salvage this problem, they proposed a column-wise only normalization and derived a provably-convergent iterative algorithm (called column-space method) [8]. Oliensis and Hartley also observed situations where the iterations fell into a limiting cycle.

1. Note that the original version of the Sturm-Triggs method reported in [2] is non-iterative, which crucially relies on accurate depth estimation from pair-wise epipolar geometry.
and never converged [9]. Hartley and Zisserman [10] discussed when the popular choice of initialization—assumption all depths to be one—is a sensible choice for the iteration. They concluded that this assumption works only when the ratios of true depths of the different 3D point \( x_j \) remain approximately constant during a sequence. Consequently, to make the Sturm-Triggs iteration work, the ground truth solution has to be rather close to the affine case.

Even worse, a recent complete theoretical analysis delivers even more negative messages [9]. Oliensis and Hartley showed that (1) the simplest iterative extension of the Sturm-Triggs algorithm (SIESTA) without normalization/balancing will always converge to the trivial solution; (2) paper [8]’s provably-convergent iteration method will generally converge to a useless solution; (3) applying both column-wise and row-wise normalization may possibly run into unstable state during iteration. They also provided a remedy, i.e., a regularization-based iterative algorithm (called CIESTA) that can converge to a stable solution, albeit the solution is biased–biased towards all estimated depths being close to one.

When the actual configurations for cameras and points are far away from affine configuration, the Sturm-Triggs algorithm and its extensions tend to fail to converge to good solution. Yet, it would be nicer if a projective factorization algorithm can be made free from these theoretical drawbacks and be useful in practice.

Meanwhile, in most real-world SfM applications, due to self occlusions, points behind cameras (i.e., chirality) or limitations of the feature trackers, missing data are inevitable. However the Tomasi-Kanade’s factorization framework and the Sturm-Triggs algorithm could not be extended to deal with missing data naturally. This constitutes a major drawback of factorization-based methods. Substantially, the unknown projective depths will make the problem much harder.

Furthermore, outliers are unavoidable in real image projective reconstruction due to mismatches/mistracks (with the tracker unaware), causing measurement matrix to have some entries with gross errors (a corrupted measurement matrix). Even after some robust estimators such as RANSAC (Random Sample Consensus [11]) have been applied, there are still small proportion of outliers.

Therefore a practical factorization framework should be able to handle complete measurements, incomplete measurements, and outliers elegantly. Our new method, to be presented next, is one of this kind. Part of the work was published in ECCV’10 [12]. However, this journal version contains nontrivial and substantial extensions, in both theory and practical solution techniques.

1.1 Related Work in Scene Reconstruction Pipeline

The factorization based structure and motion recovery method belongs to the family of batch reconstruction, where the goal is to reconstruct a scene in one step rather than in an incremental way. Besides the batch reconstruction method, another family of reconstruction methods is incremental or sequential reconstruction methods. Theses methods initialize from two or three views and add points and cameras via triangulation and camera resection. After adding each view, bundle adjustment [13] is used to refine the reconstruction. These methods have successfully built 3D models from large scale unordered Internet community image collections [14] [15] [16].

However, these incremental methods tend to be computationally intensive, making repeated use of bundle adjustment [13] as well as outlier rejection to remove inconsistent measurements. Additionally, these methods do not treat all images equally as in factorization framework, producing different results depending on the order of the photos being considered.

Recently, Crandall et al. [17] showed that robust global objective functions can be used in SfM and optimized using discrete belief propagation techniques. Olsson and Enqvist [18] presented a non-incremental approach to SfM based on robustly computing global rotations from relative geometries and feeding these into the known-rotation framework to initialize bundle adjustment.

1.2 Outline of Our Approach

In this paper, we propose a unified framework for projective factorization, which stays away from all the above mentioned theoretical traps. Our method is global and no initial guess is needed. Our key idea is that: mathematically, instead of formulating the problem as matrix factorization, we reformulate the problem as element-wise factorization that solves the projective depth matrix \( \Lambda \), projection matrix \( P \) and structure \( X \) simultaneously. This simple change offers many benefits in analyzing and solving the problem. Our method can handle complete measurements, incomplete measurements, outliers and degenerate case all in a natural and unified manner.

The numerical implementation of our approach is efficient, as we formulate the problem as semi-definite programming (SDP) for which powerful off-the-shelf solvers are available. To address the scalability issue, we adopt an alternating direction continuation based algorithm, leading to much more efficient computations. The proposed algorithm is more suitable for real world large-scale SfM applications.

2 Element-Wise Factorization

In this section, we present our element-wise factorization framework for projective multi-view structure and motion under complete measurement case. We analyze various balancing strategies. Additionally, we give analysis on the existence and uniqueness of our formulation.

2.1 Problem Statement

The problem of projective multi-view structure and motion seeks to simultaneously solve the unknown projective depths, the unknown camera motions and the
unknown structure. Compared with affine factorization, this is a much harder problem, mainly because depths are not known a priori. Of course, one can compute these projective depths by other means, e.g., via a string of fundamental matrices or trifocal tensors, via a reference plane, etc. However, such a two-step approach would diminish the elegance of the factorization algorithm.

Consider \( n \) stationary 3D points \( x_i = [x_j, y_j, z_j]^T, j = 1, \ldots, n \) observed by all \( m \) perspective cameras \( P_i, i = 1, \ldots, m \), under perspective camera model, the \( j \)-th 3D point \( x_j \) is projected onto image point \( m_{ij} = [u_{ij}, v_{ij}, 1]^T \) by

\[
m_{ij} = \frac{1}{\lambda_{ij}} P_i x_j,
\]

where \( \lambda_{ij} \) is a scale factor, commonly called “projective depth”. It is easy to see that \( \lambda_{ij} = 1 \) when the camera reduces to an affine camera.

Collecting the image measurements over all frames, we form a measurement matrix \( M = [m_{ij}] \) of size \( 3m \times n \). Now the relationship (1) can be compactly written as:

\[
M = \left[ \frac{1}{\lambda_{ij}} \right]_{ij} \otimes (PX),
\]

where \( P \in \mathbb{R}^{3m \times 4} \) and \( X \in \mathbb{R}^{4 \times n} \) are properly stacked projection matrix and structure matrix. Note that each row of the inverse depth matrix is repeated 3 times. We must have a rank at most 4, matrix \( M \) also yields a decomposable measurement matrix \( [4] \), where \( \mu_i, \eta_j \) are arbitrary coefficients. Usually, to handle ambiguity and avoid possible trivial solutions (e.g., all depths being zero, or all but 4 columns of the depth matrix being zero, etc.), some kind of column-wise and row-wise normalization/balancing is necessary.

The Sturm-Triggs type iterative algorithms [2], [3] solve the projective factorization problem through alternation: (1) fix \( \Lambda \), solve for \( P \) and \( X \) via SVD; (2) fix \( P \) and \( X \), solve for \( \Lambda \) via least squares; (3) alternate between the above two steps till converge. In [2], the rescaled measurement matrix \( \tilde{W} \) is normalized through normalizing each column and triplet of rows iteratively.

Triggs [3] added a third balancing stage to the Sturm-Triggs type iterative algorithms that rescales the projective depth \( \lambda_{ij} \) to make them close to 1. This lessens the algorithm’s bias and helps to steer it away from trivial minima. The balancing can be done in two passes, by first scaling \( \lambda_{ij} \) for each image \( i \) as \( \lambda_{ij} \to \alpha_i \lambda_{ij}, \sum_{j=1}^{n} |\lambda_{ij}|^2 = n \) and then scaling \( \lambda_{ij} \) for each point \( j \) as \( \lambda_{ij} \to \beta_j \lambda_{ij}, \sum_{i=1}^{m} |\lambda_{ij}|^2 = m \), where \( \alpha_i, \beta_j \) are scaling coefficients to determine.

Optionally, this procedure can be repeated several times or iterated to converge. Convergence of this balancing strategy is stated in the following theorem:

**Theorem 2.1:** For element-wise nonzero matrix \( \Lambda \), the \( L_2 \) norm constrained balancing strategy \( \sum_{j=1}^{n} |\lambda_{ij}|^2 = n, \sum_{i=1}^{m} |\lambda_{ij}|^2 = m \) is guaranteed to converge.

In stead of using \( L_2 \) norm balancing, we propose to apply \( L_1 \) norm balancing strategy by first scaling \( \lambda_{ij} \to \alpha_i \lambda_{ij} \) for each image \( i \) such that \( \sum_{j=1}^{n} |\lambda_{ij}| = n \) and then scaling \( \lambda_{ij} \to \beta_j \lambda_{ij} \) for each point \( j \) such that \( \sum_{i=1}^{m} |\lambda_{ij}| = m \). When all the projective depths \( \lambda_{ij} \) are positive, the above \( L_1 \) norm balancing strategy reduces to the following column-sum and row-sum balancing strategy:

\[
\lambda_{ij} \to \alpha_i \lambda_{ij}, \sum_{j=1}^{n} \lambda_{ij} = n;
\lambda_{ij} \to \beta_j \lambda_{ij}, \sum_{i=1}^{m} \lambda_{ij} = m.
\]

For the \( L_1 \) norm balancing strategy, we have the following convergence theorem:

**Theorem 2.2:** For element-wise nonzero matrix \( \Lambda \), the \( L_1 \) norm constrained balancing strategy \( \sum_{j=1}^{n} |\lambda_{ij}| = n, \sum_{i=1}^{m} |\lambda_{ij}| = m \) is guaranteed to converge.

Proofs of Theorem 2.1 and Theorem 2.2 are given in supplement material. Obviously, convergence of the column-sum and row-sum constrained balancing strategy is guaranteed in Theorem 2.2 too. Column-sum and row-sum balancing strategy will be incorporated with our formulation due to its linearity.

2.3 Element-wise Factorization Formulation

Here, we formulate the projective SfM problem as an element-wise factorization (i.e., Hadamard factorization) problem as opposed to the conventionally adopted matrix factorization.

2.3.1 Our Main Idea

We repeat the basic equation for projective factorization: \( \tilde{W} = \Lambda \odot M = PX \). Recall that \( \Lambda \) is the only input, and the task is to solve for both \( \Lambda \) and \( \tilde{W} \). Note that \( \odot \) denotes element-wise multiplication, therefore we can view the problem as an *element-wise factorization* problem, i.e.,

- Given measurement matrix \( M \), find two matrices \( \Lambda \) and \( \tilde{W} \) such that \( \tilde{W} = \Lambda \odot M \).

At first glance, this seems to be an impossible task, as the system is severely under-constrained, i.e. the number
of unknowns to infer is twice as many as the given measurements in $M$. However, for the particular problem of element-wise factorization, we have extra conditions on the unknown matrices which may hopefully sufficiently constrain the system.

Firstly, according to the projective imaging model $\hat{w} = PX$, we have $\text{rank}(\hat{w}) \leq \min(\text{rank}(P), \text{rank}(X)) \leq 4$. Thus the re-scaled measurement matrix $\hat{w}$ has a rank at most $4$. This is true for noise-free case (we will further relax this constraint in the actual computation). Under degenerate case, the rank of $\hat{w}$ will decrease in further.

Secondly, the rank of $\Lambda$ is also at most $4$. This is easy to observe, since $\Lambda$ is a sub-matrix of $\hat{w}$, hence, $\text{rank}(\hat{w}) \leq 4, \text{rank}(\Lambda) \leq \text{rank}(\hat{w}) \Rightarrow \text{rank}(\Lambda) \leq 4$. These two constraints decrease the degrees of the solution dramatically. Therefore, instead of finding two general condition matrices, we aim to find two low rank matrices $\hat{w}$ and $\Lambda$ such that $\hat{w} = \Lambda \odot M$.

Thirdly, all visible points' projective depths must be positive. This is nothing but the chirality constraint commonly used in multi-view geometry [10]. In other words, visible points must lie in front of the camera. Even noise and outliers do not violate this constraint.

Fourthly, all the columns and rows of $\Lambda$ have been normalized to have (average) unit sum that is $\sum_i \lambda_{ij} = m, \sum_j \lambda_{ij} = n$. These column-sum and row-sum constraints play two roles: (1) rule out trivial solutions such as all $\lambda_{ij} = 0$, all columns but four are zeros and etc; (2) rule out scale ambiguity in the factorization.

These constraints are in fact very mild, reasonable, and not restrictive. Roughly speaking, these extra conditions are expected to supply the system with sufficient constraints, making the element-wise factorization problem well-posed and hence solvable. A rigorous proof is provided in supplemental material.

### 2.3.2 Formulation

Taking all the constraints into consideration, mathematically, element-wise factorization for projective multi-view structure and motion under complete measurements case is formulated as:

Find $\hat{w}, \Lambda$, such that,

\[
\hat{w} = \Lambda \odot M,
\]

\[
\text{rank}(\hat{w}) \leq 4,
\]

\[
\sum_i \lambda_{ij} = m, j = 1, \cdots, n,
\]

\[
\sum_j \lambda_{ij} = n, i = 1, \cdots, m,
\]

\[
\lambda_{ij} > 0.
\]

(5)

We do not need to enforce the low rank constraint $\text{rank}(\Lambda) \leq 4$ on projective depths as it is naturally embodied in low rank constraint on $\hat{w}$. The constraint $\lambda_{ij} > 0$ represents the chirality constraint. The column-sum and row-sum constraints are included to avoid trivial solutions and deal with scale ambiguity.

Complete analysis on the existence and uniqueness of solution to our formulation Eq-(5) are presented in supplemental material.

### 3 SDP FORMULATION

In Section 2, we have formulated projective reconstruction problem as a novel element-wise factorization problem. To solve the element-wise factorization problem, we will relax it to nuclear norm minimization, which can be solved as an SDP.

#### 3.1 Rank Minimization

Recently rank minimization has been widely utilized in computer vision for low rank fitting problem [19] [20] [21]. Under noiseless and complete measurement case, our element-wise factorization formulation Eq-(5) with rank constraint $\text{rank}(\hat{w}) \leq 4$ is equivalent to the following rank minimization problem:

\[
\begin{align*}
\text{Minimize } & \text{rank}(\hat{w}), \\
\text{subject to } & \hat{w} = \Lambda \odot M, \\
& \Lambda^T 1_m = m 1_n, \quad \Lambda 1_n = n 1_m, \\
& \lambda_{ij} > 0.
\end{align*}
\]

(6)

where $1_m, 1_n$ are vectors with 1 element-wisely of length $m$ and $n$ respectively. Once this problem is solved, we can use $\Lambda$ as the estimated depth matrix, and $\hat{w}$ as the rescaled measurement matrix.

#### 3.2 Nuclear Norm Minimization

To solve rank minimization problem exactly is intractable in general, and in fact NP-hard [22]. To overcome this problem, nuclear-norm has been introduced as the tightest convex relaxation of the rank function. The nuclear norm of a matrix $X \in \mathbb{R}^{m \times n}$ is defined as

\[
||X||_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X),
\]

(7)

where $\sigma_i(X)$ is the $i$-th singular value of $X$.

Recently, using nuclear norm minimization to solve rank minimization problem has received considerable attentions, in particular in the research area of compressive sensing, also this relaxation has been used in computer vision and machine learning [19], [20]. One surprising result is that for a large class of matrices satisfying some “incoherency” or “restricted isometry” properties, nuclear norm minimization actually gives an exact solution. In other words, the relaxation gap is zero.

In this paper, we use the nuclear norm as a convex surrogate (a relaxation) to the rank function, mainly for the purpose of approximately solving our projective factorization problem. It is also worth noting that, long before the emergence of the compressive sensing research, nuclear-norm heuristics has already been known and widely used for solving rank minimization [23].

**Theorem 3.1:** The convex envelope\(^2\) of $\text{rank}(X)$ on the set $S = \{X \in \mathbb{R}^{m \times n}, ||X|| \leq 1\}$ is the nuclear norm $||X||_*$, where $||X|| = \sigma_1(X)$ is the operator norm of $X$ [24].

2. The convex envelope of $f: C \rightarrow \mathbb{R}$ is defined as the largest convex function $g$ such that $g(x) \leq f(x)$ for all $x \in C$. This means that among all convex functions, $g$ is the one that is closest (point-wise) to $f$. 

---

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
As our objective function aims to minimize the nuclear norm of \( \tilde{w} = \Lambda \odot M \), scaling of \( \Lambda \) will change the nuclear norm but not the rank. The column-sum and row-sum normalization constraints fix this scaling freedom and make all the solutions comparable.

Using nuclear norm heuristics, we replace the original objective function \( \text{rank}(\tilde{w}) \) with \( \|\tilde{w}\|_* \), and reach the following nuclear norm minimization formulation:

\[
\min \|\tilde{w}\|_*, \text{ subject to,} \\
\tilde{w} = \Lambda \odot M, \\
A^T1_m = m1_n, \quad A1_n = n1_m, \\
\lambda_{ij} > 0.
\] (8)

Furthermore, the nuclear norm minimization \( \min \|\tilde{w}\|_* \) can be rewritten as an equivalent SDP problem:

\[
\min \frac{1}{2}(\text{tr}(X) + \text{tr}(Y)) \\
\text{s.t.} \begin{pmatrix} X & \tilde{w} \\ \tilde{w}^T & Y \end{pmatrix} \succeq 0.
\]

Such an equivalence is grounded on the following theorem (ref. [23]).

**Theorem 3.2:** Let \( A \in \mathbb{R}^{n \times n} \) be a given matrix and \( t \in \mathbb{R} \), then we have \( \|A\|_* \leq t \) if and only if there exist two symmetric matrices \( B = B^T \in \mathbb{R}^{n \times n} \) and \( C = C^T \in \mathbb{R}^{n \times n} \) such that \( \text{tr}(B) + \text{tr}(C) \leq 2t \) and

\[
\begin{pmatrix} B & A \\ A^T & C \end{pmatrix} \succeq 0.
\]

There are other formulations of the nuclear norm based on factorization of \( Z \) into two matrices \( U \) and \( V \). An illustrative result is shown in Fig. 1 in logarithm. We demonstrate the singular value distribution of ground truth, affine approximation (\( \lambda_{ij} = 1 \)), result of our method and result of our method after low-rank projection. Our method outputs reprojection error 6.59 \times 10^{-3} \) pixels before low-rank projection and 3.68 \times 10^{-9} \) pixels after low-rank projection. Our method achieves almost “perfect recovery” result even the ground truth configuration is far away from affine approximation.

### 4 Numerical Implementation

To address the scalability issue in the convex optimization, we design algorithms for large scale nuclear norm minimization based on recent progress in compressive sensing (such as [28], [29], [30], [31]).

#### 4.1 Fixed Point Continuation

First, the Lagrangian version for Eq-(8) without enforcing balancing constraints is:

\[
\mu\|\tilde{w}\|_* + \frac{1}{2}\|\tilde{w} - \Lambda \odot M\|^2_F,
\] (11)

where \( \mu \) is a trade-off parameter. The residual entry \( \frac{1}{2}\|\tilde{w} - \Lambda \odot M\|^2_F \) can be equivalently expressed as:

\[
\frac{1}{2}\|\tilde{w} - \Lambda \odot M\|^2_F = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (w_{ij}^T w_{ij} - (m_{ij}^T m_{ij})^2),
\] (12)

where the equality originates from the fact that the \( \lambda_{ij} \) minimizes the residual is defined as:

\[
\lambda_{ij} = \min_{\lambda_{ij}} \|w_{ij} - \lambda_{ij} m_{ij}\|^2_F = \frac{w_{ij}^T m_{ij}}{m_{ij}^T m_{ij}},
\] (13)
Denote \( g(w) \) as the gradient of \( \frac{1}{2} \|W - \Lambda \circ M\|^2_F \). At the position \( w \), as \( w_{ij} \) influences residual \( \frac{1}{2} \|W - \Lambda \circ M\|^2_F \) elementwisely, the gradient \( g(w) \) can be computed elementwise as:

\[
g(w_{ij}) = \frac{\partial}{\partial w_{ij}} \left( \frac{1}{2} \|w_{ij} - \lambda_{ij} m_{ij}\|^2_F \right) = w_{ij} - \frac{m_{ij}^T w_{ij}}{m_{ij}^T m_{ij}} m_{ij}. \tag{14}
\]

Our adopted fixed point continuation based algorithm for solving (11) can be expressed in the following two-line algorithm:

\[
\begin{align*}
Y(k) &= \tilde{W}(k) - \tau g(Y(k)), \\
\tilde{W}(k+1) &= S_{\tau \mu}(Y(k)),
\end{align*}
\tag{15}
\]

where \( \tau \) is the step size in each iteration and \( S_{\tau \mu}(\cdot) \) is the matrix shrinkage operator [29].

To get rid of the trivial solution \( \tilde{W} = 0 \), after updating \( Y(k) \), we firstly recover the projective depth \( \lambda_{ij}^{(k+1)} = \frac{\langle w_{ij}^T m_{ij} \rangle}{m_{ij}^T m_{ij}} \), then we balance \( \lambda_{ij}^{(k+1)} = [\lambda_{ij}^{(k+1)}]_{ij} \) to satisfy the column-sum and row-sum constraints. This is done by iteratively column and row normalization.

### 4.2 Alternating Direction Continuation

The above strategy works well in practice. However during each iteration, it involves computing SVD of the rescaled measurement matrix \( \tilde{W} \) of size \( 3m \times n \). To further speed up the implementation, we resort to Augmented Lagrangian multiplier method which has been widely used in low-rank representation [31, 32].

Based on Lemma 3.3, we slightly modify our formulation Eq.(8) following the low rank parametrization [22]:

\[
\text{Minimize} \quad \frac{1}{2} \|LR^T - \Lambda \circ M\|^2_F + \frac{\mu}{2} \|L\|^2_F + \|R\|^2_F, \tag{16}
\]

s.t. \( \Lambda^T 1_m = m1_n, \quad \Lambda 1_n = n1_m, \quad \lambda_{ij} > 0 \),

where \( L \in \mathbb{R}^{3m \times 4}, R \in \mathbb{R}^{n \times 4} \) give the explicit low rank representation. In the formulation, we apply a continuation technique (i.e., homotopy approach) for accelerating the convergence of the algorithm. The parameter \( \eta_\mu \) determines the rate of reduction of consecutive \( \mu_i \),

\[
\eta_{\mu+1} = \max\{\eta_{\mu_1}, \mu\}, \quad l = 1, \ldots, L - 1. \tag{17}
\]

To develop an efficient numerical solver for our formulation Eq.(16), we resort to the classic alternating direction method. We introduce Lagrangian multipliers \( \Gamma, \Upsilon \) and \( \Lambda \) to remove the equality constraints. The resulting augmented Lagrangian formulation states as:

\[
\begin{align*}
\mathcal{L}(\Lambda, L, R, \Gamma, \Upsilon) &= \frac{1}{2} \|LR^T - \Lambda \circ M\|^2_F + \frac{\mu}{2} \|L\|^2_F + \|R\|^2_F \\
&\quad \quad + \langle \Gamma, \Lambda^T 1_m - m1_n \rangle + \langle \Upsilon, \Lambda 1_n - n1_m \rangle + \frac{\beta}{2} \|\Lambda^T 1_m - m1_n\|^2_F + \|\Lambda 1_n - n1_m\|^2_F,
\end{align*}
\tag{18}
\]

where \( \beta > 0 \) is a penalty parameter. We minimize Eq.(18) with respect to \( L, R \) and \( \Lambda \) one at a time while fixing the others, and then update the Lagrangian multipliers \( \Gamma \) and \( \Upsilon \) as:

\[
\begin{align*}
\Gamma + \beta (\Lambda^T 1_m - m1_n), \\
\Upsilon + \beta (\Lambda 1_n - n1_m), \\
\beta_+ &\left\langle \min(\beta, \rho \beta) \right\rangle.
\end{align*}
\tag{20}
\]

The update of \( L \) and \( R \) have closed-form solutions as:

\[
\begin{align*}
L_+ &= (\Lambda \circ M) R (R^T R + \mu I)^{-1}, \\
R_+ &= (\Lambda \circ M)^T L (L^T L + \mu I)^{-1}.
\end{align*}
\tag{21}
\]

The update of projective depth \( \lambda_{ij} \) Eq.(21) has a closed-form solution too. However, the equation system is of size \( 3mn \times 3mn \). To further speed up the implementation, we propose to avoid solving for this linear equation system. We compute the projective depth elementwisely from the updated rescaled measurement matrix \( \tilde{W}_+ = L_+ R^T_+ \) and measurements \( M \) as in Eq.(13).

Then we balance the projective depth matrix \( \Lambda \) to satisfy the column-sum and row-sum constraints. In this way, we even do not need to update the Lagrangian multipliers \( \Gamma, \Upsilon \) and \( \beta \). Additionally, we only solve matrix inverse (small scale \( 4 \times 4 \) instead of computing SVD of large scale matrix as opposed to FPC based implementation).

Therefore, we obtain alternating direction continuation based Algorithm 1 for element-wise factorization under complete measurements case. As all the image measurements have been normalized first, we set \( \mu_0 = 1 \), and reduction rate \( \eta_\mu = 0.1 \) for complete measurements case. Alternating direction continuation method is exclusively used in experiments except being noted.

### 4.3 Theoretical analysis

In this subsection, we give theoretical analysis to our numerical implementations of proposed element-wise factorization formulation, i.e., computational complexity, convergence and relationship with existing work.

#### 4.3.1 Computational Complexity

The proposed algorithm has a fixed number of outer iteration (10 in all our experiments). During each outer iteration, it iteratively solves linear equation updating \( L, R \) and \( \Lambda \). Here we give a demonstrative example to show the number of inner iteration in each outer iteration in Fig. 2(a), where the number of inner iteration dramatically decreases from around 40 to less than 5.

To demonstrate the computational complexity of our algorithm, we conducted experiments under various number of points while fixing the number of cameras to 200 and also experiments under various number of cameras while fixing the number of points as 2000.
1: Initialization: Given $\Lambda^{(0)}$, set $\Psi^{(0)} = \Lambda^{(0)} \odot M$, compute $L^{(0)}, R^{(0)}$ through low rank projection, select $\epsilon, \mu$ and $\eta, \mu_1$ fix the sequence of $\mu$ as in Eq-(17).

2: for $\mu = \mu_1, \mu_2, \cdots, \mu_L$ do

3: while Not converged do

4: Update $L$ as $L^{(k+1)} = (\Lambda \odot M)R(R^T R + \mu I)^{-1}$;

5: Update $R$ as $R^{(k+1)} = (\Lambda \odot M)^T L(L^T L + \mu I)^{-1}$;

6: Update $\Psi$ as $\Psi^{(k+1)} = L^{(k+1)}R^{(k+1)}T$;

7: Update the projective depth element-wisely as $\lambda_{ij}^{(k+1)} = (w_{ij}^TM_{ij})/(m_{ij}^TM_{ij})$;

8: Balance the projective depth matrix $\Lambda^{(k+1)}$ under column-sum and row-sum constraints.

9: Compute stopping criteria $\varphi^{(k+1)} = \frac{\|w^{(k+1)} - w^{(k)}\|_F}{\max\{1, \|w^k\|_F\}}$;

10: if $\varphi^{(k+1)} < \epsilon$ then

11: Inner iteration converged;

12: end if

13: end while

14: end for

**Algorithm 1:** Element-wise factorization via alternating direction continuation.

Experimental results are illustrated in Fig. 2(b) and Fig. 2(c). Empirically, time consumption of our proposed algorithm grows linearly with respect to the number of points and the number of cameras.

![Fig. 2](a) Demonstrative example of outer and inner iteration, where 200 cameras observe 7000 points. (b) Time consumption under various number of points; (c) Time consumption under various number of cameras.

### 4.3.2 Convergence

Our numerical implementations originate from element-wise factorization formulation, which avoids the trivial solutions discussed in [9]. By introducing column-sum and row-sum constraints, our element-wise factorization formulation and its numerical implementations at least rule out trivial solutions such as all but four columns of projective depths being zero, one row of projective depths being zero and etc. Note that the alternating direction continuation solution is low rank solution, while the SDP solution and fixed point continuation solution have to be projected to low rank subspace. Convergence analysis for general alternating direction continuation method and low rank parametrization have been given in [31], [33], [34] both theoretically and empirically.

In Fig. 3, we illustrate the continuation property of the alternation direction continuation algorithm on Corridor sequence. We give the continuation parameter sequence $\mu_i$ in Fig. 3(a), number of inner iteration during each outer iteration in Fig. 3(b) and reprojection error after each outer iteration in Fig. 3(c). Empirically, with the decrease of continuation parameter $\mu$, the solution decreases the reprojection error and approaches the solution with $\mu = 0$. In our implementation, the final $\mu = 0.1^{10}$.

![Fig. 3](a) (b) (c)

**Fig. 3.** Convergence of the alternating direction continuation method on Corridor sequence. (a) Continuation parameter $\mu$; (b) Number of inner iteration during each outer iteration; (c) Reprojection error after each outer iteration.

### 4.3.3 Relationship with existing work

In the above numerical implementations, the fixed point continuation algorithm minimizes nuclear norm through first order gradient descent. The alternating direction continuation algorithm is an adaption of augmented Lagrangian multiplier method to our element-wise factorization formulation, which solves the convex semi-definite programming problem with non-convex low-rank factorization form, enhancing computational speed while maintaining equivalence with SDP. It is based on the low rank parametrization for nuclear norm minimization [22], and has also been applied in low rank SDP [33] and matrix factorization [34].

PowerFactorization [35] also uses an alternated least-squares scheme to solve for the motion and shape matrices, which has also been extended to solve rank constrained linear matrix equation [36]. Our proposed alternating direction continuation method is similar to PowerFactorization [35] apparently in the sense of alternation. The main differences are: 1) PowerFactorization originates from solving SVD with missing data through alternative least squares, while our alternating direction continuation method originates from low rank parametrization and augmented Lagrangian multipliers method; 2) In our method, continuation step is introduced to speed up the implementation; 3) Balancing of the projective depth is enforced during each iteration to rule out trivial solutions.

### 5 Extensions

In the above sections, we have formulated and solved the problem of element-wise factorization under complete measurements case. In this section, we extend the
proposed framework to deal with missing data, outliers and degenerate case.

5.1 Dealing with Missing Data

In most real-world SfM applications, missing data are inevitable, due to e.g. self-occlusion, points behind cameras (i.e., cheirality) etc. Missing data lead to an incomplete measurement matrix, however simple SVD cannot directly perform on an incomplete matrix. This constitutes a major drawback of factorization-based methods.

For affine (camera) factorization, many missing-data handling ideas have been proposed as Buchanan and Fitzgibbon summarized in [37]. Angst, Zach and Pollefeys [25] proposed generalized trace norm and showed how to incorporate the prior on the camera motion and structure. Del Bue et al. [38] formulated the problem of bilinear factorization and solved the problem of SfM under affine camera model. Olsson and Oskarsson [39] proposed to deal with missing data using matrix completion which is similar to our work but restricted to affine camera model.

Unfortunately, relatively less work was reported for projective factorization with missing data. Most existing works either rely on alternation [6], [40], or assume the depths are pre-computed by other means (reducing to affine case) [41], [42]. Some methods are heuristic or ad hoc, and have no longer the factorization flavor [43].

One of the reasons is probably the complexity of the problem: as we have seen, before introducing missing data, the unknown projective depths have already made the factorization framework very hard.

In this subsection, we extend our element-wise factorization formulation for complete measurements to incomplete measurements case.

Under the circumstance of projective factorization with missing data, we are given an incomplete measurement matrix \( M = [m_{ij}] \) with missing data, where the missing elements are completed with 0. To index the missing pattern and elements, we define a \( 0 - 1 \) mask matrix \( \Omega \) as

\[
\Omega = [\omega_{ij}], \quad \omega_{ij} = \begin{cases} 1 & \text{if } m_{ij} \text{ is available}, \\ 0 & \text{if } m_{ij} \text{ is missing}. \end{cases}
\]  

(27)

With these notations, the projective imaging process with missing data can be compactly written as: \( \Lambda \odot M = \Omega \odot \tilde{W} \). Now our task becomes:

- Given an incomplete measurement matrix \( M \) and the corresponding mask matrix \( \Omega \), find a completed low-rank matrix \( \tilde{W} \) and projective depth matrix \( \Lambda \) such that \( \Lambda \odot M = \Omega \odot \tilde{W} \).

For those missing positions, we do not need to estimate the corresponding depths, so we set \( \lambda_{ij} = 1 \) whenever \( \omega_{ij} = 0 \). This setting guarantees the column-sum and row-sum constraints on the available measurements and all the projective depths possess appropriate scales.

Applying the nuclear norm heuristics, our SDP formulation for element-wise factorization with missing data is achieved by slightly modifying (10):

\[
\begin{align*}
\min_{\tilde{W}} & \quad \frac{1}{2} (\text{tr}(X) + \text{tr}(Y)) \\
\text{s.t.} & \quad \left( \frac{X}{Y} \right)_{\Omega} \preceq 0,
\end{align*}
\]  

(28)

Once this SDP converges, the resultant \( \tilde{W} \) is a completed \( 3m \times n \) full matrix with no entries missing. Moreover, we can even read out a completed projective depth matrix just as the sub-matrix of \( \tilde{W} \) formed by every the third rows of \( \tilde{W} \).

Our alternating direction continuation based algorithm can be extended to missing data, detailed analysis and algorithm are presented in supplemental material.

5.2 Dealing with Outliers

Another recurring practical issue in real world SfM applications is the outlier problem. Different from the missing data case, for the outlier case, we know that some entries of the measurement matrix \( M \) are contaminated by gross errors (i.e., wrong matches), but we do not know where they are.

For conventional projective factorization methods, there is no easy and unified way to deal with outliers. Most published works are based on some pre-processing such as using RANSAC [11], Ke and Kanade [44] formulated matrix factorization as an \( L_1 \) norm minimization problem. Eriksson and Hengel [45] represented a generalization of the Wiberg algorithm calculating the low-rank factorization of a matrix in the presence of missing data. However, these methods are designed for general low-rank matrix approximation problem and work for affine camera model naturally.

Here, we show how our element-wise factorization formulation can handle outlier problem nicely and uniformly. Under projective factorization framework, the input measurement matrix \( M \) is not of the form of low-rank + outliers but the rescaled measurement matrix \( \tilde{W} \) is of the form of low-rank + outliers. The main difficulty is that we have to estimate the projective depths that transform the input measurement matrix to low-rank + outliers which can be recast as low rank and sparsity decomposition problem [46], [47].

First, we give basic assumptions on outliers handling: 1) Outliers do not violate the cheirality constraint, which means even the outlier positions have positive projective depths; 2) Outliers are sparse. As pre-processing work such as RANSAC [11] or outlier removal has been applied, there is only small proportion of outliers in the measurement matrix.

In this problem, we aim to recover the low rank rescaled measurement matrix \( \tilde{W} \) and sparse outlier pattern \( E \) given partial observable measurement matrix \( M \), where the missing pattern is denoted as \( \Omega \). The basic
imaging equation can be compactly written as $\Omega \odot (W + E) = \Lambda \odot M$. Our target is to find projective depth matrix $\Lambda$, rescaled measurement matrix $W$ and outlier pattern $E$ such that $W$ possesses minimal rank and $E$ is as sparse as possible. To quantify the sparseness of $E$, we use its $L_1$-norm $\|X\|_1 = \sum_{i,j} |X_{ij}|$ as a relaxation of its $L_0$-norm$^4$. Combine the nuclear-norm heuristics for $W$ and $L_1$ norm heuristics for $E$, the objective function is set as

$$\min_{W,E,\Lambda} \|W\|_* + \eta \|E\|_1,$$

where $\eta$ is a trade-off parameter.

By introducing auxiliary matrix $Z \in \mathbb{R}^{3m \times n}$, $\|E\|_1$ is equivalent to $tr(Z \cdot 1_{n \times 3m})$, $-Z_{ij} \leq E_{ij} \leq Z_{ij}$, $v(i,j)$ as in [46]. Taking all the constraints into consideration, the final formulation for element-wise factorization with outliers becomes:

$$\min_{Z \in \mathbb{R}^{3m \times n}} \frac{1}{2} \left( \|X \cdot \Omega - W\|_F^2 + \eta \|E\|_1 \right),$$

subject to:

$$Z \odot (W + E) = \Lambda \odot M,$$

$$\Lambda^T 1_m = m 1_n, \quad \Lambda 1_n = n 1_m,$$

$$-Z_{ij} \leq E_{ij} \leq Z_{ij}, \quad v(i,j),$$

$$Z_{ij} = 0, i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

$$\lambda_{ij} > 0, \quad \omega_{ij} = 1,$$

$$\lambda_{ij} > 0, \quad \omega_{ij} = 0,$$

where $1_{n \times 3m} \in \mathbb{R}^{n \times 3m}$ is a matrix having 1 as its every entry. The above model is thus an SDP formulation and can be solved using off-the-shelf solvers. Our alternating direction continuation based algorithm can be extended to handle outliers, detailed analysis and algorithm are presented in supplemental material.

5.3 Degenerate Case

Since we have changed the goal from finding a rank $\leq 4$ matrix to rank minimization, our method is applicable to other cases as well. For example, if all the scene points lie on a plane or the camera undergoes some special motion (such as planar motion), this constitutes a "degenerate" case causing rank $< 4$, which may hinder conventional factorization methods. However, our method is immune to these situations, as confirmed by our experiments.

6 Experimental Results

In this section, we provide extensive experimental results on both synthetic data and real images for complete measurements case, as well as incomplete measurements case, outlier case and degenerate case.

Reprojection error in the image and relative projective depth error (if the ground-truth depths are known) are used to evaluate the performance. Relative depth error is defined as $\varepsilon = \frac{\|A - \hat{A}\|_F}{\|A\|_F}$, where $\Lambda$ is the ground truth projective depth matrix after balancing and $\hat{\Lambda}$ is the recovered projective depth matrix after balancing.

Once we have upgraded the projective reconstruction to Euclidean reconstruction (Euclidean upgrading is discussed in supplemental material), we can evaluate the performance of our algorithms in the aspect of 3D reconstruction. Given the ground truth 3D structure $X$, the recovered structure $\hat{X}$ after registration, the 3D reconstruction error is defined as $\varepsilon = \frac{\|\hat{X} - X\|_F}{\|X\|_F}$. The rotation estimation error (angle) is defined as $\theta_i = \cos(\frac{\text{trace}(R_i^T \hat{R}_i)}{2}, where $R_i$ is the ground truth rotation and $\hat{R}_i$ is the estimated rotation (after coordinate registration).

All the experiments were conducted with our alternating direction continuation implementation except for outliers handling using Matlab R2012a without any optimization on a machine with Core i7 3.4 GHz processor and 8GB of memory under Windows 7 operating system. The image coordinate has always been normalized as suggested by [48].

6.1 Synthetic Experiments

6.1.1 Exact Recovery

As we have relaxed the formulation from low-rank fitting problem to nuclear norm minimization, there is one question left: to what extent can our formulation and algorithms recover the structure and motion exactly? Theoretically, there is a relaxation gap as has been analyzed in the above sections. To cross this gap, we project the result to low rank space. In this part, we conduct extensive experiments on synthetic data and give statistical results.

In these synthetic experiments, we randomly generated 100 points within a cube of $[-30, 30]^3$ in space and 30 perspective cameras were simulated. The image size is set as $800 \times 800$. The camera parameters are set as follows: the focal lengths are set randomly between 900 and 1100, the principal point is set at the image center, and the skew is zero.

To measure the variation of projective depths, we define the depth variation as $\tau = \max_{ij} (\lambda_{ij})/\min_{ij} (\lambda_{ij})$, i.e. the ratio between the maximal depth and the minimal depth after balancing.

Statistical results on synthetic images are counted from more than 1700 random experiments as demonstrated in Fig. 4. We observe that no matter what level of the depth variation, our algorithm can always output results with reprojection error less than $3.5 \times 10^{-6}$ pixels while most of the reprojection errors are less than $1 \times 10^{-8}$ pixels, which can be determined as almost “exact recovery”.

6.1.2 Noise Sensitivity

In this part, we provide experimental validation for the proposed algorithms in the aspect of reprojection error.

---

4. We use $\|E\|_0 = \|\Lambda \odot E\|_0$, since $\lambda_{ij} > 0$. 
and 3D reconstruction under various levels of Gaussian noise. To obtain dataset with ground truth, we use the Notre Dame dataset [14]. This set is a community photo collection consisting of 595 images and 277887 3D points and has been reconstructed and bundle-adjusted, resulting in estimates of all the camera matrices, which we take to represent ground truth. The real image measurements are highly sparse (less than 1%), we project all the points to all the cameras and choose the image measurements with positive depths. We utilize 50 cameras and 400 points, which is a middle size problem however SDP solvers such as SDPT3 and SeDuMi can not solve. The focal lengths vary from 600 pixels to 79820 pixels. Gaussian noise with levels ranging from 0 pixels to 5.0 pixels is added to the measurements. For each noise level, 50 trails are averaged to obtain statistical results.

Fig. 5(a) shows one illustrative example of ground truth 3D structure and recovered structure after registration. Fig. 5(b) and Fig. 5(c) demonstrate the reprojection error and relative depth estimation error under various Gaussian noise levels. Fig. 5(d) demonstrates the relative 3D structure error. Fig. 5(e) demonstrates the rotation estimation error in degrees and Fig. 5(f) demonstrates the relative error in focal length estimation. From these figures, we observe that for noiseless data, our algorithm recovers the camera motion and structure exactly and all the evaluation metrics increase with the increase of Gaussian noise level.

### 6.1.3 Large Depth Variation

As analyzed in the previous sections, for configurations near to affine camera model ($\lambda_{ij} = 1$), iterative algorithms tend to succeed. However for configurations far away from this setting, these algorithms tend to fail in converging or converge to trivial solutions. In this part, we explore this aspect. We test on two cases, the first one is that all the depth variations are within $[1, 5]$ i.e small variation, the other is that all the depth variations are within $[5, 30]$ i.e. large variation.

We compare our algorithm with state-of-the-art iterative methods as CIESTA [9], SIESTA with $L_2$ norm balancing, SIESTA with $L_1$ norm balancing and Column Space method [40]. To compare all these results, we use the form of reprojection error histogram, which gives the distribution of reprojection error and shows statistical property for each algorithm.

Statistical results on more than 1700 randomly configurations are illustrated in Fig. 6. Fig. 6(a) and Fig. 6(b) illustrate depth variation distributions for small depth variation and large depth variation respectively. Fig. 6(c) to Fig. 6(f) demonstrate the results for our proposed method, CIESTA, column space method, SIESTA with $L_1$ balancing and SIESTA with $L_2$ balancing under small and large depth variations respectively. As we expect, our method outperforms all state-of-the-art iterative methods by a significant margin. This can be explained that our method is a global solution and does not depend on initialization. However all other algorithms highly depend on initialization, where affine camera model is widely used as initialization which is not the case for large depth variation.

### 6.1.4 Synthetic Images: Missing Data

To evaluate the performance of our algorithm on incomplete measurements, we generated synthetic data sets as before followed by removing various proportions of 2D image measurements to simulate missing data. On the $50 \times 400$ Notre Dame synthetic dataset, we randomly generated the missing pattern according to uniform distribution, where the missing ratios vary from 10% to 70%. Gaussian noise is added to the measurements. For each missing ratio and Gaussian noise level, 50 trails are repeated.

Statistical results for various noise levels under different missing ratios are demonstrated in Fig. 7. From the figure, we can reach the conclusion that for noiseless measurements, our algorithm can always recover the structure and motion perfectly no matter what level
the missing ratio is. For noisy measurements, the performances (reprojection error and structure estimation error) decrease with the increase of noise level and missing ratio.

Fig. 7. Performance evaluation under various ratios of missing data. (a) Reprojection error VS Missing ratio; (b) Relative 3D structure error VS Missing ratio.

6.1.5 Synthetic Images: Outlier Handling
To illustrate the performance of our method for projective factorization with outliers, we generated the following illustrative example first. The configuration is 15 cameras observing 25 points, leading to measurement matrix of size $30 \times 25$. The outlier pattern, i.e., sparse component $E$ is generated by choosing a support set of size $k$ uniformly at random, $E$ is a matrix with independent Bernoulli $\pm 1$ entries while the ground truth measurements are normalized coordinates with zero mean and standard variation $\sqrt{2}$. In the experiments, we set the outlier ratio as 5% and 10%, which is a realistic percentage of outliers, assuming that a method such as RANSAC [11] has been applied to remove most outliers.

We evaluate the performance of outlier handling in the view of outlier detection. Precision and recall are used to qualify the performance, which are defined as follows:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad \text{Precision} = \frac{TP}{TP + FP},$$

where $TP$ is the true positive, $FP$ is the false positive and $FN$ is the false negative.

Illustrative example is shown in Fig. 8, where the outlier ratio is 10% while 10% of the measurements are missing. Our method outputs result with $\text{Recall} = 0.9473$, $\text{Precision} = 0.9643$, the relative depth error is 6.00% while the relative error in low rank component recovery is 7.68%. Our algorithm not only detects the outliers.
but also completes the measurements and recovers the projective depths.

![Image](image_url)

Fig. 8. Illustrative example for outliers and missing data handling, where the outlier ratio is 10% and the missing ratio is 10%. (a) Input incomplete measurement matrix, (b) Recovered measurement matrix, (c) Ground truth outlier pattern, (d) Recovered outlier pattern.

We conducted experiments under various outlier ratios and missing data ratios up to 30% to obtain statistical result. For each outlier ratio and missing data ratio combination, we repeat experiments 50 times. Statistical results of outliers and missing data handling simultaneously are demonstrated in Fig. 9.

![Image](image_url)

Fig. 9. Statistical result for outliers and missing data handling. (a) Outlier detection precision, (b) Outlier detection recall, (c) Projective depth recovery error, (d) Measurement recovery error.

6.2 Real Images Experiments

6.2.1 Real Images: Complete Data

In this part, we evaluate the proposed methods on real images, where the real images used for complete data are shown in Fig. 10. Reprojection errors for these images are illustrated in Tab. 1. Our method obtains comparable results if not superior.

We also implemented Euclidean upgrading for the Chair sequence. Orthogonal relationships in the original scene are measured on the reconstructed 3D scene as shown in Fig. 11. The average measured angle is 86.9832° for the 9 right angles in the scene.

![Image](image_url)

Fig. 10. Real image sequences. (a) Teabox [49], (b) Corridor, (c) Dinosaur.

![Image](image_url)

Fig. 11. Real image chair sequence. (a) One frame of the sequence, (b) Euclidean upgrading result of the Chair sequence. The average value for nine right angles is 86.9832°.

6.2.2 Real Images: Missing Data

We tested our algorithm for projective factorization with missing data on the dataset shown in Fig. 10(b) and 10(c). The Dinosaur sequence Fig. 10(c) is arguably the most popular dataset used to evaluate rigid SfM algorithms and is conventionally used as an example of affine factorization. Here we however solve it as a projective factorization with missing data problem. Our experiments considered the 36 × 319 measurement case. Experimental results are demonstrated in Tab. 2, where we gave the results of our proposed method and the results after bundle adjustment (denoted as “+BA”).

3D reconstruction results on the Dinosaur and Corridor sequences are demonstrated in Fig. 12.

![Image](image_url)

Fig. 12. 3D reconstruction results. (a) Dinosaur, (b) Corridor.

6.2.3 Real Images: Degenerate Scene

Experimental results on degenerate scene (a calibration pattern [49]) are illustrated in Fig. 13. Fig. 13(a) shows one illustrative image of the calibration pattern and

5. Available at [http://www.robots.ox.ac.uk/~vgg/](http://www.robots.ox.ac.uk/~vgg/).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Corridor(11 × 104)</td>
<td>0.4328</td>
<td>0.4296</td>
<td>0.4501</td>
<td>0.4327</td>
<td>0.3763</td>
</tr>
<tr>
<td>Tebox(2 × 10)</td>
<td>4.482e-04</td>
<td>4.4872e-04</td>
<td>5.126e-04</td>
<td>4.482e-4</td>
<td>3.8071e-4</td>
</tr>
<tr>
<td>Chair(6 × 9)</td>
<td>1.3584</td>
<td>1.3723</td>
<td>1.4759</td>
<td>1.3568</td>
<td>1.2545</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Performance evaluation for real image complete data (Reprojection error in pixels)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Proposed method</th>
<th>+BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dinosaur(36 × 319, 76.92% Missing)</td>
<td>1.7510</td>
<td>1.5893</td>
</tr>
<tr>
<td>Corridor(11 × 737, 50.23% Missing)</td>
<td>0.4217</td>
<td>0.4003</td>
</tr>
</tbody>
</table>

Fig. 13(b) illustrates the corresponding recovered balanced projective depth. Fig. 13(c) shows the performance comparison between our method, SIESTA, CIESTA and Column space method under various number of points. Finally, Fig. 13(d) gives the Euclidean reconstruction.

7 Conclusions

In this paper, we have proposed a new method for projective multi-view structure and motion factorization, using a novel element-wise factorization formulation. We represent the problem as an SDP and solve it efficiently and (approximately) globally. Furthermore, we adopt fixed point continuation based and alternating direction continuation based algorithms to address the scalability issue, which prepare our method well for practical large scale applications. The results obtained by our new method are either comparable with or superior to the conventional iteration-based methods—when these methods also work. In many other practical cases when these conventional methods do not work well, our method still produces satisfactory results.

Acknowledgments

We would like to thank Prof. Richard Hartley for his invaluable support and discussions. We would like to thank the anonymous reviewers of TPAMI and ECCV 2010 for constructive comments that improves the manuscript. This work was supported in part by NSFC (60736007, 61171154), NSF of Shaanxi Province (2010JZ011), Australian Research Council (ARC) discovery grant and ARC linkage grant.

References


Yuchao Dai is currently a Postdoctoral Fellow with the Research School of Computer Science at the Australian National University, Canberra, Australia. He received the B.E. degree, M.E degree and Ph.D. degree all in signal and information processing from Northwestern Polytechnical University, Xian, China, in 2005, 2008 and 2012, respectively. He was a visiting student at ANU from Oct. 2008 to Oct. 2009 with the support of the China Scholarship Council. His research interests include structure from motion, multi-view geometry, compressive sensing and optimization. He won the best paper award in CVPR 2012.

Hongdong Li (M’01) is a Fellow with the Australian National University. He was also seconded to NICTA (National ICT Australia) as a Senior Research Scientist. His research interests include geometric computer vision, structure from motion, image restoration and bionic eyes, and optimization methods. He received MSc. and Ph.D. degree in Electronics Engineering from Zhejiang University in 2000. After teaching in the same university for three years he joined the ANU/NICTA teaching and doing researches in Computer Vision. He constantly serves on the Programme Committees (or as a reviewer) for all the major international Computer Vision conferences and journals, such as ICCV, CVPR, ECCV, PAMI and IJCV. He is an Area Chair for ICCV’13. He was a recepient for the CVPR’12 Best Paper Award.

Mingyi He received the B.Sc. and M.Sc. degrees in electronic engineering from Northwestern Polytechnical University, Xian, China, in 1982 and 1985, respectively, and the Ph.D. degree in signal and information processing from Xidian University, Xian, in 1994. Professor He is the Founder and Director of the Shaanxi Key Laboratory of Information Acquisition and Processing, a professor in School of Electronics and Information, and the Director and Chief Scientist of the Center for Earth Observation Research, Northwestern Polytechnical University. He won 9 scientific prizes from China and the best paper award in CVPR 2012. He has published over 280 papers and 5 books and has made contributions to hyperspectral remote-sensing data processing, visual information processing, neural networks, and 3D technique.