Consensus of Non-Rigid Reconstructions

Minsik Lee[†*] Jungchan Cho[‡*] Songhwai Oh[‡]
Division of EE, Hanyang University, Korea[†]
Department of ECE, ASRI, Seoul National University, Korea[‡]
mleepaper@hanyang.ac.kr {cjc83, songhwai}@snu.ac.kr

Abstract

Recently, there have been many progresses for the problem of non-rigid structure reconstruction based on 2D trajectories, but it is still challenging to deal with complex deformations or restricted view ranges. Promising alternatives are the piecewise reconstruction approaches, which divide trajectories into several local parts and stitch their individual reconstructions to produce an entire 3D structure. These methods show the state-of-the-art performance, however, most of them are specialized for relatively smooth surfaces and some are quite complicated. Meanwhile, it has been reported numerously in the field of pattern recognition that obtaining consensus from many weak hypotheses can give a strong, powerful result. Inspired by these reports, in this paper, we push the concept of part-based reconstruction to the limit: Instead of considering the parts as explicitly-divided local patches, we draw a large number of small random trajectory sets. From their individual reconstructions, we pull out a statistic of each 3D point to retrieve a strong reconstruction, of which the procedure can be expressed as a sparse l_1-norm minimization problem. In order to resolve the reflection ambiguity between weak (and possibly bad) reconstructions, we propose a novel optimization framework which only involves a single eigenvalue decomposition. The proposed method can be applied to any type of data and outperforms the existing methods for the benchmark sequences, even though it is composed of a few, simple steps. Furthermore, it is easily parallelizable, which is another advantage.

1. Introduction

There have been many remarkable advances in non-rigid structure from motion (NRSfM), a problem to reconstruct 3D structures from a set of 2D trajectories, during the last decade. The ambiguity in global rotations, pointed out by Xiao *et al*. [30], was proven to be resolvable with the rank-3 constraint [4]. An algorithm to find an approximate rank-3 solution followed in [10], combined with an efficient structure reconstruction algorithm which is concisely expressed by a convex rank-minimization problem. On the other hand, there have been many attempts to incorporate the temporal dependence prior in the reconstruction process [5, 17], one of which [17] already showed good performances for the popular benchmark data sets. The difficulty of deciding the number of shape bases has been resolved by the Procrustean distributions [20, 21], which are shape distributions without global rigid transformations, and online reconstruction became possible by processing data sequentially based on the continuum mechanics [1, 2]. Finally, most recent proposals on dense structure recovery [13, 15] have made the practicality of NRSfM convincing.

Even with all these progresses, however, non-rigid structure recovery is still a difficult problem when there are complex deformations or when the range of viewpoint is restricted. Existing methods are usually based on linear or Gaussian shape models, which are not strong enough to handle complicated shape changes. The situation is worse for the restricted view range cases, *i.e*., the reconstruction itself becomes completely impossible for this case because these methods rely on global rigid transformations. Regarding this, some of the state-of-the-art results achieved by the existing methods for benchmark data sets are mainly attributed by the artificially added rotations. For example, the benchmark sets proposed in [4] contain unrealistic artificial rotations, of which the total amount is about 1500 – 5500 degrees depending on the data. This problem can be confirmed in the experiments of Section 4, where the performances of the existing methods degrade severely for realistic view ranges.

A possible alternative to these problems is the piecewise reconstruction approach [11, 12, 24]. This approach tries to divide trajectories into several parts, which are local patches of the entire structure. Fayad *et al*. [11] used the quadratic deformation model to reconstruct overlapping patches and connected them based on a smoothness assumption, and its heuristic patch assignment was later replaced by a graph-

*Authors contributed equally.

cut formulation in [24]. In [12], a rigid segment assumption was used instead of the smooth surface assumption to handle articulated data. However, most of these assume a smooth surface for the object of interest, which restricts the range of applications. Although [12] does not assume a smooth surface, but the rigid segment assumption also restricts the possible uses. A more important drawback of these methods is that most of them have to iterate between part assignment and part reconstruction, which can be a heavy procedure. This can also be detrimental for parallelization, even though it could be a strong point for part-based reconstruction.

Nevertheless, the part-based approach can still be promising for the aforementioned issues due to the following reasons: Obviously, a part may have simpler deformations than the global shape, as argued in [11, 12, 24]. The second reason, which was not seriously taken by other work but we claim to be more important, is that a part may have more rigid changes than the global one, because many of deformations can be seen as rigid motions of local parts. A most obvious (and extreme) example is an articulated object, say, a human body. Note that the movements of arms and legs will be considered as deformations in a holistic view, but if we divide the body into a part for each bone, then these movements can now be seen as the motions of rigid objects. Even for an object that is not articulated, we can expect to have similar effects.

To maximize these expectations, in this paper, we take a different view on the part-based reconstruction strategy. There have been many successful examples in pattern recognition that the consensus of many weak hypotheses gives a strong result, such as boosting techniques [28], random forests [9], and Hough voting [14]. In these approaches, an individual hypothesis does not have much discriminative power and often gives bad answers to the problem, although their combination gives a high recognition rate. In this spirit, we consider the part-based reconstruction problem as deriving consensus from numerous 'weak' reconstructions. A large amount of random trajectory sets, each of which contains only a few trajectories, are sampled from the data, and their 3D structures are reconstructed individually by a weak reconstructor. In this process, we do not care about how correctly the parts are assigned or how accurate their reconstructions are, but only how efficiently they can be reconstructed. After that, we derive a statistic of each trajectory from the weak results to generate a strong reconstruction. Since there can be some bad reconstructions or outliers in the weak reconstructions, median is chosen for the statistic, which can be elegantly expressed by an l_1-norm minimization problem. Since there are reflection ambiguities between different weak (and possibly bad) reconstructions, we resolve them before obtaining the consensus, by a novel optimization problem which is essentially a Rayleigh quotient [19].

The difference between the proposed method and the other part-based schemes is apparent in that the others can be viewed as stitching problems, and they try to find good part assignments by using many assumptions, *e.g.*, the minimum description length principle or strictly connected, overlapping segments. On the other hand, the proposed method can be considered as a bagging or bootstrap aggregation [8] approach because it generates a large number of small, random reconstructions and processes them afterwards, which is much simpler and more flexible. The proposed method is composed of a few, very simple steps and yet, it shows the state-of-the-art performance for the benchmark data sets. An advantage of the proposed algorithm is that the types of data and weak reconstructor are not restricted. Furthermore, it is easy to parallelize because the proposed method mainly consists of a large number of small reconstructions, which are performed only once for each trajectory group and are not repeated again for the entire process. A parallelized version can be actually very beneficial, because the weak reconstruction step is the most time-consuming part of the proposed method.

The remainder of this paper is organized as follows: We present random sampling and weak reconstruction procedures in Section 2. The details of obtaining consensus is explained in Section 3. The experimental results follow in Section 4 and finally, we conclude the paper in Section 5.

2. Weak reconstructions

In this section, we explain the sampling and weak reconstruction procedures for trajectory groups. Before that, we define some basic notations: $\mathbf{D}_i \in \mathbb{R}^{2\times p}$, $1 \le i \le f$, is the input 2D points of the ith frame observed by an orthographic camera. The coordinates of $\mathbf{D}_i$ are assumed to be translated so that the centroid is at the origin. $G = \{g_k\}$ is a set of random trajectory groups, and the jth trajectory, $1 \le j \le p$, is included in the kth group if $j \in g_k$. $\mathbf{z}_{ik}$ is a p-dimensional row vector, which is the reconstructed z coordinates of the ith frame for the points in the kth group. We assume that the centroid of $\mathbf{z}_{ik}$ for the points in g_k is at the origin and the elements of all the other points, which are not included in g_k, are zero, without loss of generality. For a matrix $\mathbf{A}$ with p columns, $\mathbf{A}|_{g_k}$ denotes a submatrix of $\mathbf{A}$ that is composed of the columns indexed by g_k, throughout this paper. Similarly, the jth column vector of $\mathbf{A}$ will be denoted as $\mathbf{A}|_j$ and the columns from j to j' as $\mathbf{A}|_{j:j'}$. $\|\cdot\|$ on matrices indicates the Frobenius norm in this paper.

2.1. Sampling trajectory sets

The purposes of part-based reconstruction are to reduce the complexity of deformations and to increase the range of view angles. To meet these purposes, it is better to sample a trajectory group so that the points are close to each other

and share similar movements. A natural way is to make the trajectories in a group follow a Gaussian distribution:

$$P(\mathcal{V}_k) \propto \prod_{j \in g_k} \exp\left(-\frac{\lambda}{2} \sum_i \|\mathbf{D}_i|_j - \mathbf{m}_i\|^2\right), \quad \mathbf{m}_i \triangleq \frac{1}{n_g} \sum_{j \in g_k} \mathbf{D}_i|_j, \tag{1}$$

where $\mathcal{V}_k = \{\mathbf{D}_i|g_k | 0 \le i \le f\}$, n_g is the number of trajectories in a single group[1] and λ is a pre-defined parameter. However, evaluating this probability for every possible combination is not realistic. Instead, we can manipulate $\log(P(\mathcal{V}_k))$ as

$$-\frac{\lambda}{2n_g} \sum_{j,j' \in g_k} \sum_i \|\mathbf{D}_i|_j - \mathbf{D}_i|_{j'}\|^2 \tag{2}$$

up to a normalization constant. This makes us easy to sample a trajectory group: If we have already chosen some trajectories for g_k and want to pick another one, then we can perform a weighted sampling based on the weights

$$w_j = \begin{cases} \exp\left(-\frac{\lambda}{2n_g} \sum_{j' \in g_k} l_{jj'}\right) & \text{if } j \notin g_k, \\ 0 & \text{otherwise,} \end{cases} \quad l_{jj'} \triangleq \sum_i \|\mathbf{D}_i|_j - \mathbf{D}_i|_{j'}\|^2. \tag{3}$$

The only exception is the first trajectory, which is selected based on a uniform distribution. The pairwise distance $l_{jj'}$ can be pre-computed to speed up the process.

This procedure is much simpler than the part-assignment schemes in other part-based reconstruction methods. It does not require any constraints but simply selects similar trajectories for a group. Of course, one can put some prior knowledge on the parts to improve the procedure but, nonetheless, this simple version was good enough to achieve the state-of-the-art performance for the benchmark sets. Note that this procedure compares trajectories but it has nothing to do with the trajectory-basis methods [5], because they are not enforced to be represented by low-frequency bases. Thus, shuffling the order of frames will not affect the outcome.

A trajectory can be included in multiple groups, and for the purpose of the proposed method, we must have a lot of groups that share the same point. Hence, we repeat the sampling process until all the points are included in at least m_g groups. In practice, we found out that normalizing the 2D coordinates beforehand for each frame so that their covariance becomes isotropic, *i.e.*, an identity matrix, can improve the performance of the final reconstruction.

[1]We assume that all groups have the same number of samples.

2.2. Design of weak reconstructor

Like any other part-based methods, the proposed method can use most of the reconstruction schemes as a weak reconstructor. The criteria for choosing the weak reconstructor in this work are two-fold: (i) It has to be simple and concise, although (ii) it is based on well-established theories. The block matrix method (BMM) proposed in [10] meets these criteria, hence, we adopt this algorithm as a weak reconstructor, with some modifications for efficiency.

Here, we briefly review BMM and explain our modifications. As for most of the NRSfM schemes, BMM is composed of two steps, *i.e.*, calculating 3D rotations and estimating 3D shapes. The basic system of equations for shape-basis-based NRSfM is expressed as [10, 17, 30]

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_f \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & & & \\ & \mathbf{R}_2 & & \\ & & \ddots & \\ & & & \mathbf{R}_f \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_f \end{bmatrix} = \mathbf{RS} = \mathbf{R}(\mathbf{C} \otimes \mathbf{I})\mathbf{B} \triangleq \mathbf{QB}, \tag{4}$$

where $\mathbf{R}_i \in \mathbb{R}^{2\times 3}$ and $\mathbf{S}_i \in \mathbb{R}^{3\times p}$ are the rotation and the aligned 3D shape, respectively, of the ith frame. Because there are assumed to be only a few ($= K$) shape bases, $\mathbf{S}$ is decomposed into two thin matrices as in the above equation, where $\mathbf{B} \in \mathbb{R}^{3K\times p}$ consists of the shape bases and $\mathbf{C} \in \mathbb{R}^{f\times K}$ the corresponding coefficients. We can utilize the singular value decomposition (SVD) to find a $3K$-rank approximation of $\mathbf{D} \triangleq \mathbf{Q'B'}$ and it is likely that, for some $\mathbf{F} \in \mathbb{R}^{3K\times 3K}$, they satisfy

$$\mathbf{Q'F} \approx \mathbf{Q} = \mathbf{R}(\mathbf{C} \otimes \mathbf{I}), \qquad \mathbf{F}^{-1}\mathbf{B'} \approx \mathbf{B}. \tag{5}$$

From (5), we can find out that $\mathbf{Q'F}|_{1:3} \approx [c_{11}\mathbf{R}_1^T \ \dots \ c_{f1}\mathbf{R}_f^T]^T$ where c_{ij} is the (i,j)th element of $\mathbf{C}$. Each 2×3 block of $\mathbf{Q'F}|_{1:3}$ is nothing but a scaled orthogonal matrix, of which the outer product is given as

$$\mathbf{Q}_i'\mathbf{F}|_{1:3}\mathbf{F}|_{1:3}^T\mathbf{Q}_i'^T = c_{i1}^2\mathbf{I}, \tag{6}$$

where $\mathbf{Q}_i'$ is the ith pair of consecutive rows in $\mathbf{Q'}$. From this, we have two linear equations on the rank-3 positive semidefinite (PSD) matrix $\mathbf{M} \triangleq \mathbf{F}|_{1:3}\mathbf{F}|_{1:3}^T$ for each i, one from the off-diagonal elements and the other by comparing the diagonal elements [30]. $\mathbf{M}$ can be solved by minimizing the squared errors of these linear equations, which can be expressed as $\|\mathbf{A}\,\mathrm{vec}(\mathbf{M})\|^2$ based on an appropriate matrix $\mathbf{A}$ derived from (6) and a vectorization operator. However, this is a non-convex problem due to the rank-3 constraint on $\mathbf{M}$. In [10], to bypass the difficulty of non-convex optimization, the rank of $\mathbf{M}$ is minimized with the equations from (6) as constraints to make it close to a rank-3 matrix, but this method can be quite unstable depending on the data.

Instead, we follow the approach in [17] which directly solves the non-convex problem. The reason we adopt this rotation-calculation scheme is because it does not require the knowledge of the exact K, in that the rotation problem is solved repeatedly with increasing K until there is little improvement on the orthogonality cost. In [17], a general-purpose optimization tool is applied to a formulation based on $\mathbf{F}|_{1:3}$, which can be slow. Instead, in this paper, we provide a proximal gradient [6] formulation for this problem[2]: Based on the cost $\|\mathbf{A}\,\mathrm{vec}(\mathbf{M})\|^2$, which is in the form of linear least squares, the following subproblem is solved in each iteration:

$$\min_{\mathbf{M}} \quad \|\mathbf{M}-\mathbf{M}'\|^2 \qquad \text{s.t.} \quad \mathrm{rank}(\mathbf{M})=3, \tag{7}$$

where $\mathbf{M}'$ is given as

$$\mathrm{vec}(\mathbf{M}') = \mathrm{vec}(\mathbf{M}_0) - \mathbf{A}^T\mathbf{A}\,\mathrm{vec}(\mathbf{M}_0)/L_{\mathbf{A}}. \tag{8}$$

Here, $\mathbf{M}_0$ is the previous estimate of $\mathbf{M}$ and $L_{\mathbf{A}}$ is the Lipschitz constant, *i.e.*, the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$. One thing to note is that, since the original cost function has a trivial solution, we add another constraint on $\mathbf{M}$ about its scale. The rank-3 solution to the above problem can be easily found by an eigenvalue decomposition. This eigenvalue decomposition is not heavy and can be performed in a short instance, because in this work, the size of $\mathbf{M}$ is no more than $3n_g$ and the size of each group is small. After finding $\mathbf{M}$, $\mathbf{F}|_{1:3}$ is recovered to find the rotations.

The rotation matrix $\mathbf{R}_i$ we have been dealing with so far is a 2×3 Stiefel matrix. The third axis $\acute{\mathbf{R}}_i \in \mathbb{R}^{1\times 3}$ can be found by the cross product of the first and second rows of $\mathbf{R}_i$, and we denote the full 3D rotation as $\breve{\mathbf{R}}_i = \begin{bmatrix}\mathbf{R}_i^T & \acute{\mathbf{R}}_i^T\end{bmatrix}^T$. Based on the estimated rotations, BMM retrieves the (aligned) 3D shapes by solving the following convex rank-minimization problem [10]:

$$\min_{\mathbf{S}} \quad \left\|\mathbf{S}^{\#}\right\|_* \qquad \text{s.t.} \quad \mathbf{D}=\mathbf{R}\mathbf{S}. \tag{9}$$

Here, $\|\cdot\|_*$ denotes the nuclear norm and $\mathbf{S}^{\#}$ is the rearranged version of $\mathbf{S}$, *i.e.*, $\mathbf{S}^{\#} = \begin{bmatrix}\mathrm{vec}(\mathbf{S}_1) & \dots & \mathrm{vec}(\mathbf{S}_f)\end{bmatrix}$. In this work, this process is applied to a small group of trajectories, of which the deformation is possibly simple and small, thus the 'mean' 3D structure will be the dominant component in $\mathbf{S}^{\#}$. Hence, we modify (9) as

$$\min_{\mathbf{S}} \quad \left\|\mathbf{S}^{\#}\mathbf{P}\right\|_* \qquad \text{s.t.} \quad \mathbf{D}=\mathbf{R}\mathbf{S}. \tag{10}$$

Here, $\mathbf{P}$ is introduced to remove the mean component from $\mathbf{S}^{\#}$. In this paper, $\mathbf{P}$ is defined to be an orthogonal projection with $\mathbf{1}$, a vector of ones, being its only null space. For example, in the above problem, $\mathbf{P}$ is $\left(\mathbf{I}-\mathbf{1}\mathbf{1}^T/f\right)$.

[2]The convergence of a nonconvex proximal algorithm has been extensively studied [7]: In summary, the projection onto a fixed-rank matrix can be casted into a semi-algebraic cost function, satisfying the KL property, and so our proximal algorithm always converges to a critical point.

A continuation method has been used in [10] to solve the above equality-constrained problem, which is very slow. Instead, in this paper, we use the augmented Lagrangian method (ALM) [22] to solve (10). By an introduction of an auxiliary variable $\tilde{\mathbf{S}} = \mathbf{S}^{\#}\mathbf{P}$, we can solve this problem by alternatingly updating $\tilde{\mathbf{S}}$ and $\mathbf{S}$ according to the alternating directional method of multipliers (ADMM) [22], which can be thought as a non-inner-loop version of ALM. Updates of $\tilde{\mathbf{S}}$ and $\mathbf{S}$ can be obtained by performing the singular value thresholding [22] and by solving a linear least squares problem, respectively. Since this problem is a convex optimization, ADMM guarantees a global optimal solution. The size of this problem is also small for our case, thus the SVD operation required for the singular value thresholding does not cost much computation.

We are using this modified version of BMM to reconstruct each trajectory group g_k, hence, $\mathbf{D}_i|_{g_k}$ is actually used instead of $\mathbf{D}_i$ for the weak reconstructor. After finding the solution $\hat{\mathbf{S}}_i$ of the aligned 3D shapes, $\mathbf{z}_{ik}$ is set to $\mathbf{z}_{ik}|_{g_k} = \acute{\mathbf{R}}_i\hat{\mathbf{S}}_i\mathbf{P}$ and all the elements not in g_k to zeros[3].

3. Obtaining consensus

In this section, we explain the procedure of obtaining consensus from weak reconstructions. Some additional notations are used in this section: Let $\bar{\mathbf{z}}_i \in \mathbb{R}^{1\times p}$ be the z-coordinates of the strong reconstruction for the ith frame. $\mathbf{e}_k$ is a p-dimensional row vector of which the element is one if its index is in g_k and zero otherwise. $\mathbf{W}_k \in \mathbb{R}^{p\times p}$ is defined based on $\mathbf{e}_k$ as

$$\mathbf{W}_k = \mathrm{diag}(\mathbf{e}_k) - \frac{1}{n_g}\mathbf{e}_k^T\mathbf{e}_k, \tag{11}$$

where $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix of which the diagonal elements are the elements of a vector $\mathbf{x}$. The role of $\mathbf{W}_k$ is to retrieve the g_k part of a p-dimensional vector with removing the translation component, *i.e.*, $\mathbf{x}'\mathbf{W}_k = \mathbf{x}'|_{g_k}\mathbf{P}$ where $\mathbf{x}'$ is a p-dimensional row vector and $\mathbf{P}$ here is n_g-dimensional. Some relations hold for $\mathbf{e}_k$, $\mathbf{W}_k$, and $\mathbf{z}_{ik}$, *i.e.*,

$$\begin{aligned} &\mathbf{1}^T\mathbf{W}_k = \mathbf{0}^T, \quad && \mathbf{e}_k\mathbf{W}_k = \mathbf{0}^T, \quad && \mathbf{z}_{ik}\mathbf{e}_k^T = 0, \\ &\mathbf{z}_{ik}\mathbf{W}_k = \mathbf{z}_{ik}, && \mathbf{z}_{ik}\mathbf{1} = 0, && \mathbf{W}_k^2 = \mathbf{W}_k, \end{aligned} \tag{12}$$

where $\mathbf{0}$ is a column vector of zeros.

3.1. Resolving reflection ambiguities

Before explaining the procedure of calculating the strong reconstruction, we have to resolve an important issue in the weak reconstructions, which arises due to the nature of orthographic NRSfM [4]: $\breve{\mathbf{R}}_i$ may have some reflection ambiguities for the signs of $\mathbf{R}_i$ and $\acute{\mathbf{R}}_i$, which can later lead to those of the reconstructed z-coordinates.

[3]Note that $\mathbf{P}$ here is p-dimensional.

There can be several ways to resolve this issue, but in this paper, we take the advantage of our previously-mentioned intuition that the aligned 3D shapes have a dominant mean component. If $\mathbf{B}_l$, the lth basis shape, is the dominant component, then the coefficients c_{il} associating with this basis should have the same sign. In practice, we solve the following optimization to resolve the sign differences in $\mathbf{R}_i$:

$$\min_{q_i,\acute{\mathbf{B}}} \sum_i \left\| \acute{\mathbf{B}} - q_i \mathbf{R}_i^T \mathbf{D}_i \right\|^2 \quad \text{s.t. } q_i \in \{-1, 1\}. \tag{13}$$

This problem can be solved alternatingly for $\acute{\mathbf{B}}$ and q_i. $\acute{\mathbf{B}}$ is updated simply to the average of $q_i \mathbf{R}_i^T \mathbf{D}_i$, and q_i is updated to $\operatorname{sign}\left(\left\langle \acute{\mathbf{B}}, \mathbf{R}_i^T \mathbf{D}_i \right\rangle\right)$[4], in each iteration. This procedure converges within a few step in practice, and $\mathbf{R}_i$ can be updated afterwards as $\mathbf{R}_i \leftarrow \hat{q}_i \mathbf{R}_i$ based on the solution $\hat{q}_i$. If the sign differences between $\mathbf{R}_i$'s are resolved, those for $\acute{\mathbf{R}}_i$'s can be easily resolved by enforcing $\breve{\mathbf{R}}_i$ to be in the special orthogonal group, *i.e.*, $\det(\breve{\mathbf{R}}_i) = 1$. In fact, setting $\acute{\mathbf{R}}_i$ to the cross product of the first and second rows of $\mathbf{R}_i$ already makes $\det(\breve{\mathbf{R}}_i) = 1$. Note that this procedure should be performed before the computation of $\mathbf{S}$, because the mean removal term added in (10) assumes that the directions of all $\mathbf{S}_i$'s coincide.

Care should be taken in understanding this procedure, because this only resolves the relative differences between the signs and not the entire reflection ambiguity. Therefore, the entire (weak) reconstruction may have the correct z-coordinates, or all of them might be inverted. This means that the reconstructions of different trajectory groups might have different directions for the z-coordinates, as in the other part-based reconstruction schemes.

A possible approach for resolving these sign differences is to compare the Euclidean distance between two groups and correct the sign of one group for another one by one, as in [11]. However, this is not a good idea for two reasons: (i) This approach gives a greedy solution to an overall non-convex sign problem, and (ii) the weak reconstructions are supposed to be rough solutions and there might be many bad ones or outliers. Therefore, correcting patches one by one might not give a good estimate and even cumulate errors, which corresponds with our empirical experiences. This necessitates a method to handle the sign problem based on global information, rather than local information.

The overall sign problem can be given as follows: Let us define $\mathbf{m}_i$ as a rough estimate of $\bar{\mathbf{z}}_i$, and $\mathbf{M}$ and $\mathbf{Z}_k$ as $f \times p$ matrices of which the ith rows are $\mathbf{m}_i$ and $\mathbf{z}_{ik}$, respectively, then optimizing

$$\min_{\mathbf{M}, r_k} \frac{1}{2} \sum_k \|\mathbf{M}\mathbf{W}_k - r_k \mathbf{Z}_k\|^2 \quad \text{s.t. } r_k \in \{-1, 1\} \tag{14}$$

can give the solution of the relative sign r_k of a group. As mentioned earlier, this is a non-convex problem and there can be bad local optimums.[5] In order to resolve this issue, we can substitute r_k with an estimate that could make the problem easier to solve. If we assume that most of $\mathbf{Z}_k$'s are moderately accurate, then they must be similar to $\mathbf{M}\mathbf{W}_k$'s up to sign changes. Therefore, we choose

$$\frac{\operatorname{tr}(\mathbf{M}\mathbf{W}_k \mathbf{Z}_k^T)}{\|\mathbf{Z}_k\|^2} = \frac{\operatorname{tr}(\mathbf{M}\mathbf{Z}_k^T)}{\|\mathbf{Z}_k\|^2} \approx r_k \tag{15}$$

for the estimate of r_k. Then, (14) is changed to

$$\min_{\mathbf{M}} \quad \frac{1}{2} \sum_k \left\| \mathbf{M}\mathbf{W}_k - \frac{\operatorname{tr}(\mathbf{M}\mathbf{Z}_k^T)}{\|\mathbf{Z}_k\|^2} \mathbf{Z}_k \right\|^2, \tag{16}$$

which can be also expressed as

$$\min_{\mathbf{M}} \quad \operatorname{vec}(\mathbf{M})^T (\mathbf{W} \otimes \mathbf{I} - \mathbf{V}^T \mathbf{V}) \operatorname{vec}(\mathbf{M}), \tag{17}$$

where $\mathbf{W} \triangleq \sum \mathbf{W}_k$ and

$$\mathbf{V} = \begin{bmatrix} \operatorname{vec}(\mathbf{Z}_1)/\|\mathbf{Z}_1\| & \operatorname{vec}(\mathbf{Z}_2)/\|\mathbf{Z}_2\| & \cdots \end{bmatrix}^T. \tag{18}$$

We can verify that the Hessian of (17) is PSD:

Proposition 1. $\mathbf{H} \triangleq \mathbf{W} \otimes \mathbf{I} - \mathbf{V}^T \mathbf{V}$ *is PSD,* i.e., $\mathbf{0} \preceq \mathbf{H} \preceq \bar{m}_g \mathbf{I}$ *where* $\bar{m}_g$ *is the maximum number of groups that a trajectory belongs to.*

The proof can be found in the supplementary material. The solution of $\mathbf{M}$ can be trivial, so we introduce a scale constraint on $\mathbf{M}$ as

$$\min_{\mathbf{M}} \operatorname{vec}(\mathbf{M})^T \mathbf{H} \operatorname{vec}(\mathbf{M}) \quad \text{s.t. } \|\mathbf{M}\|^2 = 1, \tag{19}$$

which is basically a Rayleigh quotient [19].[6] The solution to a Rayleigh quotient can be easily found by an eigenvalue decomposition, which makes the problem very simple. However, there is still an issue here: the size of $\mathbf{H}$ is $pf \times pf$, which can be pretty large, so this might ruin the efficiency of the proposed method.

This issue can be bypassed by finding only a single eigenvector that is necessary. First, we alter the form of (19) to a maximization problem:

$$\max_{\mathbf{M}} \operatorname{vec}(\mathbf{M})^T (\bar{m}_g \mathbf{I} - \mathbf{H}) \operatorname{vec}(\mathbf{M}) \text{ s.t. } \|\mathbf{M}\|^2 = 1. \tag{20}$$

[4] In this paper, the signum function is a right-continuous function.

[5] In fact, this is equivalent to the problem in (13). For (13), solving the non-convex problem directly gives a good solution (and is much more efficient actually), because our dominant-mean assumption fits the real situation very well. However, for different groups, the non-convex formulation did not work well because of the outlying weak reconstructions.

[6] This version also gives a trivial solution, *i.e.*, $\mathbf{M} = \frac{1}{p}\mathbf{1}$, and there should be another constraint $\mathbf{M}\mathbf{1} = \mathbf{0}$. We do not deal with this fact explicitly for the sake of simplicity. Nevertheless, it can be verified that the proposed method yields a solution that satisfies this additional constraint.

Here, $\bar{m}_g\mathbf{I}$ has been added to the cost function to make the Hessian PSD without changing the solution, based on Proposition 1. Since the Hessian is PSD, we can apply a power iteration to find the eigenvector of the largest eigenvalue, which is the only thing we need. Exploiting the fact that $\mathbf{H}$ is composed of sparse and low-rank matrices, A time-efficient procedure for updating $\mathbf{M}$ can be given as

$$\begin{aligned} \mathbf{M}' &\leftarrow \bar{m}_g\mathbf{M} - \mathbf{M}\mathbf{W}, \\ \mathrm{vec}(\mathbf{M}') &\leftarrow \mathrm{vec}(\mathbf{M}') + \mathbf{V}^T(\mathbf{V}\,\mathrm{vec}(\mathbf{M})), \\ \mathbf{M} &\leftarrow \mathbf{M}'/\|\mathbf{M}'\|. \end{aligned} \tag{21}$$

Note that $\mathrm{vec}(\mathbf{M})$ is initialized to the right singular vector of the largest singular value of the fat matrix $\mathbf{V}$ beforehand.

In fact, this is the least time-consuming step except the sampling step, thanks to this efficient formulation. After finding $\mathbf{M}$, the relative sign of each group can be found as $r_k = \mathrm{sign}(\mathrm{tr}(\mathbf{M}\mathbf{Z}_k^T))$, according to (15).

3.2. Finding a strong reconstruction

Now, since we have resolved the reflection ambiguity between different groups, consensus can be obtained from them. Our approach is to derive a statistic from the groups and the choice of the statistic is median, because there can be some bad reconstructions or outliers among the groups. It is well-known that median can be expressed as the solution of a convex problem as

$$\mathrm{med}(\{x_i\}) = \arg\min_x \sum |x - x_i|, \tag{22}$$

which is related to the $l1$-norm. We might try to minimize the $l1$-error between $\mathbf{z}_{ik}$ of each group and $\bar{\mathbf{z}}_i|_{g_k}$ to find the strong reconstruction. However, there is a translation ambiguity in $\mathbf{z}_{ik}$, hence we have to compare them to $\bar{\mathbf{z}}_i\mathbf{W}_k$ instead. Thus, our problem can be described as

$$\min_{\bar{\mathbf{z}}_i} \quad \sum_k \|\mathbf{z}_{ik} - \bar{\mathbf{z}}_i\mathbf{W}_k\|_1, \tag{23}$$

Note that this problem can be solved independently for each i, *i.e.*, we can solve this problem for each frame separately. The above problem can be equivalently express as

$$\min_{\bar{\mathbf{z}}_i, t_{ik}} \quad \sum_k \|(\mathbf{z}_{ik} + t_{ik}\mathbf{1}^T - \bar{\mathbf{z}}_i) \odot \mathbf{e}_k\|_1 \tag{24}$$

with additional variables t_{ik}'s ($\in \mathbb{R}$). Here, $\odot$ denotes the Hadamard product. In reality, we solve the latter formulation because this can reduce the complexity of the optimization, which will be explained later.

The cost function of (24) is the same as

$$\sum_k \|\mathbf{z}_{ik}|_{g_k} + t_{ik}\mathbf{1}^T - \bar{\mathbf{z}}_i|_{g_k}\|_1. \tag{25}$$

Let us assume

$$\begin{aligned} \mathbf{y}_i &\triangleq \begin{bmatrix} \mathbf{z}_{i1}|_{g_1} & \mathbf{z}_{i2}|_{g_2} & \dots \end{bmatrix}^T, \\ \mathbf{t}_i &\triangleq \begin{bmatrix} t_{i1} & t_{i2} & \dots \end{bmatrix}^T, \quad \mathbf{E} \triangleq \begin{bmatrix} \mathbf{E}_1^T & \mathbf{E}_2^T & \dots \end{bmatrix}^T, \end{aligned} \tag{26}$$

where $\mathbf{E}_k$ is a submatrix of $\mathrm{diag}(\mathbf{e}_k)$ that includes all the nonzero rows. Then, (25) can be expressed as

$$\min_{\bar{\mathbf{z}}_i, \mathbf{t}_i} \quad \left\| \mathbf{y}_i - \begin{bmatrix} \mathbf{E} & -\mathbf{1} \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{z}}_i^T \\ \mathbf{t}_i \end{bmatrix} \right\|_1. \tag{27}$$

To eliminate the translation ambiguity in $\bar{\mathbf{z}}_i$, we enforce the centroid of $\bar{\mathbf{z}}_i$ to be at the origin by altering the problem as

$$\begin{aligned} \min_{\bar{\mathbf{z}}_i, \mathbf{t}_i} \quad & \left\| \begin{bmatrix} \mathbf{y}_i \\ 0 \end{bmatrix} - \begin{bmatrix} \mathbf{E} & -\mathbf{1} \otimes \mathbf{I} \\ \mathbf{1}^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \bar{\mathbf{z}}_i^T \\ \mathbf{t}_i \end{bmatrix} \right\|_1 \\ & \triangleq \|\mathbf{y}_i' - \mathbf{E}'\bar{\mathbf{z}}_i'\|_1. \end{aligned} \tag{28}$$

The above problem can be efficiently solved by ADMM, by introducing an auxiliary variable $\mathbf{u}_i = \mathbf{y}_i' - \mathbf{E}'\bar{\mathbf{z}}_i'$. In each iteration, $\bar{\mathbf{z}}_i'$ and $\mathbf{u}_i$ can be updated by solving a linear least squares and by performing a soft thresholding [6], respectively. Note that $\mathbf{E}'$ is a sparse matrix with ones and zeros, and each row contains only two ones except the last row. We can take advantage of this fact in solving the linear least squares to speed up the computation, which was the reason for choosing (24). In fact, this step takes much less time than the weak reconstruction step.

4. Experimental Results

We have compared the performance of the proposed method with other existing methods, such as EM-PPCA [27], MP [23], CSF2 [17], EM-PND [20], and BMM [10]. Note that the performances of the part-based reconstruction methods in [11] and [24] were also quoted from the papers for comparison. We have tested various data sets including the popular benchmark sets [5, 16, 27], as well as the dense sequences from [29], the real data sets from [25], and the optical flow results [3][7] of the human back video [24].

Note that, however, the benchmark sets used in [5] (drink, pickup, stretch, and yoga) contains synthetic camera rotations, five degrees per frame, of which the total amounts are from about 1500 to 5500 degrees. These camera motions can be considered very fast and unrealistic, which was added to the data artificially because it was quite hard in the early stages of NRSfM research to achieve good performance on realistic data. This, however, should be overcome for an NRSfM scheme in order to be practical. Therefore, we have altered these benchmark sets as follows: We have removed these synthetic rotations from the data sets and added smaller ones, starting from the one-fourth position of the entire frames and ending at the half position of the

[7] We would like to thank Dr. Antonio Agudo for providing this data.

Table 1. Average reconstruction errors for the benchmark data sets [16, 27].

data set	EM-PPCA	MP	CSF2	EM-PND	BMM	proposed					[24]	[11]
						mean	std.	max	min	# groups		
dance	0.2096	0.3968	0.1370	0.1834	0.1454	**0.0759**	0.0020	0.0803	0.0724	478.2	–	–
capoeira	0.4476	0.4118	0.3005	0.3312	0.2465	**0.1725**	0.0010	0.1745	0.1699	271.6	–	–
walking	0.1248	0.2488	0.0709	0.0465	0.0862	**0.0396**	0.0003	0.0403	0.0389	356.6	–	–
face	0.0209	0.0320	0.0209	**0.0165**	0.0233	0.0248	0.0003	0.0255	0.0241	296.8	–	–
shark	0.0366	0.0874	0.0551	**0.0134**	0.1669	0.0832	0.0000	0.0832	0.0831	563.9	–	–
flag	0.1325	0.1761	0.1609	0.1313	0.1741	0.0387	0.0007	0.0396	0.0378	3088.3	0.0325	**0.0159**
pace	0.0892	0.1399	0.0882	0.0736	0.0892	**0.0648**	0.0004	0.0653	0.0643	8053.0	–	–

Table 2. Average reconstruction errors for the data sets in [5] with realistic rotations.

data set		EM-PPCA	MP	CSF2	EM-PND	BMM	proposed				
total rotation	sequence						mean	std.	max	min	# groups
60°	drink	0.1934	0.1744	0.0993	0.0952	**0.0411**	0.0431	0.0003	0.0438	0.0424	265.0
	pickup	2.3232	1.8799	0.2821	0.5371	0.1580	**0.1281**	0.0017	0.1339	0.1252	269.9
	stretch	1.2543	1.7585	0.1277	0.1356	0.0971	**0.0939**	0.0016	0.0973	0.0899	289.0
	yoga	0.4905	0.4939	0.2909	0.4063	0.2463	**0.1845**	0.0036	0.1946	0.1777	264.4
90°	drink	0.2108	0.1952	0.0712	0.0371	0.0919	**0.0353**	0.0001	0.0357	0.0350	263.3
	pickup	0.5140	0.7674	0.1825	0.2650	0.1011	**0.0918**	0.0023	0.0981	0.0882	271.0
	stretch	0.2657	0.2725	0.1523	0.1098	**0.0773**	0.0797	0.0013	0.0827	0.0771	273.9
	yoga	0.6949	0.5445	0.2221	0.3020	0.1686	**0.1190**	0.0026	0.1243	0.1118	285.8
120°	drink	0.2261	0.3026	0.0649	0.0753	0.0424	**0.0304**	0.0001	0.0306	0.0301	261.4
	pickup	0.6891	0.6996	0.2838	0.3635	0.1104	**0.0964**	0.0012	0.0991	0.0940	273.1
	stretch	0.4655	0.3326	0.1576	0.2858	0.0930	**0.0846**	0.0012	0.0883	0.0826	289.4
	yoga	0.4756	0.4359	0.1785	0.3497	0.1309	**0.1115**	0.0024	0.1180	0.1056	261.8

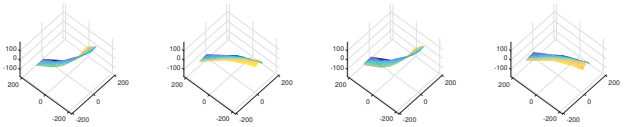
frames. This means that there were standing-still periods at the beginning and the ending of the data sets. The total amount of rotations were set to various values; 60, 90 and 120 degrees. These camera motions are much more practical and reflect real situations. The rotations were made around the y-axis, as like the original synthetic rotations.

The parameters of the proposed method were set as follows; $n_g = 10$, $m_g = 50$, and $\lambda = 0.1$.[8] The parameters of the other methods were set as in their papers. The number of bases K for CSF2 and BMM were tuned to achieve the best performance, unless there were corresponding experiments in their paper. Since the proposed method relies on random sampling, we repeated the experiment 50 times for each data sets, except the flag and pace sequences [29] because of the increased processing time due to the large data size. One thing to note is that the rotation calculation scheme (so-called "rectification" algorithm) in the shape-basis approach can be unstable for co-planar structures, which is not an exception for the proposed weak reconstructor. Hence, we have modified the rotation calculation step for the cloth-like objects (flag [29], two-cloths, and tear [25]): The Tomasi-Kanade factorization [26] was applied to a trajectory group to find a rigid 3D approximation and a closest two-dimensional plane was calculated by SVD. Then, this plane was rotated for each frame minimizing the reprojection error to find the rotation estimates.

The measure of performance was the (normalized) reconstruction error, *i.e.*, $e_{3D} = \sum_i \|\mathbf{X}_i - \hat{\mathbf{X}}_i\| / \|\hat{\mathbf{X}}_i\|$ where $\mathbf{X}_i$ and $\hat{\mathbf{X}}_i$ are the 3D coordinates of the reconstructed and the ground truth shapes, respectively, in the ith frame. Note that there can be reflection ambiguities, and especially for most of the existing schemes, all the frames may have different reflections. Therefore, the errors have been measured also on the inverted shapes for each frame and the smaller ones were picked. Note that, only for the flag sequence, the errors were evaluated after aligning the reconstructed shapes to the ground truth based on the Procrustes alignment [18], as in the practice in [11, 24].

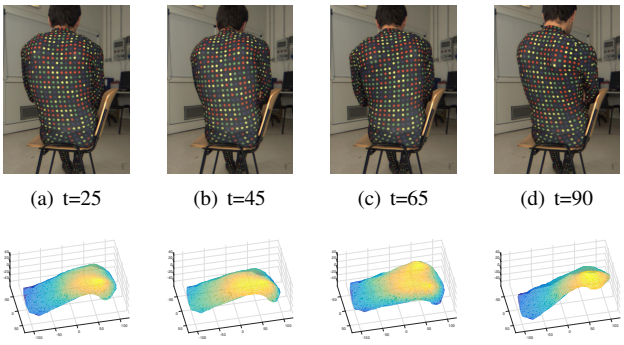
Table 1 shows the performance for the benchmark sets from [16, 27] and the dense sets, flag and pace, from [29]. Here, the performance for many data sets was not denoted for the other part-based schemes [11, 24], because they are specialized for smooth surfaces. Note that, in this table, the proposed method gives the best performance for the dance, capoeira, walking, and pace sequences. These data sets all have either complex, large deformations or restricted view points, which confirms the effectiveness of the proposed algorithm to the aforementioned issues of NRSfM. Especially, the dance and capoeira sequences have been the difficult ones for many NRSfM schemes, and the proposed method achieves dramatic improvements for them. On the other hand, the face and shark sequences have very simple deformations with enough view angles, hence in this case, all-trajectory-based schemes can give better results than our method that relies on randomly sampled parts. Nevertheless, the performance of the proposed algorithm is comparable, which is good enough for practical uses. The flag sequence is basically a cloth floating in the wind, and the other methods specialized for smooth surfaces give better performance. Yet, the proposed method shows a competitive result even though its procedure is much simpler. The

[8] More experiments on different parameters can be found in the supplementary material.



(a) t=200 (b) t=400 (c) t=200 (d) t=400

Figure 1. Reconstruction examples of the flag sequence. (a) and (b) are ground truth shapes, while (c) and (d) are their reconstructions. Procrustes alignment was *not* applied for this visualization.



(a) t=25 (b) t=45 (c) t=65 (d) t=90

(e) t=25 (f) t=45 (g) t=65 (h) t=90

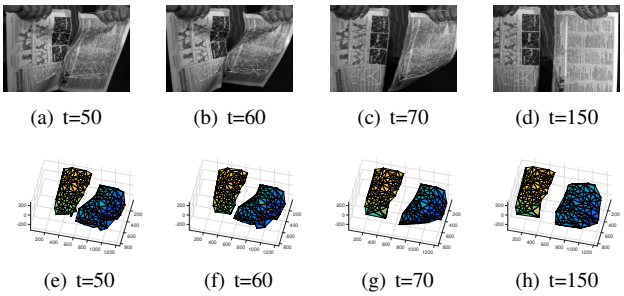
Figure 2. Reconstruction examples of the back sequence.

reconstruction examples of the flag sequence are shown in Fig. 1, which confirm the reasonable performance of the proposed method. A good thing about the proposed algorithm is that its performance has low standard deviations, which attests its reliability.

Table 2 shows the performance for the benchmarks sets from [5] with realistic rigid motions. In this table, we can see that the proposed method achieves the best performance for most of the cases. The performance of the other methods often gets severely bad with the realistic rigid motions, even though the deformations in these data sets are not much complicated. This verifies the bias in the existing algorithms towards the well-conditioned examples with large rigid motions. On the other hand, the proposed method shows steady performance for all of the cases.

Finally, we have conducted experiments for the real data sets [3, 25].[9] The reconstructed results are shown in Figs. 2, 3 and 4. These figures show that the proposed algorithm is capable of reconstructing real scenes with no problem. An important aspect of the proposed method can be seen in the tear sequence, where a piece of paper is teared apart into two pieces. Even though the object is dividing into two parts, the proposed algorithm can handle the reconstruction because of its part-based nature. This gives a possibility of bypassing object or part segmentation problem arises in NRSfM [13], which is not the main scope of this paper.

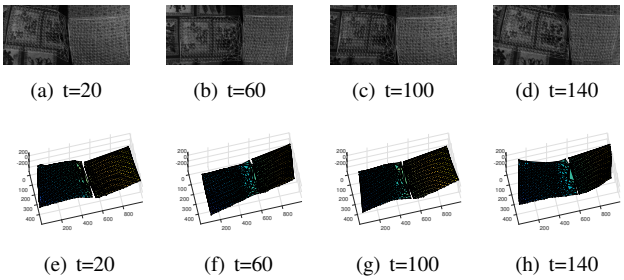
[9]We could not evaluated the quantitative performance for these data because there were no ground truths. The videos of reconstructed results are provided in the supplementary material.



(a) t=50 (b) t=60 (c) t=70 (d) t=150

(e) t=50 (f) t=60 (g) t=70 (h) t=150

Figure 3. Reconstruction examples of the tear sequence.



(a) t=20 (b) t=60 (c) t=100 (d) t=140

(e) t=20 (f) t=60 (g) t=100 (h) t=140

Figure 4. Reconstruction examples of the two cloths sequence.

5. Conclusion

In this paper, we proposed a novel part-based NRSfM framework, which considers the parts as weak reconstructions and obtains consensus from them. The proposed method is composed of four steps; sampling, weak reconstructions, reflection correction, and calculation of statistics. It does not have any strong assumption about the sampled parts, which makes the sampling method very simple. BMM has been adopted and modified for the weak reconstructions, and a single run of the eigenvalue decomposition can resolve the reflection ambiguities. The final step of the algorithm can be formulated as $l1$-norm minimization, which can be efficiently solved by ADMM. The most time-consuming step, weak reconstructions, can be easily parallelized, which is a useful advantage. The proposed method does not have an assumption on the type of data and gives the state-of-the-art results for the benchmark sets. The proposed framework is very simple and flexible, and there are good chances that many other important issues in NRSfM could be efficiently dealt with; *e.g.*, simultaneous object segmentation and reconstruction, handling missing points, sequential reconstruction, and so on, which will be addressed in future work.

Acknowledgment

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1006269).

References

[1] A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good Vibrations: A Modal Analysis Approach for Sequential Non-Rigid Structure from Motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2014.

[2] A. Agudo and F. Moreno-Noguer. Learning Shape, Motion and Elastic Models in Force Space. In *Proc. IEEE Int'l Conf. Computer Vision*, December 2015.

[3] A. Agudo and F. Moreno-Noguer. Simultaneous Pose and Non-Rigid Shape with Particle Dynamics. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 201k.

[4] I. Akhter, Y. Seikh, and S. Khan. In Defense of Orthonormality Constraints for Nonrigid Structure from Motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2009.

[5] I. Akhter, Y. Seikh, S. Khan, and T. Kanade. Trajectory Space: A Dual Representation for Nonrigid Structure from Motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(7):1442–1456, July 2011.

[6] A. Beck and M. Teboulle. A Fast Iterative Shrinkage-thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Science*, 2(1):183–202, 2009.

[7] J. Bolte, S. Sabach, and M. Teboulle. Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems. *Mathematical Programming*, 146(1):459–494, August 2014.

[8] L. Breiman. Bagging Predictors. *Machine Learning*, 24(2):123–140, August 1996.

[9] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001.

[10] Y. Dai, H. Li, and M. He. A Simple Prior-free Method for Non-Rigid Structure-from-Motion Factorization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2012.

[11] J. Fayad, L. Agapito, and A. D. Bue. Piecewise Quadratic Reconstruction of Non-Rigid Surface from Monocular Sequences. In *Proc. European Conference on Computer Vision*, September 2010.

[12] J. Fayad, C. Russell, and L. Agapito. Automated Articulated Structure and 3D Shape Recovery from Point Correspondences. In *Proc. IEEE Int'l Conf. Computer Vision*, November 2011.

[13] K. Fragkiadaki, M. Salas, P. Arbelaez, and J. Malik. Grouping-Based Low-Rank Trajectory Completion and 3D Reconstruction. In *Proc. Neural Information Processing Systems*, December 2014.

[14] J. Gall and V. Lempitsky. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013.

[15] R. Garg, A. Roussos, and L. Agapito. Dense Variational Reconstruction of Non-rigid Surfaces from Monocular Video. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2013.

[16] P. F. U. Gotardo and A. M. Martinez. Kernel Non-Rigid Structure from Motion. In *Proc. IEEE Int'l Conf. Computer Vision*, November 2011.

[17] P. F. U. Gotardo and A. M. Martinez. Non-Rigid Structure from Motion with Complementary Rank-3 Spaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2011.

[18] J. C. Gower. Generalized Procrustes Analysis. *Psychometrika*, 40(1):33–51, March 1975.

[19] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge university press, 2012.

[20] M. Lee, J. Cho, C.-H. Choi, and S. Oh. Procrustean Normal Distribution for Non-Rigid Structure from Motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2013.

[21] M. Lee, C.-H. Choi, and S. Oh. A Procrustean Markov Process for Non-rigid Structure Recovery. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2014.

[22] Z. Lin, M. Chen, L. Wu, and Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-rank Matrices. Technical Report UILU-ENG-09-2215, University of Illinois at Urbana-Champaign, November 2009.

[23] M. Paladini, A. Del Bue, M. Stošić, M. Dodig, J. ao Xavier, and L. Agapito. Factorization for Non-Rigid and Articulated Structure using Metric Projections. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2009.

[24] C. Russell, J. Fayad, and L. Agapito. Energy Based Multiple Model Fitting for Non-rigid Structure from Motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2011.

[25] J. Taylor, A. Jepson, and K. Kutulakos. Non-rigid Structure from Locally-rigid Motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2010.

[26] C. Tomasi and T. Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *Int'l J. Computer Vision*, 9(2):137–154, November 1992.

[27] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(5):878–892, May 2008.

[28] P. Viola and M. J. Jones. Robust Real-time Face Detection. *Int'l J. Computer Vision*, 57(2):137–154, May 2004.

[29] R. White, K. Crane, and D. Forsyth. Capturing and Animating Occluded Cloth. In *ACM Trans. Graphics (SIGGRAPH)*, August 2007.

[30] J. Xiao, J. Chai, and T. Kanade. A Closed-Form Solution to Non-Rigid Shape and Motion Recovery. *Int'l J. Computer Vision*, 67(2):233–246, April 2006.