2013 International Conference on Computer and Robot Vision

Robust Structure from Motion of Nonrigid Objects in the Presence of Outlying and Missing Data

Guanghui Wang, John S. Zelek
Department of Systems Design Engineering
University of Waterloo
200 University Avenue West, Waterloo, Ontario N2L 3G1
Email: {g29wang, jzelek}@uwaterloo.ca

Q. M. Jonathan Wu
Department of Electrical and Computer Engineering
University of Windsor
401 Sunset Ave, Windsor, Ontario N9B 3P4
Email: jwu@uwindsor.ca

***Abstract*—The paper focuses on robust 3D structure from motion of nonrigid objects from uncalibrated image sequences. A new affine factorization algorithm is first proposed to avoid the difficulty in image alignment for imperfect data, followed by a robust factorization scheme to handle outlying and missing data. The novelty and main contributions of the paper are as follows: (i) as a new addition to previous nonrigid structure from motion, the proposed factorization algorithm can perfectly handle imperfect tracking data; (ii) it is demonstrated that the image reprojection residuals are in general proportional to the error magnitude of the tracked features. Thus, the outliers can be detected directly from the residuals, which are then used to estimate the uncertainties of the image measurement; and (iii) the robust factorization scheme is proved empirically to be more efficient and more accurate than other robust algorithms. The proposed approach is validated and evaluated by extensive experiments on synthetic data and real image sequences.**

***Keywords*-computer vision, structure from motion, nonrigid factorization, robust algorithm**

I. Introduction

Structure from motion (SfM) is an important task in computer vision. Over the past two to three decades, considerable progress has been made in SfM and the results have been successfully applied in robot navigation, industrial inspection, autonomous vehicles, and digital entertainment. The typical approach for SfM is the factorization algorithm which was first proposed by Tomasi and Kanade [19]. The algorithm assumes that the tracking matrix of an image sequence is available and deals uniformly with the data from all images. Thus, it is more robust and more accurate than the methods that use only two or three images [15][21].

The main idea of the factorization algorithm is to decompose the tracking matrix into the motion and structure components simultaneously by Singular Value Decomposition (SVD) with low-rank approximation. Most of the studies on the problem assume an affine camera model due to its linearity [10]. Christy and Horaud [5] extended the method to a perspective camera model by incrementally performing the affine factorization of a scaled tracking matrix. Triggs [21] proposed a full projective factorization algorithm with projective depths recovered from epipolar geometry. The method was further studied and different iterative schemes were proposed to recover the projective depths by minimizing image reprojection errors [14].

The factorization algorithm was extended to nonrigid SfM by assuming that the 3D shape of a nonrigid object can be modeled as a weighted linear combination of a set of shape bases [4]. Thus, the shape bases and camera motions are factorized simultaneously for all time instants under a rank-$3k$ constraint of the tracking matrix. The method has been extensively investigated and developed in [3] [20]. Recently, Rabaud and Belongie [17] relaxed the Bregler's assumption and developed a manifold-learning framework to solve the problem. Yan and Pollefeys [25] extended the factorization approach to recover the structure of articulated objects. Akhter *et al.* [2] proposed a dual approach to describe the nonrigid structure in trajectory space by a linear combination of basis trajectories.

Most factorization methods assume that all features are tracked across the sequence. In the presence of missing data, SVD factorization cannot be used directly, researches proposed to solve the motion and shape matrices alternatively, such as the alternative factorization [12], power factorization [8], and factor analysis [7]. In practice, outlying data are inevitable during the process of feature tracking, as a consequence, performance of the algorithm will degrade. Some popular strategies to handle outliers in computer vision field are RANSAC, Least Median of Squares [9], and other similar hypothesise-and-test frameworks [18]. However, these methods are usually designed for two or three views and they are computational expensive.

In recent years, the problem of robust factorization has received a lot of attention [27]. Aguitar and Moura [1] proposed a scalar-weighted SVD algorithm that minimizes the weighted square errors. Gruber and Weiss [7] formulated the problem as a factor analysis and derived an Expectation Maximization (EM) algorithm to enhance the robustness to missing data and uncertainties. Zelnik-Manor *et al.* [28] defined a new type of motion consistency based on temporal consistency, and applied it to multi-body factorization with

978-0-7695-4983-5/13 $26.00 © 2013 IEEE
DOI 10.1109/CRV.2013.39

directional uncertainty. Zaharescu and Horaud [27] introduced a Gaussian mixture model and incorporate it with the EM algorithm. Huynh *et al.* [11] proposed an iterative approach to correct the outliers with 'pseudo' observations.

Ke and Kanade [12] designed a robust algorithm to handle outliers by minimizing a $L1$-norm of the reprojection errors. Eriksson and Hengel [6] introduced the $L1$-norm to the Wiberg algorithm to handle missing data and outliers. Okatani *et al.* [13] proposed to incorporate a damping factor into the Wiberg method to solve the problem. Yu *et al.* [26] presented a Quadratic Program formulation for robust multi-model fitting of geometric structures. Wang *et al.* [24] proposed an adaptive kernel-scale weighted hypotheses to segment multiple-structure data even in the presence of a large number of outliers. Paladini *et al.* [16] developed an alternating bilinear approach to SfM by introducing a globally optimal projection step of the motion matrices onto the manifold of metric constraints. Wang *et al.* [23] proposed a spatial-and-temporal-weighted factorization approach to handle significant noise in the measurement.

The above robust algorithms are initially designed for SfM of rigid objects. To the best of our knowledge, few studies have been carried out for nonrigid scenarios. In this paper, a robust nonrigid factorization approach is reported. The outlying data are detected from a new viewpoint via image reprojection residuals by exploring the fact that the reprojection residuals are largely proportional to the measurement errors. The paper first proposed a rank-$(3k+1)$ factorization algorithm to avoid the problem of image registration for imperfect data, followed by an alternative weighted factorization algorithm to handle the missing features and image uncertainty. Finally, a robust factorization scheme is proposed to deal with outliers.

II. Background of Nonrigid Factorization

Under affine projection, a 3D point $\mathbf{X}_j$ is projected onto $\mathbf{x}_{ij}$ in frame i according to the imaging equation

$$\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{X}_j + \mathbf{c}_i \tag{1}$$

where $\mathbf{A}_i$ is a 2×3 affine projection matrix; the translation term $\mathbf{c}_i$ is the image of the centroid of all space points. Let $\mathbf{S}_i = [\mathbf{X}_1, \cdots, \mathbf{X}_n]$ be the 3D structure associated with frame i, the structure may be different at different instants. In nonrigid SfM, we usually follow Bregler's assumption that $\mathbf{S}_i = \sum_{l=1}^{k} \omega_{il} \mathbf{B}_l$, where the nonrigid structure is assumed to be a linear combination of a set of rigid shape bases $\mathbf{B}_l$ [4]. Under this assumption, the imaging process of one image can be modeled as

$$\begin{aligned}
\mathbf{W}_i &= [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in}] = \mathbf{A}_i \mathbf{S}_i + [\mathbf{c}_i, \cdots, \mathbf{c}_i] \\
&= [\omega_{i1}\mathbf{A}_i, \cdots, \omega_{ik}\mathbf{A}_i] \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix} + [\mathbf{c}_i, \cdots, \mathbf{c}_i]
\end{aligned}$$

It is easy to verify that if all image points in each frame are registered to the centroid and relative image coordinates are employed, the translation term vanishes, i.e., $\mathbf{c}_i = \mathbf{0}$. Consequently, the nonrigid factorization under affine camera model is expressed as

$$\underbrace{\begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \cdots & \mathbf{x}_{mn} \end{bmatrix}}_{\mathbf{W}_{2m \times n}} = \underbrace{\begin{bmatrix} \omega_{11}\mathbf{A}_1 & \cdots & \omega_{1k}\mathbf{A}_1 \\ \vdots & \ddots & \vdots \\ \omega_{m1}\mathbf{A}_m & \cdots & \omega_{mk}\mathbf{A}_m \end{bmatrix}}_{\mathbf{M}_{2m \times 3k}} \underbrace{\begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix}}_{\mathbf{B}_{3k \times n}} \tag{2}$$

Structure from motion is a reverse problem. Suppose the tracking matrix $\mathbf{W}$ is available, our purpose is to recover the camera motion parameters in $\mathbf{M}$ and the 3D structure from the shape matrix $\mathbf{B}_i$. It is obvious from (2) that the rank of the tracking matrix $\mathbf{W}$ is at most $3k$. Previous studies on nonrigid SfM are based on the rank-$3k$ constraint due to its simplicity, and the factorization can be easily obtained via SVD decomposition by truncating its rank to $3k$.

III. Affine Factorization without Registration

One critical condition for equation (2) is that all image measurements are registered to the corresponding centroid of each frame. When the tracking matrix contains outliers and/or missing data, it is impossible to reliably retrieve the centroid. As will be shown in the experiments, the miscalculation of the centroid will cause a significant error to the final solutions. Previous studies were either ignoring this problem or hallucinating the missing points with pseudo observations, which may lead to a biased estimation. In this section, a rank-$(3k+1)$ factorization algorithm is proposed to solve this problem.

A. Rank-$(3k+1)$ Affine Factorization

Let us formulate the affine imaging process (1) in the following form

$$\mathbf{x}_{ij} = [\mathbf{A}_i | \mathbf{c}_i] \, \tilde{\mathbf{X}}_j \tag{3}$$

where $\tilde{\mathbf{X}}_j = [\mathbf{X}_j^T, t_j]^T$ is a 4-dimensional homogeneous expression of $\mathbf{X}_j$. Let $\tilde{\mathbf{S}}_i = \begin{bmatrix} \mathbf{S}_i \\ \mathbf{t}_i^T \end{bmatrix}$ be the homogeneous form of the deformable structure, then the imaging process of frame i can be written as

$$\begin{aligned}
\mathbf{W}_i &= [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in}] = [\mathbf{A}_i | \mathbf{c}_i] \begin{bmatrix} \sum_{l=1}^{k} \omega_{il} \mathbf{B}_l \\ \mathbf{t}_i^T \end{bmatrix} \\
&= [\omega_{i1}\mathbf{A}_i, \cdots, \omega_{ik}\mathbf{A}_i, \mathbf{c}_i] \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \\ \mathbf{t}_i^T \end{bmatrix}
\end{aligned}$$

Thus, the structure and motion factorization for the entire sequence is formulated as follows.

$$\underbrace{\begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \cdots & \mathbf{x}_{mn} \end{bmatrix}}_{\mathbf{W}_{2m \times n}} = \underbrace{\begin{bmatrix} \omega_{11}\mathbf{A}_1 & \cdots & \omega_{1k}\mathbf{A}_1 & \mathbf{c}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \omega_{m1}\mathbf{A}_m & \cdots & \omega_{mk}\mathbf{A}_m & \mathbf{c}_m \end{bmatrix}}_{\mathbf{M}_{2m \times (3k+1)}} \underbrace{\begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \\ \mathbf{t}_i^T \end{bmatrix}}_{\mathbf{B}_{(3k+1) \times n}} \tag{4}$$

Obviously, the rank of the tracking matrix becomes $3k+1$ in this case. Given the tracking matrix, the factorization can be easily obtained via SVD decomposition and imposing rank-$(3k+1)$ constraint. The expression (4) does not require any image registration thus can directly work with outlying and missing data.

Both factorization algorithms (2) and (4) can be equivalently denoted as the following minimization scheme.

$$f(\mathbf{M},\mathbf{S}) = \underset{\mathbf{M},\mathbf{S}}{\operatorname{argmin}} \|\mathbf{W}-\mathbf{M}\mathbf{S}\|_F^2 \tag{5}$$

By enforcing different rank constraints, the Frobenius norm of (5) corresponding to the algorithms (2) and (4) would be

$$E_{3k} = \sum_{i=3k+1}^{N} \sigma_i^2, \;\; E_{3k+1} = \sum_{i=3k+2}^{N} \sigma_i^2 \tag{6}$$

where $\sigma_i, i = 1, \cdots, N$ are singular values of the tracking matrix in descending order, and $N = \min(2m, n)$. Clearly, the error difference by the two algorithm is σ_{3k+1}^2. For noise free data, if all image points are registered to the centroid, then, $\sigma_i = 0, \forall i > 3k$, the equations (2) and (4) are actually equivalent. However, in the presence of outlying and missing data, the image centroid cannot be accurately recovered, the rank-$3k$ algorithm (2) will yield a big error since σ_{3k+1} does not approach zero in this situation.

B. Euclidean Upgrading Matrix

Suppose $\mathbf{W} = \hat{\mathbf{M}}\hat{\mathbf{B}}$ is a set of factorization result of (4). Obviously, the decomposition is not unique since it is only defined up to a nonsingular linear $\mathbf{H} \in \mathbb{R}^{(3k+1)\times(3k+1)}$ as $\mathbf{M} = \hat{\mathbf{M}}\mathbf{H}$ and $\mathbf{S} = \mathbf{H}^{-1}\hat{\mathbf{S}}$. The recovery of the upgrading matrix is different with that in the rank-$3k$ factorization.

Let us write the $(3k+1)\times(3k+1)$ upgrading matrix in the following form.

$$\mathbf{H} = [\mathbf{H}_1, \cdots, \mathbf{H}_k | \mathbf{h}_{3k+1}] \tag{7}$$

where $\mathbf{H}_l \in \mathbb{R}^{(3k+1)\times 3} (l = 1, \cdots, k)$ denotes the l-th triple columns of $\mathbf{H}$, and $\mathbf{h}_{3k+1}$ denotes the last column of $\mathbf{H}$. Suppose $\hat{\mathbf{M}}_i$ is the i-th two-row submatrix of $\hat{\mathbf{M}}$, then the upgraded motion matrix can be written as

$$\mathbf{M}_i = \hat{\mathbf{M}}_i\mathbf{H} = [\hat{\mathbf{M}}_i\mathbf{H}_1, \cdots, \hat{\mathbf{M}}_i\mathbf{H}_k | \hat{\mathbf{M}}_i\mathbf{h}_{3k+1}] \tag{8}$$

Comparing the above equation with (4), we have

$$\hat{\mathbf{M}}_i\mathbf{H}_l = \omega_{il}\mathbf{A}_i = \omega_{il} f_i \left[\begin{array}{c} \mathbf{r}_{i1}^T \\ \mathbf{r}_{i2}^T \end{array} \right] \tag{9}$$

where f_i is the focal length of the camera, $\mathbf{r}_{i1}^T$ and $\mathbf{r}_{i2}^T$ are the first two rows of the rotation matrix. Denote $\mathbf{Q}_l = \mathbf{H}_l\mathbf{H}_l^T$, then, $\mathbf{Q}_l$ can be constrained from (9) as

$$\hat{\mathbf{M}}_i\mathbf{Q}_l\hat{\mathbf{M}}_i^T = \big(\hat{\mathbf{M}}_i\mathbf{H}_l\big)\big(\hat{\mathbf{M}}_i\mathbf{H}_l\big)^T = \omega_{il}^2 f_i^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{10}$$

The matrix $\mathbf{Q}_l$ has $9k$ degree-of-freedom since it is a $(3k+1)\times(3k+1)$ positive semidefinite symmetric matrix defined up to a scale. The above equation provides two constraints to $\mathbf{Q}_l$, thus it can be linearly solved via least squares by stacking the constraints (10) frame by frame. Furthermore, the submatrix $\mathbf{H}_l$ can be decomposed from $\mathbf{Q}_l$ via extended Cholesky decomposition [22].

From equations (4) and (8), it is easy to prove that the last column of the upgrading matrix $\mathbf{h}_{3k+1}$ only influences the translation from the world coordinate system to the image system. Under a given coordinate system, different values of $\mathbf{h}_{3k+1}$ will only alter the origin of the world system, however, it does not change the Euclidean structure of the reconstructed points. Therefore, $\mathbf{h}_{3k+1}$ can be set freely as any $(3k+1)$-vector that is independent of the columns of $\{\mathbf{H}_l, l = 1, \cdots, k\}$ such that the resulted upgrading matrix is nonsingular.

After recovering the Euclidean upgrading matrix, the camera parameters, motions, shape bases, and deformation weights can be easily decomposed from the upgraded motion matrix $\hat{\mathbf{M}}\mathbf{H}$ and shape matrix $\mathbf{H}^{-1}\hat{\mathbf{B}}$.

IV. Alternative Factorization

Since SVD decomposition cannot directly work with missing data, researchers proposed different alternative approaches to handle missing data [8]. In this section, a two-step alternative factorization algorithm is introduced.

A. Alternative Factorization Algorithm

The basic idea of the two-step factorization is to minimize the cost function (5) over $\mathbf{S}$ and $\mathbf{M}$ alternatively until convergence, while leaving the other one fixed, i.e.,

$$f(\mathbf{S}) = \underset{\mathbf{S}}{\operatorname{argmin}} \|\mathbf{W}-\mathbf{M}\mathbf{S}\|_F^2 \tag{11}$$

$$f(\mathbf{M}) = \underset{\mathbf{M}}{\operatorname{argmin}} \|\mathbf{W}-\mathbf{M}\mathbf{S}\|_F^2 \tag{12}$$

Each cost function of the algorithm is indeed a convex function thus a global minimum can be found. The algorithm converges very fast if the tracking matrix is close to rank-$(3k+1)$ even with a random initialization. Different to SVD decomposition, the minimization process is carried out by least squares (LS).

Rewrite the cost function (11) as follows.

$$f(\mathbf{s}_j) = \underset{\mathbf{s}_j}{\operatorname{argmin}} \|\mathbf{w}_j-\mathbf{M}\mathbf{s}_j\|_F^2 \tag{13}$$

where $\mathbf{w}_j$ is the j-th column of $\mathbf{W}$, and $\mathbf{s}_j$ is the j-th column of $\mathbf{S}$. The LS solution of $\mathbf{S}$ can be given by

$$\mathbf{s}_j = \mathbf{M}^\dagger\mathbf{w}_j, \; j = 1, \cdots, n \tag{14}$$

where $\mathbf{M}^\dagger = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ is the Moore-Penrose pseudoinverse of $\mathbf{M}$. Missing data can be easily handled by the algorithm. For example, if some entries in the tracking matrix $\mathbf{w}_j$ are unavailable, one can simply delete those elements in $\mathbf{w}_j$ and the corresponding columns in $\mathbf{M}^\dagger$, then $\mathbf{s}_j$ can still be solved from (14) in the least-square sense using the remaining data. Similarly, the motion matrix is solved from the second cost function (12) as follows.

$$\mathbf{m}_i^T = \mathbf{w}_i^T \mathbf{S}^\dagger, \ i = 1, \cdots, m \tag{15}$$

where the pseudoinverse $\mathbf{S}^\dagger = \mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1}$, $\mathbf{m}_i^T$ and $\mathbf{w}_i^T$ denote the i-th row of the matrices $\mathbf{M}$ and $\mathbf{W}$, respectively.

B. Alternative Weighted Factorization

Measurement errors are inevitable in the process of feature detection and tracking. In order to increase the robustness of the algorithm, one common practice is to introduce a weight matrix into the cost function.

$$f(\mathbf{M}, \mathbf{S}) = \underset{\mathbf{M},\mathbf{S}}{\operatorname{argmin}} \|\boldsymbol{\Sigma} \otimes (\mathbf{W} - \mathbf{M}\mathbf{S})\|_F^2 \tag{16}$$

where $'\otimes'$ denotes the Hadamard product of element-by-element multiplication; $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ is the weight matrix whose entries are derived from the confidence of the image measurements. Many researchers have proposed different schemes to solve the problem [27]. In this section, the solutions of (16) are obtained using the alternative factorization algorithm as follows.

$$\begin{aligned} f(\mathbf{S}) &= \underset{\mathbf{s}_j}{\operatorname{argmin}} \|\boldsymbol{\Sigma}_j \otimes (\mathbf{w}_j - \mathbf{M}\mathbf{s}_j)\|_F^2 && (17) \\ f(\mathbf{M}) &= \underset{\mathbf{m}_i^T}{\operatorname{argmin}} \|\boldsymbol{\Sigma}_i^T \otimes (\mathbf{w}_i^T - \mathbf{m}_i^T \mathbf{S})\|_F^2 && (18) \end{aligned}$$

where $\boldsymbol{\Sigma}_j$ and $\boldsymbol{\Sigma}_i^T$ denote the j-th column and i-th row of $\boldsymbol{\Sigma}$, respectively. The close-form solutions of the shape and motion matrices are obtained by least squares.

$$\mathbf{s}_j = (\operatorname{diag}(\boldsymbol{\Sigma}_j)\mathbf{M})^\dagger \operatorname{diag}(\boldsymbol{\Sigma}_j)\mathbf{w}_j, \ j = 1, ..., n \tag{19}$$

$$\mathbf{m}_i^T = \mathbf{w}_i^T \operatorname{diag}(\boldsymbol{\Sigma}_i^T) \left(\mathbf{S}\operatorname{diag}(\boldsymbol{\Sigma}_i^T)\right)^\dagger, \ i = 1, ..., m \tag{20}$$

where $'\operatorname{diag}(\bullet)'$ stands for the diagonal matrix generated by a vector. The algorithm alternatively updates the shape and motion matrices until convenience. In case of missing data, one can simply delete the corresponding elements in equations (19) and (20), respectively.

V. Outlier Detection and Robust Factorization

Based on the foregoing proposed factorization algorithm, A fast and practical scheme for outlier detection is discussed in this section.

A. Outlier Detection Scheme

The best fit model of the factorization algorithm is obtained by minimizing the sum of squared residuals between the observed data and the fitted values provided by the model. Extensive empirical studies show that the least-square solutions are usually reasonable even in the presence of certain amount of outliers, and the reprojection residuals of the outlying data are significantly larger than those associated with inliers.

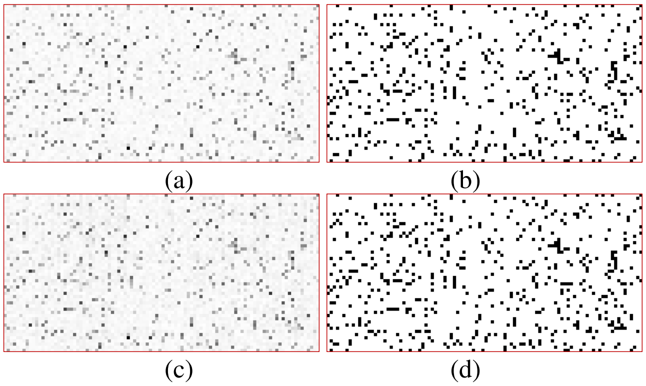


Figure 1. (a) Illustration of the normalized absolute value of the added Gaussian noise, where the intensity of each pixel corresponds the error magnitude at that point; (b) the distribution of the real added outliers; (c) the normalized value of the reprojection errors; (d) the outliers segmented from the reprojection errors by applying a single threshhold. Only 50 frames of 100 points are showed in the image.

Suppose $\hat{\mathbf{M}}$ and $\hat{\mathbf{S}}$ are a set of initial solutions of the motion and structure matrices, the reprojection residuals can be computed by reprojecting the solutions back onto all images. Let us define a residual matrix as follows.

$$\mathbf{E} = \mathbf{W} - \hat{\mathbf{M}}\hat{\mathbf{S}} = \begin{bmatrix} \mathbf{e}_{11} & \cdots & \mathbf{e}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{e}_{m1} & \cdots & \mathbf{e}_{mn} \end{bmatrix}_{2m \times n} \tag{21}$$

where

$$\mathbf{e}_{ij} = \mathbf{x}_{ij} - \hat{\mathbf{M}}_i \hat{\mathbf{s}}_j = \begin{bmatrix} \Delta u_{ij} \\ \Delta v_{ij} \end{bmatrix} \tag{22}$$

is the residual of point (i, j) in both image directions. The reprojection error of a point is defined as $\|\mathbf{e}_{ij}\|$, which is the Euclidean norm of the residual at that point.

Bellow is an example of the residual matrix and reprojection errors. Using the synthetic data in Section VI, we added 3-unit Gaussian noise and 10% outliers (significant noise greater than 5-unit) to the images. Then, the reprojection error is estimated via back-projection of the solutions using rank-$(3k+1)$ factorization. The real added noise and the reprojection error are illustrated in Fig.1 as grayscale images, where the gray level of each pixel corresponds to the inverse magnitude of the error on that point, lower gray level (black points) stands for larger error. It is evident that the reprojection error and the added noise have similar distribution. The ground truth of the real added outliers is depicted as a binary image in Fig.1 (b). Fig.1 (d) shows the binarized image of the reprojection errors by simply applying a global threshold. Surprisingly, almost all outliers are successfully detected by a single threshhold.

B. Implementation details

Inspired by the above observation, an efficient outlier detection and robust factorization scheme is developed based on the reprojection residuals. The computational details of the proposed scheme are as follows.

Robust Factorization Algorithm

```
Input: Tracking matrix of the sequence
1. Perform rank-($3k+1$) factorization (use alternative factorization in case of missing data) to obtain an initial solutions of $\hat{\mathbf{M}}$ and $\hat{\mathbf{B}}$.
2. Estimate the reprojection residuals (21) from initial solutions.
3. Determine an outlier threshhold and eliminate the outliers.
4. Refine the solutions from the inliers via the alternative factorization algorithm.
5. Estimate the weight matrix $\mathbf{\Sigma}$ from the refined solutions.
6. Perform the weighted factorization using the inliers.
7. Recover the upgrading matrix $\mathbf{H}$ and upgrade the solutions to the Euclidean space: $\mathbf{M} = \hat{\mathbf{M}}\mathbf{H}$, $\mathbf{B} = \mathbf{H}^{-1}\hat{\mathbf{B}}$.
8. Recover the Euclidean structure $\mathbf{S}_i$ and motion parameters corresponding to each frame.
Output: 3D structure and camera motion parameters
```

The alternative factorization algorithm is employed in steps 4 and 6 to handle missing data, while the initial values are obtained from the previous steps. Although the alternative factorization can work with random initialization, a reliable initial values can speed up its convergence.

Two important parameters are required in the robust algorithm: one is the outlier threshhold, the other is the weight matrix. The following will discuss how to recover these parameters.

C. Parameter Estimation

Assuming Gaussian image noise, it is easy to prove that the reprojection residuals also follow Gaussian distribution. Fig.3 shows an example from the synthetic data in Section VI. 3-unit Gaussian noise and 10% outliers were added to the synthetic images, and the residual matrix was calculated from (21). As shown in Fig.3, the residuals are obviously follow Gaussian distribution. Thus, the outlier threshhold can be determined from the distribution of the residuals.

Let $\mathcal{V}(\mathbf{E})$ be a $2mn$-vector formed by the residual matrix $\mathbf{E}$, suppose μ and σ are the mean and standard deviation of $\mathcal{V}(\mathbf{E})$, then the outlier threshhold can be chosen as follows.

$$\theta = \kappa\sigma \tag{23}$$

where κ is a parameter which is usually set from 3.0 to 5.0. Register $\mathcal{V}(\mathbf{E})$ with respect to its mean μ, then the outliers are classified via the following criteria.

$$|\Delta u_{ij} - \mu| > \theta \text{ or } |\Delta v_{ij} - \mu| > \theta \text{ or } \left\| \begin{smallmatrix} \Delta u_{ij} - \mu \\ \Delta v_{ij} - \mu \end{smallmatrix} \right\| > \theta \tag{24}$$

Since the outliers have heavily influence to the estimation of the mean and standard deviation due to their large deviations, the mean is practically estimated from the data that are less than the median value of $|\mathcal{V}(\mathbf{E})|$.

$$\mu = \text{mean}\{\mathcal{V}'(\mathbf{E}) | \mathcal{V}'(\mathbf{E}) < \text{median}(|\mathcal{V}(\mathbf{E})|)\} \tag{25}$$

while the standard deviation is estimated from the median absolute deviation (MAD)

$$\sigma = 1.4826 \text{ median}(|\mathcal{V}(\mathbf{E}) - \text{median}(\mathcal{V}(\mathbf{E}))|) \tag{26}$$

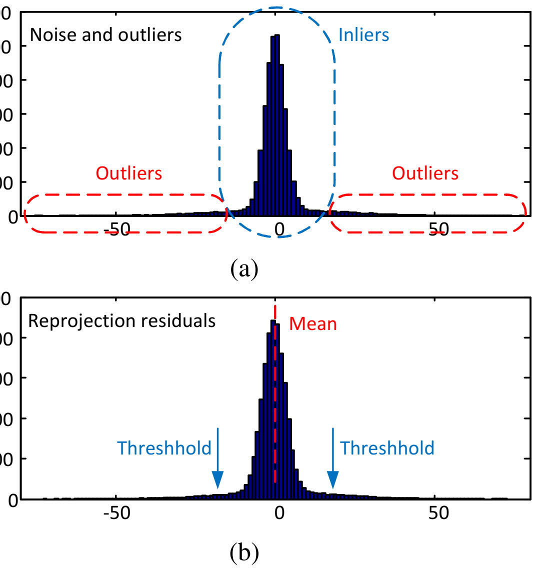


Figure 2. (a) Histogram distribution of the real added noise and outliers; (b) histogram distribution of the reprojection residuals.

The above computation is resistant to outliers and thus guarantees a robust estimation of the mean and standard deviation of the residuals.

The weight matrix is determined from the uncertainty of each feature based on the information such as sharpness and intensity contrast around its neighborhood [1][19]. The uncertainty is usually estimated during the process of feature detection and tracking or given as prior information. Nonetheless, this information is unavailable in many applications. In our early study [23], it was demonstrated that image uncertainty is generally proportional to the magnitude of reprojection residuals. The points with larger residuals have higher uncertainties, and vice versa. Based on this fact, the weight of each point is estimated directly from the residual matrix as follows.

$$\omega_{ij} = \frac{1}{\mathcal{N}} \exp\left(-\frac{\mathbf{E}_{ij}^2}{2\sigma^2}\right) \tag{27}$$

where the standard deviation σ is estimated from the median absolute deviation (26) using the data after eliminating the outliers, $\mathbf{E}_{ij}$ is the (i,j)-th element of the residual matrix (21), and $\mathcal{N}$ is a normalization scale. Clearly, the weight of a point is directional, it may have different values at different coordinate directions based on its residual. The points with higher residuals get lower weights, and the weights of missing data and outliers are set at zeros.

VI. Evaluations on Synthetic Data

The proposed technique was tested and evaluated extensively on synthetic data.

During the simulation, we generated a deformable space cube, which was composed of 21 evenly distributed rigid

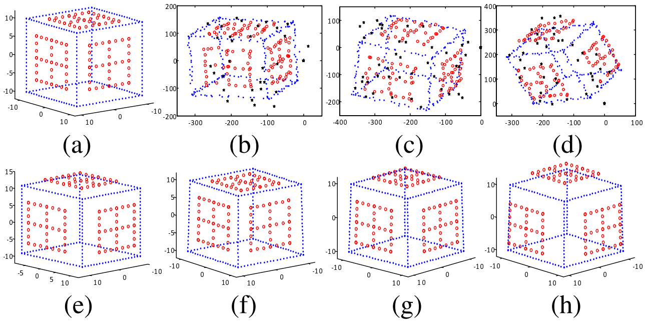


Figure 3. (a) (e) Two simulated space cubes with three sets of moving points; (b) (c) (d) three synthetic images with noise and outliers (black stars); (f) (g) (h) the reconstructed 3D structures corresponding to the three images (b), (c), and (d).

points on each side and three sets of dynamic points (33×3 points) on the adjacent surfaces of the cube that were moving outward. There are 252 space points in total as shown in Fig.3. Using the synthetic cube, 100 images were generated by affine projection with randomly selected camera parameter. Each image corresponds to a different 3D structure. The image resolution is 800×800 units and Gaussian white noise is added to the synthetic images.

A. *Influence of Image Centroid*

The influence of the centroid was evaluated in this test. We deliberately deviated the centroid of the features in each image from 1% to 9% in steps of 2%, then registered all image points to the deviated centroid. This is a simulation of the situation that the centroid could not be reliably recovered due to missing and outlying data.

Using the misaligned data, the motion and shape matrices were recovered using the rank-$(3k+1)$ factorization algorithm and its rank-$3k$ counterpart. The performance of different algorithms were evaluated and compared by means of the bellow defined mean reprojection variance.

$$E_{rv} = \frac{1}{mn}\|\mathbf{W}_0 - \hat{\mathbf{M}}\hat{\mathbf{S}}\|_F^2 \tag{28}$$

where $\mathbf{W}_0$ is the noise-free tracking matrix; $\hat{\mathbf{M}}$ and $\hat{\mathbf{S}}$ are the estimated motion and shape matrices, respectively. In order to obtain a statistically meaningful comparison, 100 independent tests were performed at each noise level. The mean reprojection variance with respect to different centroid deviations at different noise levels is shown in Fig. 4, where the noise level is defined as the standard deviation of the Gaussian noise.

As shown in Fig. 4, the miscalculated centroid has no influence to the proposed rank-$(3k+1)$ algorithm, however, it has an extremely large impact to the performance of rank-$3k$ based approach. The test indicates that the influence caused by the centroid error is far more significant than that caused by the image noise. Thus, it is better to choose rank-$(3k+1)$ affine factorization in practice, especially in cases that the centroid cannot be reliably recovered due to the presence of missing data and/or outliers.

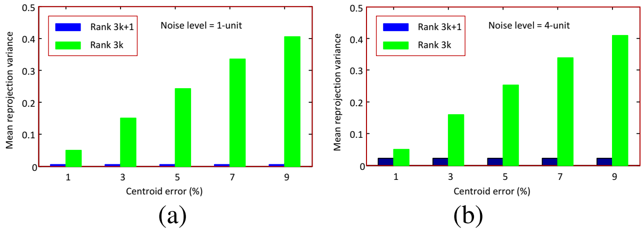


Figure 4. The mean reprojection variance with respect to different centroid deviations at the noise level of (a) 1-unit and (b) 4-unit.

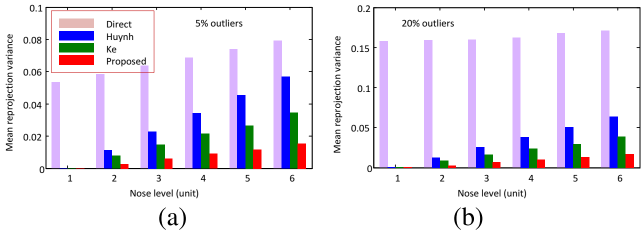


Figure 5. The mean reprojection variance with respect to different noise levels and outliers. (a) 5% outliers; (b) 20% outliers.

B. *Performance of the Robust Algorithm*

For the above simulated image sequence, Gaussian noise was added to each image point and the noise level was varied from 1-unit to 5-unit in steps of 1. In the mean time, 10% outliers were added to the tracking matrix. Using the contaminated data, the foregoing proposed robust algorithm was employed to recover the motion and shape matrices. Fig.3 shows three noise and outlier corrupted images and the corresponding 3D structures recovered by the proposed approach. It is evident that the deformable cube structures are correctly retrieved.

As a comparison, the direct nonrigid factorization algorithm without outlier rejection [20] and two successful robust algorithms in the literature were implemented as well, one is an outlier correction scheme proposed by Huynh *et al.* [11], the other one is proposed by Ke and Kanade [12] based on minimization of the $L1$ norm. The two robust algorithms were employed to recover the nonrigid structure using the same data in the test. The mean reprojection variance at different noise levels and outliers ratios is shown in Fig.5.

The results in Fig.5 were evaluated from 100 independent tests, and the reprojection variance was estimated only using the original inlying data so as to provide a fair comparison. Obviously, the proposed scheme outperforms other algorithms in terms of accuracy. The direct factorization algorithm yields significantly large errors due to the influence of outliers, and the error increases with the increase of the amount of outliers. The experiment also shows that all three robust algorithms are resilient to outliers, as can be

Table I
REAL COMPUTATION TIME OF DIFFERENT ALGORITHMS (SECOND)

Frame no.	50	100	150	200	250	300
Huynh	3.62	12.25	23.18	41.93	70.54	99.27
Ke	5.48	32.61	90.17	176.42	303.26	527.98
Proposed	15.76	56.08	79.86	136.14	212.39	303.46

seen in Fig. 5, the ratio of outliers has little influence to the reprojection variance of the three robust algorithms.

The complexity of different approaches was compared in the above test. All algorithms were implemented using Matlab on a Lenovo T500 laptop with 2,26GHz Intel Core Duo CPU. The frame number was varied from 50 to 300 in steps of 50 so as to generate different sizes of the tracking data, and 10% outliers were added to the data. Table I shows the real computation time of different algorithms. Obviously, the complexity of the proposed scheme lies in between of 'Huynh' and 'Ke', but it yields the best accuracy. The minimization of $L1$ norm in 'Ke' is computationally more intensive than the alternative factorization algorithm. Huynh's method does not include the step of weighted factorization, this is why it is relatively fast but it yields the lowest accuracy among the three algorithms.

VII. EVALUATIONS ON REAL SEQUENCES

The method was tested on many real image sequences. The results on two data sets are reported here.

The first test was on a dinosaur sequence sequence from the literature [2]. The sequence consists of 231 images with various movement and deformation of a dinosaur model. The image resolution is 570×338 pixel and 49 features were tracked across the sequence. In order to test the robustness of the algorithm, an additional 8% outliers were added to the tracking data as shown in Fig.6.

Using the proposed approach, all outliers were successfully detected, however, a few tracked features were also eliminated due to large tracking errors. The proposed approach was employed to remove the outliers and recover the motion and structure matrices. Then, the solutions were upgraded to the Euclidean space. Fig.6 shows the reconstructed structure and wireframe at different viewpoints. It can be seen from the results that the deformed structure has been correctly recovered from the corrupted data, and the reconstructed VRML model is visually realistic.

The histogram distribution of the reprojection residual matrix (21) with outliers is shown in Fig.7. The residuals are largely conform to the assumption of normal distribution. As can be seen from the histogram, the outliers are obviously distinguished from inliers, the computed threshhold is shown in the figure. After rejecting outliers, the histogram distribution of the residuals produced by the final solutions is also shown in Fig.7. Obviously, the residual error is reduced significantly by the proposed approach. The final mean reprojection error given by the proposed approach is 0.597. In comparison, the reprojection errors by the algorithms of 'Huynh' and 'Ke' are 0.926 and 0.733, respectively. The proposed scheme outperforms other approaches.

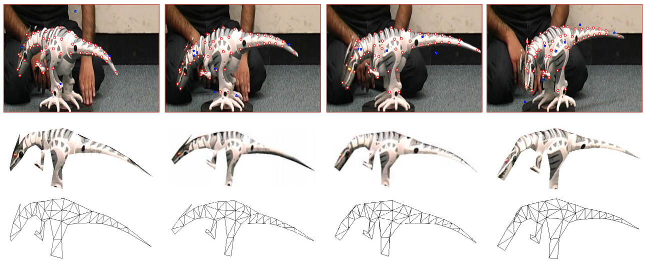
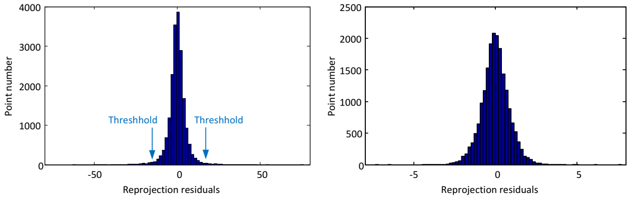


Figure 6. Test results of the dinosaur sequence. (top) Four frames from the sequence overlaid with the tracked features (red circles) and added outliers (blue stars); (middle) the corresponding 3D VRML models from different viewpoints; (bottom) the associated wireframes.



(a) (b)

Figure 7. The histogram distribution of the residual matrix of the dinosaur sequence. (a) Before outlier rejection; (b) after outlier rejection.

The second test was on a face sequence with different facial expressions. The sequence was downloaded from FGnet at http://www-prima.inrialpes.fr/FGnet/html/home.html, and 200 images from the sequence were used in the test. The image resolution is 720×576 with 68 automatically tracked feature points using the active appearance model (AAM). For test purpose, 8% outliers were added to the tracking data as shown in Fig.8.

The proposed robust algorithm was used to recover the Euclidean structure of the face. Fig.8 shows the reconstructed VRML models of four frames and the corresponding wireframes from different viewpoints. As demonstrated in the results, different facial expressions have been correctly recovered by the proposed approach. The reprojection errors obtained from 'Huynh', 'Ke', and the proposed algorithms are 0,697, 0.581, and 0.453, respectively.

VIII. CONCLUSION

The paper first proposed a rank-$(3k + 1)$ factorization algorithm which has been proved to be more accurate and more widely applicable than classic rank-$3k$ nonrigid factorization, especially in the case that the feature centroid could not be reliably recovered due to the presence of missing and outlying data. Then, an alternatively weighted factorization algorithm was presented to reduce the influence

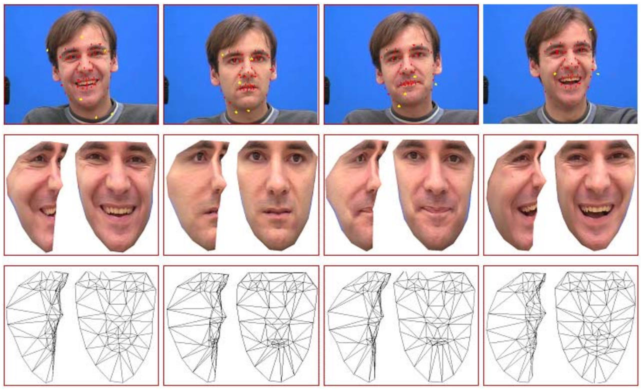


Figure 8. Test results of the face sequence. (top) Four frames from the sequence overlaid with the tracked features (red circles) and added outliers (blue stars); (middle) the corresponding 3D VRML models from different viewpoints; (bottom) the associated wireframes.

of large image noise. Finally, a robust factorization scheme was designed to deal with corrupted data containing outliers and missing points. The proposed technique requires no prior information of the error distribution in the tracking data. Extensive tests and evaluations demonstrated its advantages over previous methods.

IX. Acknowledgement

The work is supported in part by National Natural Science Foundation of China (NSFC) under grant no. 61273282 and Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] P. M. Q. Aguiar and J. M. F. Moura. Rank 1 weighted factorization for 3D structure recovery: Algorithms and performance analysis. *T-PAMI*, 25(9):1134–1049, 2003.

[2] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *IEEE T-PAMI*, 33(7):1442–1456, 2011.

[3] M. Brand. Morphable 3D models from video. In *Proc. of CVPR*, volume 2, pages 456–463, 2001.

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering nonrigid 3D shape from image streams. In *Proc. of CVPR*, vol. 2, pages 690–696, 2000.

[5] S. Christy and R. Horaud. Euclidean shape and motion from multiple perspective views by affine iterations. *IEEE T-PAMI*, 18(11):1098–1104, 1996.

[6] A. Eriksson and A. van den Hengel. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L1 norm. *CVPR*, 771–778, 2010.

[7] A. Gruber and Y. Weiss. Multibody factorization with uncertainty and missing data using the em algorithm. In *Proc. of CVPR*, pages 707–714, 2004.

[8] R. Hartley and F. Schaffalizky. Powerfactorization: 3D reconstruction with missing or uncertain data. *Australia-Japan Advanced Workshop on Computer Vision*, 2003

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[10] R. Hartley and R. Vidal. Perspective nonrigid shape and motion recovery. In *ECCV (1)*, pages 276–289. 2008.

[11] D. Q. Huynh, R. Hartley, and A. Heyden. Outlier correction in image sequences for the affine camera. In *Proc. ICCV*, pages 585–590, 2003.

[12] Q. Ke and T. Kanade. Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *CVPR*, pages 739–746, 2005.

[13] T. Okatani, T. Yoshida, and K. Deguchi. Efficient algorithm for low-rank matrix factorization with missing components and performance comparison of latest algorithms. In *ICCV*, pages 842–849, 2011.

[14] J. Oliensis and R. Hartley. Iterative extensions of the Sturm/Triggs algorithm: Convergence and nonconvergence. *IEEE T-PAMI*, 29(12):2217–2233, 2007.

[15] K. E. Ozden, K. Schindler, and L.Van Gool. Multibody structure-from-motion in practice. *IEEE T-PAMI*, 32(6):1134–1141, 2010.

[16] M. Paladini, *et al.*. Optimal metric projections for deformable and articulated structure-from-motion. *IJCV*, 96(2):252–276, 2012.

[17] V. Rabaud and S. Belongie. Re-thinking non-rigid structure from motion. In *CVPR*, 2008.

[18] D. Scaramuzza. 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *IJCV*, 95(1):74–85, 2011.

[19] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9(2):137–154, November 1992.

[20] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE T-PAMI*, 30(5):878–892, 2008.

[21] B. Triggs. Factorization methods for projective structure and motion. In *Proc. CVPR*, pages 845–851, 1996.

[22] G. Wang and J. Wu. Quasi-perspective projection model: Theory and application to structure and motion factorization from uncalibrated image sequences. *IJCV*, 87(3):213–234, 2010.

[23] G. Wang, J. Zelek and J. Wu. Structure and motion recovery based on spatial-and-temporal-weighted factorization. *IEEE T-CSVT*, 22(11): 1590–1603 2012.

[24] H. Wang, T.-J. Chin, and D. Suter. Simultaneously fitting and segmenting multiple-structure data with outliers. *IEEE T-PAMI*, 34(6):1177–1192, 2012.

[25] J. Yan and M. Pollefeys. A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video. *T-PAMI*, 30(5):865–877, 2008.

[26] J. Yu, T.-J. Chin, and D. Suter. A global optimization approach to robust multi-model fitting. In *CVPR*, pages 2041–2048, 2011.

[27] A. Zaharescu and R. Horaud. Robust factorization methods using a gaussian/uniform mixture model. *IJCV*, 81(3):240–258, 2009.

[28] L. Zelnik-Manor, M. Machline, and M. Irani. Multi-body factorization with uncertainty: Revisiting motion consistency. *IJCV*, 68(1):27–41, 2006.