IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 23, 2017, accepted September 18, 2017, date of publication September 21, 2017,
date of current version October 12, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2755019

Robust Structure and Motion Recovery Based on Augmented Factorization

GUANGHUI WANG, (Member, IEEE)
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA
National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing 100190, China

ghwang@ku.edu

This work was supported in part by the General Research Fund of the University of Kansas under Grant 2228901, and in part by the National Natural Science Foundation of China under Grant 61573351.

ABSTRACT This paper proposes a new strategy to promote the robustness of structure from motion algorithm from uncalibrated video sequences. First, an augmented affine factorization algorithm is formulated to circumvent the difficulty in image registration with noise and outliers contaminated data. Then, an alternative weighted factorization scheme is designed to handle the missing data and measurement uncertainties in the tracking matrix. Finally, a robust strategy for structure and motion recovery is proposed to deal with outliers and large measurement noise. This paper makes the following main contributions: 1) An augmented factorization algorithm is proposed to circumvent the difficult image registration problem of previous affine factorization, and the approach is applicable to both rigid and nonrigid scenarios; 2) by employing the fact that image reprojection residuals are largely proportional to the error magnitude in the tracking data, a simple outliers detection approach is proposed; and 3) a robust factorization strategy is developed based on the distribution of the reprojection residuals. Furthermore, the proposed approach can be easily extended to nonrigid scenarios. Experiments using synthetic and real image data demonstrate the robustness and efficiency of the proposed approach over previous algorithms.

INDEX TERMS Structure and motion factorization, robust factorization, alternative factorization, outlier detection, reprojection residual.

I. INTRODUCTION

Structure from motion (SfM) is the process to find the three-dimensional structure and camera motion from a set of uncalibrated 2D images. Classical method for 3D structure recovery is stereo vision [17], which is sensitive to image noise, since stereo vision only explores limited information from two or three images. Given a sequence of images, structure from motion is a powerful method to build a consistent 3D map with the knowledge of multiple-view geometry. Over the past two decades, tremendous progress in structure from motion has been made [10], [20], [22], [23], [26], [36], [38], [55], The results of this research have a wide range of potential applications, including robot navigation and obstacle avoidance, autonomous driving, video surveillance, and environment modeling.

Structure and motion factorization algorithm, pioneered by Tomasi and Kanade [39], is an effective approach for SfM. Given a set of tracked features across the sequence, the method decomposes image measurement directly into the 3D structure and the camera motion components through a bilinear formulation using singular value decomposition (SVD). By uniformly utilizing the data from all measurement, the algorithm achieves a more reliable result than stereo vision-based methods [29], [34], [41], [54].

A linear affine camera has been adopted by most research in SfM due to its simplicity [16]. It was extended to a more accurate nonlinear perspective camera model in [9] by incrementally factorizing a scaled measurement matrix. A full projective factorization algorithm was proposed by Triggs [41] iteratively using epipolar geometry between two adjacent image pairs. Inspired by this idea, different iterative strategies to recover the projective depths were designed by minimizing back-projection errors [44]. A complete analysis of these iterative methods was presented by Oliensis and Hartley [28]. Full perspective model based approach, though accurate, is computational intensive; as a trade-off of the efficiency and accuracy, a quasi-perspective model was proposed in [45].

By assuming deformation constraints that the nonrigid 3D shape can be represented by a span of rigid bases, the factorization algorithm was extended to handle nonrigid

deformation [7], where the shape bases, combination coefficients, and motion parameters are solved simultaneously from the SVD decomposition. This idea received a lot of attention and has been extensively studied in [3], [11], [31], and [40]. A manifold-learning framework was proposed in [33] to relax the deformation assumption. Agudo *et al.* [2] proposed a sequential nonrigid factorization approach. Yan and Pollefeys [50] developed a similar framework to recover the structure of articulated objects. In a dual trajectory space, Akhter *et al.* [5] suggested a duality solution to this problem based on basis trajectories.

Most factorization algorithms are based on the SVD decomposition of the tracking matrix composed by all tracking features tracked. In case of incomplete tracking data, however, the SVD-based approach is not applicable. Different alternative factorization approaches have been proposed to handle incomplete data, such as power factorization [15], alternative factorization [21], and factor analysis [14]. In practical application, the tracked features are usually corrupted by outliers or larger errors, in this case, most algorithms will degrade or even fail. The most popular strategies to handle outliers are random sample consensus (RANSAC) [13] and its variations, least median of squares (LMedS) [17], and other similar framework based on hypothesis-and-test [35]. Most of these methods are computational intensive. Recently, some robust structure and motion factorization algorithms have been proposed [1], [6], [31], [42], [52].

A scalar-weighted factorization scheme was proposed by Aguiar and Moura [4] through minimizing weighted square errors. The robustness to measurement uncertainties was enhanced in [14] using a factor analysis in an expectation maximization (EM) framework. Zelnik-Manor *et al.* [53] proposed temporal consistency for uncertain multi-body factorization. A Gaussian mixture model was introduced by Zaharescu and Horaud [52]. The same model was also adopted in [24] to approximate the noise distribution which was then estimated by a maximum likelihood algorithm. Ke and kanade [21] proposed to use L1 norm to increase robustness. The L1 norm and a damping factor were also introduced to the Wiberg algorithm in [12] and [27] to handle outliers. The outliers in the measurement were corrected using 'pseudo' observations in [18]. Bazin *et al.* [6] developed an optimal approach based on branch-and-bound. Other robust techniques were proposed based on quadratic formulation [51], kernel-scale [49], alternating bilinear algorithm [30], and spatial-and-temporal-weighted strategy [46].

In this paper, by exploring the reprojection residuals, we propose to handle the outlying data through a new viewpoint via the distribution of image reprojection residuals. The proposed approach is based on a new augmented factorization formulation, which circumvents the errors caused by image registration of contaminated tracking data. We also develop an alternative factorization algorithm to handle incomplete data and a weighted factorization scheme to incorporate measurement uncertainties. At last, we develop a robust factorization strategy for both rigid [48] and nonrigid [47] structure and motion recovery.

The remainder of this paper is organized as follows. Some background on affine factorization is offered in Section II. The augmented factorization algorithm is elaborated in Section III. Section IV presents the alternative factorization algorithm for incomplete data. An outlier detection scheme and the robust factorization algorithm are discussed in Section V. Section VI discusses the extension to nonrigid factorization. Extensive experimental results and comparisons are presented and analyzed in Sections VII and VIII, respectively. Finally, the paper is concluded in Section IX.

II. BACKGROUND ON AFFINE STRUCTURE FROM MOTION

To facilitate our discussion, some background on affine structure and motion factorization is presented in this section. Under affine projection model, the mapping process from 3D space to 2D image can be approximated by the following equation.

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

where $\mathbf{x}_{ij} = [u_{ij}, v_{ij}]^T$ is the image in frame i of a 3D space point $\mathbf{X}_j = [x_j, y_j, z_j]^T$; $\mathbf{A}_i$ is the projection matrix of the size 2×3; and $\mathbf{c}_i$ is a translation term between the space and the image frames. Suppose there are n space points, the projection of these points in the i-th image frame can be formulated as

$$[\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{in}] = \mathbf{A}_i[\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n] + \mathbf{C}_i \tag{2}$$

where all the translation vectors are grouped in the matrix $\mathbf{C}_i = [\mathbf{c}_i, \mathbf{c}_i, \cdots, \mathbf{c}_i]$, and $i = 1, \cdots, m$ is the frame number. Stacking the imaging equations of all m frames together, we can obtain the projection of an image sequence as follows.

$$\underbrace{\begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \cdots & \mathbf{x}_{mn} \end{bmatrix}}_{\mathbf{W}_{2m\times n}} = \underbrace{\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_m \end{bmatrix}}_{\mathbf{M}_{2m\times 3}} \underbrace{[\mathbf{X}_1, \quad \cdots, \quad \mathbf{X}_n]}_{\mathbf{S}_{3\times n}} + \underbrace{\begin{bmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_m \end{bmatrix}}_{\mathbf{C}_{2m\times n}} \tag{3}$$

where m is the frame number and n is the number of features. From the affine projection (1), we can see that the origin of the world system $\mathbf{X} = [0, 0, 0]^T$ is projected to $\mathbf{x}_i = \mathbf{c}_i$, which is the translation term of that frame. Assuming the world origin is aligned with the centroid of all space points, if we register the origin of the image system to the centroid of all imaged points, we have $\mathbf{c}_i = [0, 0]^T$, which means the translation term vanishes and $\mathbf{C} = \mathbf{0}$. As a result, the projection process (3) can be simplified to the following concise form after image centroid registration.

$$\mathbf{W}_{2m\times n} = \mathbf{M}_{2m\times 3}\mathbf{S}_{3\times n} \tag{4}$$

where the $2m \times n$ matrix $\mathbf{W}$ is composed of all tracked features, we call it tracking matrix hereafter. As an inverse problem of the image formulation, the problem of structure

from motion is to recover the shape matrix $\mathbf{S}$ and the motion matrix $\mathbf{M}$ from the tracking data $\mathbf{W}$ obtained across the image sequence.

From the right side of (3), we can see that the tracking matrix is at most 3, which is highly rank-deficient. For real tracking data, due to image noise and tracking errors, the rank of $\mathbf{W}$ is far more greater than 3. Therefore, we need to find a low rank approximation of $\mathbf{W}$ by techniques like SVD decomposition. From the rank-3 approximation of the tracking matrix, the shape matrix $\mathbf{S}$ and the motion matrix $\mathbf{M}$ can be easily decomposed. Obviously, such kind of decomposition is not unique as one can always insert a nonsingular transformation matrix $\mathbf{H} \in \mathbb{R}^{3\times 3}$ as

$$\mathbf{W} = (\mathbf{MH})(\mathbf{H}^{-1}\mathbf{S}) \tag{5}$$

Since we are interested in the metric structure and motion parameters, to this end, we need to find a metric transformation matrix $\mathbf{H}$ to upgrade the structure to the Euclidean space. Certain metric constraints are employed for this purpose $\mathbf{H}$ [32], [45], once an Euclidean upgrading matrix $\mathbf{H}$ is available, the metric structure can be recovered from $\mathbf{H}^{-1}\mathbf{S}$, and the corresponding camera motions can be obtained from $\mathbf{MH}$.

III. AUGMENTED STRUCTURE AND MOTION FACTORIZATION

Previous studies on affine factorization of rigid objects are based on the formulation of (4) due to its simplicity. One necessary condition of the rank-3 affine factorization is that all imaged points should be centroid registered. However, when some tracked features are missing, or contaminated with outliers and significant noise, the centroid of image points could not be computed reliably, the error in the registration, as shown in the experiments, will result in significant deviation to the final result. This issue was overlooked by previous studies. We propose an augmented factorization algorithm to circumvent this registration problem.

A. AUGMENTED AFFINE FACTORIZATION

By adopting homogeneous representation, we can rewrite the affine projection equation (1) as below.

$$\mathbf{x}_{ij} = [\mathbf{A}_i|\mathbf{c}_i]\,\tilde{\mathbf{X}}_j \tag{6}$$

where the space point $\mathbf{X}_j$ is denoted using homogeneous coordinates as $\tilde{\mathbf{X}}_j = [\mathbf{X}_j^T, t_j]^T$. Then, the imaging process of the entire sequence can be written as

$$\underbrace{\begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \cdots & \mathbf{x}_{mn} \end{bmatrix}}_{\mathbf{W}_{2m\times n}} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & | & \mathbf{c}_1 \\ \vdots & | & \vdots \\ \mathbf{A}_m & | & \mathbf{c}_m \end{bmatrix}}_{\mathbf{M}_{2m\times 4}} \times \underbrace{\begin{bmatrix} \tilde{\mathbf{X}}_1, & \cdots, & \tilde{\mathbf{X}}_n \end{bmatrix}}_{\mathbf{S}_{4\times n}} \tag{7}$$

which can be written in a concise form as

$$\mathbf{W}_{2m\times n} = \mathbf{M}_{2m\times 4}\mathbf{S}_{4\times n} \tag{8}$$

Compared to the rank-3 factorization, the motion matrix in (7) is augmented by an extra column, while the shape matrix is augmented by an extra row. As a result, the rank of the tracking matrix becomes four, instead for three for the data after registration. We call the formulation (7) augmented factorization.

It is obvious that the expression (7) is derived directly from the affine projection model (1), which does not require image registration with respect to the centroid. Therefore, it is applicable to corrupted data with significant noise, missing entries, and outlying points. Both factorization algorithms (4) and (8) can be equivalently written as the following minimization problem.

$$f(\mathbf{M}, \mathbf{S}) = \underset{\mathbf{M},\mathbf{S}}{arg\,min}\ \|\mathbf{W} - \mathbf{MS}\|_F^2 \tag{9}$$

The major difference between (4) and (8) lies in the rank constraints. As a result, the corresponding residual errors are

$$E_3 = \sum_{i=4}^{N} \sigma_i^2, \quad E_4 = \sum_{i=5}^{N} \sigma_i^2, \tag{10}$$

respectively, where σ_i are the singular values of $\mathbf{W}$ in descending order, and $N = \min(2m, n)$ denotes the number of singular values. It is obvious that the error between the two algorithm is σ_4^2 when the tracking data is properly normalized. If all imaged points are noise free and registered accurately to the corresponding centroid, the last column of $\mathbf{M}$ in (7) will vanish because $\mathbf{c}_i = \mathbf{0}$, and the augmented expression (8) is equivalent to the rank-3 factorization (4) . Thus, (4) is a special case of the augmented factorization after registration to the centroid. Nonetheless, in case of noise and outlier corrupted data or there are missing feature, we cannot accurately recover the image centroid, the rank-3 algorithm will produce a large error since σ_4 is not close to zero.

Suppose the rank-4 decomposition of (7) yields a set of solutions $\hat{\mathbf{M}}_{m\times 4}\hat{\mathbf{S}}_{4\times n}$. Similar to rank-3 factorization, the decomposition is defined up to a nonsingular transformation as $\mathbf{MS} = (\hat{\mathbf{M}}\mathbf{H})(\mathbf{H}^{-1}\hat{\mathbf{S}})$. In the following, we will discuss how to recovery of the Euclidean upgrading matrix.

B. EUCLIDEAN UPGRADING MATRIX

The upgrading matrix $\mathbf{H}$ is a 4×4 nonsingular matrix which can be denoted as the following form.

$$\mathbf{H} = [\mathbf{H}_{1:3}|\mathbf{h}_4] \tag{11}$$

where $\mathbf{H}_{1:3}$ and $\mathbf{h}_4$ stand for the first three and the last columns of $\mathbf{H}$. Let us denote the i-th two-row of $\hat{\mathbf{M}}$ as $\hat{\mathbf{M}}_i$, which corresponds to the motion of the i-th camera, after upgrading, the metric motion matrix becomes

$$\mathbf{M}_i = \hat{\mathbf{M}}_i\mathbf{H} = \left[\hat{\mathbf{M}}_i\mathbf{H}_{1:3}|\hat{\mathbf{M}}_i\mathbf{h}_4\right] = [\mathbf{A}_i|\mathbf{c}_i] \tag{12}$$

By assuming a simplified camera model with only one parameter, i.e., the focal length f_i, the left 2×3 submatrix of $\mathbf{M}_i$ in (12) can be written as

$$\mathbf{A}_i = \hat{\mathbf{M}}_i\mathbf{H}_{1:3} = f_i \begin{bmatrix} \mathbf{r}_{i1}^T \\ \mathbf{r}_{i2}^T \end{bmatrix} \tag{13}$$

from the camera imaging process [15], where $\mathbf{r}_{i1}^T$ and $\mathbf{r}_{i2}^T$ are the first two rows of the camera's rotation matrix. Let $\mathbf{Q} = \mathbf{H}_{1:3}\mathbf{H}_{1:3}^T$, then, $\mathbf{Q}$ is constrained from (13) as

$$\hat{\mathbf{M}}_i\mathbf{Q}\hat{\mathbf{M}}_i^T = \left(\hat{\mathbf{M}}_i\mathbf{H}_{1:3}\right)\left(\hat{\mathbf{M}}_i\mathbf{H}_{1:3}\right)^T = f_i^2\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \quad (14)$$

The above equation provides two independent constraints to $\mathbf{Q}$, which is a 4×4 positive semidefinite symmetric matrix. The matrix $\mathbf{Q}$ is homogeneous and has nine degree-of-freedom, thus, a least squares solution can be obtained from five or more images. Then, according to our previous study [45], the submatrix $\mathbf{H}_{1:3}$ can be recovered from the matrix $\mathbf{Q}$ via extended Cholesky decomposition.

After recovering $\mathbf{H}_{1:3}$, the last column of the upgrading matrix is then determined straightforwardly. From the expression (12), the projection equation (6) can be written as

$$\mathbf{x}_{ij} = \hat{\mathbf{M}}_i\mathbf{H}_{1:3}\mathbf{X}_j + \hat{\mathbf{M}}_i\mathbf{h}_4 \quad (15)$$

It can be easily proved from (15) that the last column $\mathbf{h}_4$ corresponds to the translation term between the world system and the image frame for noise free case with general motion. For any given world system, the values of $\mathbf{h}_4$ make no influence to the metric structure of the reconstructed object. Therefore, we can choose any 4-vector for $\mathbf{h}_4$, as long as it is independent of the columns of $\mathbf{H}_{1:3}$, since the upgrading matrix should be nonsingular. In practice, we construct $\mathbf{h}_4$ as below.

From the SVD decomposition of $\mathbf{H}_{1:3}$

$$\begin{aligned}\mathbf{H}_{1:3} &= \mathbf{U}_{4\times4}\Sigma_{4\times3}\mathbf{V}_{3\times3}^T \\ &= [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4]\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & \sigma_3\\ 0 & 0 & 0\end{bmatrix}[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]^T \end{aligned} \quad (16)$$

where $\mathbf{U}$ and $\mathbf{V}$ are two orthonormal matrices, and Σ is a diagonal matrix of the singular values of $\mathbf{H}_{1:3}$. Then, $\mathbf{h}_4$ can be simply set as

$$\mathbf{h}_4 = \kappa_4\mathbf{u}_4 \quad (17)$$

where $\mathbf{u}_4$ is the last column of $\mathbf{U}$, with κ_4 an arbitrary scalar between the largest and the smallest singular values σ_1 and σ_3. The construction guarantees a good numerical stability in computing the inverse of $\mathbf{H}$, since the constructed matrix $\mathbf{H}$ has the same condition number as $\mathbf{H}_{1:3}$.

C. ALGORITHM OF THE AUGMENTED AFFINE FACTORIZATION

The above proposed augmented rank-4 affine factorization algorithm is summarized in Algorithm 1.

Algorithm 1 Augmented Affine Factorization

Input: Tracking data $\mathbf{W}$
1. Perform SVD decomposition of the tracking matrix;
2. Recover a set of rank-4 solutions of $\hat{\mathbf{M}}$ and $\hat{\mathbf{S}}$;
3. Estimate the metric transformation matrix $\mathbf{H}$;
4. Upgrade the result to metric space as $\mathbf{H}^{-1}\hat{\mathbf{S}}$ and $\hat{\mathbf{M}}\mathbf{H}$.

Output: Euclidean structure and camera motion

IV. ALTERNATIVE FACTORIZATION SCHEME

SVD decomposition is a convenient technique for structure and motion factorization, however, SVD only works when the tracking matrix is complete. In practice, missing data are inevitable since some features may get lost during the process of tracking due to occlusion or other factors. To this end, a two-step alternative factorization scheme is developed to handle incomplete data and image uncertainties.

A. ALTERNATIVE FACTORIZATION ALGORITHM

The essence of the structure from motion algorithm (7) is equivalent to finding a set of rank-4 solutions $\mathbf{M}$ and $\mathbf{S}$ by minimizing the following Frobenious norm.

$$\begin{aligned}&\underset{\mathbf{M},\mathbf{S}}{arg\,min}\ \|\mathbf{W} - \mathbf{MS}\|_F^2 \\ &s.t.\ \mathbf{M} \in \mathbb{R}^{2m\times4}, \mathbf{S} \in \mathbb{R}^{4\times n}\end{aligned} \quad (18)$$

To solve the problem (18), $\mathbf{S}$ and $\mathbf{M}$ can be factorized simultaneously using SVD decomposition. Alternatively, we can fix either the shape matrix or the motion matrix, and iteratively solve the other one as below.

$$f(\mathbf{S}) = \underset{\mathbf{S}}{arg\,min}\ \|\mathbf{W} - \mathbf{MS}\|_F^2 \quad (19)$$

$$f(\mathbf{M}) = \underset{\mathbf{M}}{arg\,min}\ \|\mathbf{W} - \mathbf{MS}\|_F^2 \quad (20)$$

The above algorithm is called Power Factorization [15] or alternative factorization. It can be verified that each of the above two cost functions is convex by itself, and it converges very fast even with random initial values. The idea has been adopted by several researches [15], [45].

During iteration, each step can be solved via least squares. Let us take the cost functions (19) as an example and rewrite it with respect to each feature.

$$f(\mathbf{s}_j) = \underset{\mathbf{s}_j}{arg\,min}\ \|\mathbf{w}_j - \mathbf{M}\mathbf{s}_j\|_F^2 \quad (21)$$

where $\mathbf{w}_j$ denotes the j-th column of $\mathbf{W}$, and $\mathbf{s}_j$ stands for the j-th column of $\mathbf{S}$. Thus, each column of $\mathbf{S}$ can be solved in closed form via least squares.

$$\mathbf{s}_j = \mathbf{M}^\dagger\mathbf{w}_j = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{w}_j, \quad j = 1, \cdots, n \quad (22)$$

where $\mathbf{M}^\dagger$ denotes the Moore-Penrose pseudoinverse. The least squares solution can naturally handle the missed features in tracking. For example, if some entries in $\mathbf{w}_j$ is missing, we can simply set those entries in $\mathbf{w}_j$ to zeros, then, $\mathbf{s}_j$ can still be solved from (22) in the least squares sense.

Similarly, the second cost function (20) can be rewritten with respect to each frame as

$$f(\mathbf{m}_i^T) = \underset{\mathbf{m}_i^T}{arg\,min}\ \|\mathbf{w}_i^T - \mathbf{m}_i^T\mathbf{S}\|_F^2 \quad (23)$$

Algorithm 2 Alternative Factorization Algorithm

Input: Tracking data $\mathbf{W}$
while *not convergence* **do**
  1. update the shape matrix via (22)
  2. update the motion matrix via (24)
end
Output: The shape matrix and the motion matrix

which yields the following least-square solution of the motion matrix $\mathbf{M}$.

$$\mathbf{m}_i^T = \mathbf{w}_i^T \mathbf{S}^\dagger = \mathbf{w}_i^T \mathbf{S}^T (\mathbf{S}\mathbf{S}^T)^{-1}, \quad i = 1, \cdots, m \tag{24}$$

where $\mathbf{m}_i^T$ and $\mathbf{w}_i^T$ respectively stand for the i-th row of $\mathbf{M}$ and $\mathbf{W}$, and $\mathbf{S}^\dagger$ denotes the pseudoinverse. Like in (22), we can set the missed entries in $\mathbf{w}_i^T$ to zeros.

The above alternative algorithm is summarized in Algorithm 2 with random initialization.

B. WEIGHTED FACTORIZATION

Feature tracking is a hard problem and tracking errors are inevitable in practice. If prior knowledge about distribution of the errors is available, all elements of the approximation error can be weighted by taking account of the error distribution. The basic idea is to give each image measurement a weight according to its uncertainty. Reliable features are assigned higher weights while unreliable features receive lower weights. The weighted factorization is formulated as follows.

$$\begin{aligned} &\underset{\mathbf{M},\mathbf{S}}{arg\,min} \ \|\Sigma \otimes (\mathbf{W} - \mathbf{M}\mathbf{S})\|_F^2 \\ &s.t.\ \mathbf{M} \in \mathbb{R}^{2m\times 4}, \mathbf{S} \in \mathbb{R}^{4\times n} \end{aligned} \tag{25}$$

where $'\otimes'$ stands for the Hadamard product, which is element-wise array product; $\Sigma = \{\sigma_{ij}\}$ denotes the uncertainty matrix composed by the weights of all features derived from the measurement confidence.

The general weighted factorization could not be solved analytically using the singular value decomposition. In our study, we adopt an alternative scheme, similar to [4], [19], [52], to solve $\mathbf{S}$ and $\mathbf{M}$ alternatively as follows.

$$f(\mathbf{S}) = \underset{\mathbf{s}_j}{arg\,min} \ \|\Sigma_j \otimes (\mathbf{w}_j - \mathbf{M}\mathbf{s}_j)\|_F^2 \tag{26}$$

$$f(\mathbf{M}) = \underset{\mathbf{m}_i^T}{arg\,min} \ \|\Sigma_i^T \otimes (\mathbf{w}_i^T - \mathbf{m}_i^T \mathbf{S})\|_F^2 \tag{27}$$

where Σ_j denotes the j-th column of Σ and Σ_i^T represents the i-th row. Then, a close-form solution of can be obtained in the sense of least squares.

$$\mathbf{s}_j = \left(\text{diag}(\Sigma_j)\mathbf{M}\right)^\dagger \left(\text{diag}(\Sigma_j)\mathbf{w}_j\right), \quad j = 1, \cdots, n \tag{28}$$

$$\mathbf{m}_i^T = \left(\mathbf{w}_i^T \text{diag}(\Sigma_i^T)\right)\left(\mathbf{S}\text{diag}(\Sigma_i^T)\right)^\dagger, \quad i = 1, \cdots, m \tag{29}$$

where $(\bullet)^\dagger$ denotes the pseudoinverse of a matrix, and $'\text{diag}(\bullet)'$ denotes the diagonal matrix formed from a vector. Equations (28) and (29) yield the least-square solutions of $\mathbf{S}$ and $\mathbf{M}$. Same as the alternative factorization, when there are some missing elements in the tracking matrix, one can simply set those entries in $\mathbf{w}_j$ to zeros.

The alternative weighted factorization algorithm is summarized in Algorithm 3, where the motion matrix $\mathbf{M}$ can be initialized randomly or using previous estimation, while the initial value of the weight matrix Σ, as will be discussed in next section, is estimated from the reprojection residuals.

Algorithm 3 Alternative Weighted Factorization

Input: Matrices $\mathbf{W}$, $\mathbf{M}$, and Σ
while *not convergence* **do**
  1. update the shape matrix via (28)
  2. update the motion matrix via (29)
end
Output: The shape and motion matrices

V. ROBUST FACTORIZATION STRATEGY

Based on the augmented factorization scheme proposed in the foregoing sections. An efficient strategy for outlier detection is proposed below.

A. OUTLIER REJECTION

Outliers are inevitable in practice. The most popular strategy in the computer vision field is based on the hypothesis-and-test scheme [8], [35], which are computationally intensive. We will investigate the problem from a new viewpoint via the distribution of image reprojection residuals.

Both the SVD-based and the alternative factorization-based algorithms yield a set of least-square solutions, which are achieved by minimizing the sum of the squared errors between the fitted values and the observation. Extensive experiments show that the least-square algorithms usually yield reasonable solutions, however, the residuals yield from outliers are outstandingly larger than the errors from inliers.

Suppose $\hat{\mathbf{M}}$ and $\hat{\mathbf{S}}$ are the solutions of the rank-4 factorization of a tracking data $\mathbf{W}$, by reprojecting the set of solutions back to the sequence, we can obtain the reprojection residuals, which can be organized as a matrix.

$$\mathbf{E} = \mathbf{W} - \hat{\mathbf{M}}\hat{\mathbf{S}} = \begin{bmatrix} \mathbf{e}_{11} & \cdots & \mathbf{e}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{e}_{m1} & \cdots & \mathbf{e}_{mn} \end{bmatrix}_{2m\times n} \tag{30}$$

where

$$\mathbf{e}_{ij} = \mathbf{x}_{ij} - \hat{\mathbf{M}}_i\hat{\mathbf{s}}_j = \begin{bmatrix} \Delta u_{ij} \\ \Delta v_{ij} \end{bmatrix} \tag{31}$$

is the residual of the point (i, j). The reprojection error can be defined by the Euclidean distance $\|\mathbf{e}_{ij}\|$ of the image point and its reprojection, thus, the reprojection error of the entire can be defined by the following error matrix

$$\mathbf{Err} = \begin{bmatrix} \|\mathbf{e}_{11}\| & \cdots & \|\mathbf{e}_{1n}\| \\ \vdots & \ddots & \vdots \\ \|\mathbf{e}_{m1}\| & \cdots & \|\mathbf{e}_{mn}\| \end{bmatrix}_{m\times n} \tag{32}$$

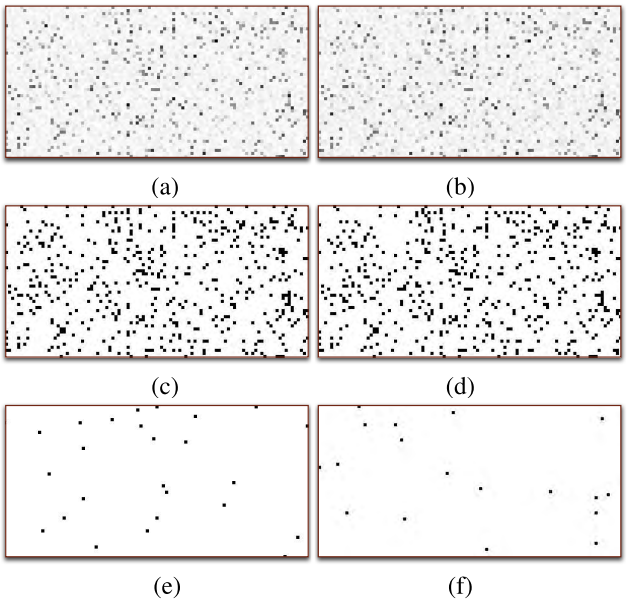

FIGURE 1. (a) The absolute values of the added noise and outliers in the tracking matrix, where the gray level of each pixel corresponds the normalized error magnitude at that point; (b) the distribution of the normalized reprojection errors (32); (c) the distribution of the added outlying data; (d) the outliers segmented from reprojection error by a single threshold; (e) the distribution of false positive error given by the thresholding; and (f) the false negative error given by the thresholding.

Fig. 1 shows the distribution of the matrix (32), where 40 images are generated from 100 random 3D space points via affine projection. The image resolution is 800 × 800, and the images are corrupted by Gaussian noise and 10% outliers. The added noise level is 3 pixels, and the outliers are simulated by random noise whose level (standard deviation of the noise) is set at 15 pixels. The real added noise and outliers are illustrated by an image as shown in Fig. 1(a), where the grayscale of each pixel corresponds to the inverse magnitude of the error on that point, the darker the pixel, the larger the error magnitude on that point. The distribution of the real added outliers is depicted as a binary image in Fig. 1(c), which correspond to the darker points in Fig. 1(a).

Using the corrupted data, a set of motion and shape matrices were estimated by employing the rank-4 factorization algorithm and the error matrix was then computed. The distribution of the reprojection error (32) is illustrated in Fig. 1(b) with each pixel corresponds to the reprojection error of that point. It is evident that the reprojection error and the real added noise have similar distribution. The points with large reprojection errors correspond to those with large noise levels. Fig. 1(d) shows the binary image of Fig. 1(b) by simply applying a global threshold to the reprojected residuals.

It is obvious from the above example that almost all outliers are successfully segmented by a single threshold. The distribution of false positive error (the inlier points being classified as outliers by the given global threshold) and the false negative error (the outliers not being detected by the thresholding) are given in Fig. 1(e) and (f), respectively. The false positive error is mainly caused by those inliers with large noise (which should be treated as outliers), while the false negative error is caused by the outliers with small deviations (which can actually be treated as inliers), these two types errors are related to the threshold, however, they will not make a big influence to the final solutions.

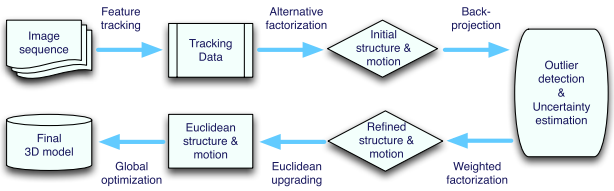

FIGURE 2. The outline of the robust structure from motion strategy.

Inspired by this observation, an intuitive outlier detection and robust factorization scheme is proposed. The flowchart of the strategy is shown in Fig. 2, and the computation details is given in Algorithm 4.

Algorithm 4 Robust Structure and Motion Factorization

Input: Tracking matrix $\mathbf{W}$

1. Normalize the tracking matrix point-wisely and image-wisely, as in [37], to improve the numerical stability;
2. Perform augmented affine factorization on the tracking matrix to obtain a set of solutions of $\hat{\mathbf{M}}$ and $\hat{\mathbf{S}}$;
3. Estimate the reprojection residuals and determine a global threshold to remove the outliers;
4. Eliminate the outliers and recalculate the matrices $\hat{\mathbf{M}}$ and $\hat{\mathbf{S}}$ using the remaining inliers via Algorithm 2;
5. Estimate the uncertainty of each inlying feature from the distribution of the reprojection residuals;
6. Refine the solutions by weighted factorization Algorithm 3;
7. Recover the metric upgrading matrix $\mathbf{H}$ and upgrade the solutions to the metric space;
8. Perform a global optimization via bundle adjustment.

Output: 3D metric structure and camera motion parameters recovered from $\mathbf{S}$ and $\mathbf{M}$, respectively

In Algorithm 4, steps 3 and 4 can be repeated for one more time to ensure a more refined inlying data and solutions. In practice, however, the repetition does not make much difference to the final results. During computation, the Algorithms 2 and 3 are employed to handle the missing data and measurement uncertainties. Concerning the initialization of the alternative algorithm, since an initial set of solutions have been obtained in the previous steps, these solutions can be used as initial values in the iteration so as to speed up the convergence of the algorithm.

B. PARAMETER ESTIMATION

Suppose the image noise is modeled by Gaussian distribution, it can be verified that the reprojection residuals (30) also follow the same distribution as the image noise, while the reprojection errors (32) follow χ^2 distribution with codimension 2. It is nature to assume that the noise at both coordinate directions in the image is independent and

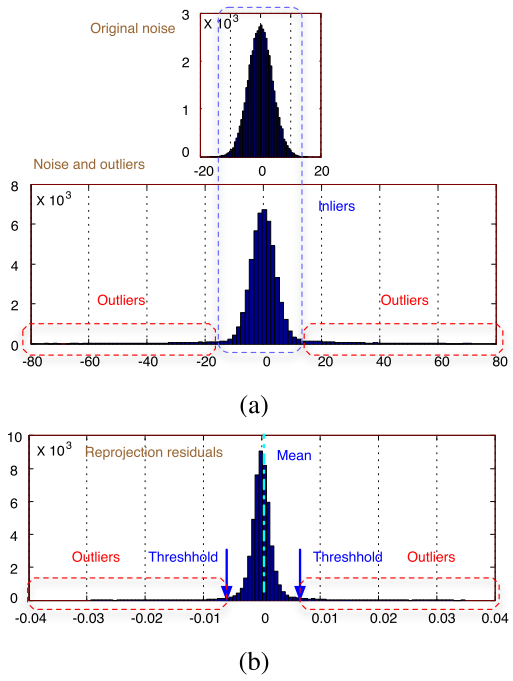

FIGURE 3. **(a) The histogram of the noise added to the sequence (upper), and the noise plus outliers (lower); and (b) the distribution of the residuals.**

identically distributed (i.i.d.). Let $\mathcal{V}(\mathbf{E})$ denotes the vector derived from the matrix $\mathbf{E}$, then, $\mathcal{V}(\mathbf{E})$ should also be Gaussian. For example, we add Gaussian noise and outliers to a tracking matrix, as shown in Fig. 3(a), then, we compute a set of solutions using the proposed augmented factorization and calculate the reprojection residuals, whose distribution is also Gaussian, as shown in Fig. 3(b).

Suppose the mean and standard deviation (STD) of $\mathcal{V}(\mathbf{E})$ are μ and σ respectively, by registering the residual vector $\mathcal{V}(\mathbf{E})$ with respect to its mean μ, we can determine the outlier threshold as below.

$$\theta = \kappa\,\sigma \tag{33}$$

where κ is a parameter, which is set at 4.0 in our experiment (the result is not sensitive to this value). The points whose reprojection errors are greater than θ after registration are classified as outliers, i.e.,

$$\text{outliers} = \left\{\mathbf{x}_{i,j} \middle| \left((\Delta u_{ij} - \mu)^2 + (\Delta v_{ij} - \mu)^2\right)^{\frac{1}{2}} > \theta\right\} \tag{34}$$

Since the residual vector $\mathcal{V}(\mathbf{E})$ contains outliers, which have significant influence to the estimation of the mean and STD since the outliers will make an extreme deviation of the result. In this study, we estimate the standard deviation using the median absolute deviation (MAD) as

$$\sigma = 1.4826\ \text{median}(|\mathcal{V}(\mathbf{E}) - \text{median}(\mathcal{V}(\mathbf{E}))|) \tag{35}$$

which is proved to be robust to outliers. The mean is calculated from the features that are smaller than the median of the residuals.

$$\mu = \text{mean}\{\mathcal{V}'(\mathbf{E}) | \mathcal{V}'(\mathbf{E}) < \text{median}(|\mathcal{V}(\mathbf{E})|)\} \tag{36}$$

The above computation usually yields a more reasonable estimation of the STD and the mean.

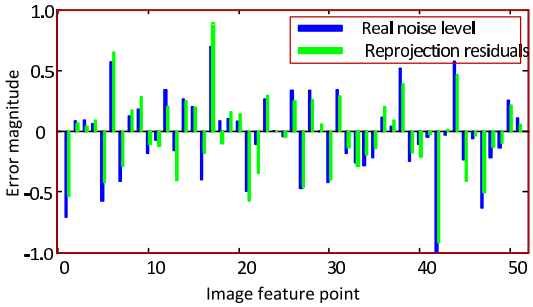

FIGURE 4. **The distribution of the added noise and the calculated reprojection residuals of 50 feature along the *u*-axis in one image.**

In previous study, the weights in weighted factorization are usually obtained from the uncertainty of the features, either isotropically [4], [39] or directionally [19], [25]. However, the uncertainty is usually hard to estimate or unavailable in practice. Through extensive experiments, we proved that the uncertainty is in general proportional to the reprojection residuals [46]. For example, from the structure and motion matrices computed at step 5, the residuals of inlying data can be estimated. As depicted in Fig. 4, the distribution of the residuals is largely close to that of the added noise. Therefore, we are suggested to estimate the weights from the reprojection residuals after outlier removal. The features with higher residual values have larger uncertainties, and thus, lower weights are assigned. We treat each coordinate direction independently and estimate the weight via the following equation.

$$\omega_{ij} = \frac{1}{\mathcal{N}} \exp\left(-\frac{e_{ij}^2}{2\sigma^2}\right) \tag{37}$$

where the weights are assigned in a shape of Gaussian, e_{ij} denotes the (i, j)-th entry of the residual (30), the σ is estimated from the MAD (35), and $\mathcal{N}$ is a scalar used for normalization. For the missed points and outliers, the corresponding weights are set as $\omega_{ij} = 0$.

VI. EXTENSION TO NONRIGID FACTORIZATION

In the preceding discussion, we assume the scene is globally rigid or static. In case of nonrigid scenarios, we follow Bregler's assumption which models the nonrigid structure using a linear model composed of a set of shape bases $\mathbf{B}_l$ [7], i.e.,

$$\mathbf{S}_i = \sum\nolimits_{l=1}^{k} \omega_{il}\mathbf{B}_l \tag{38}$$

where ω_{il} is the combination weight; and k is the number of bases. Based on the assumption (38), the projection of image i can be modeled as

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

Similar to rigid factorization, if we register all image points to the associated centroid in each frame and adopt relative image coordinates, the translation $\mathbf{c}_i = \mathbf{0}$. Consequently, the nonrigid structure and motion factorization can be modeled as

$$\underbrace{\begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \cdots & \mathbf{x}_{mn} \end{bmatrix}}_{\mathbf{W}_{2m\times n}} = \underbrace{\begin{bmatrix} \omega_{11}\mathbf{A}_1 & \cdots & \omega_{1k}\mathbf{A}_1 \\ \vdots & \ddots & \vdots \\ \omega_{m1}\mathbf{A}_m & \cdots & \omega_{mk}\mathbf{A}_m \end{bmatrix}}_{\mathbf{M}_{2m\times 3k}} \underbrace{\begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix}}_{\mathbf{B}_{3k\times n}} \tag{39}$$

It is obvious from the right side of (39) that the rank of the tracking matrix is $3k$, which is the basic assumption of previous work on nonrigid factorization. Please note that the expression (39) is obtained based on the assumption of image registration with respect to the centroid of feature measurement. However, it is impossible to recover the centroid when there are outliers and/or missing features in the measurement. Similar to the rigid case, we can employ an augmented formulation like (7) to circumvent the registration issue. By adopting homogeneous expression as (6), the above nonrigid factorization can be expressed in the following augmented form.

$$\underbrace{\begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \cdots & \mathbf{x}_{mn} \end{bmatrix}}_{\mathbf{W}_{2m\times n}} = \underbrace{\begin{bmatrix} \omega_{11}\mathbf{A}_1 & \cdots & \omega_{1k}\mathbf{A}_1 & \mathbf{c}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \omega_{m1}\mathbf{A}_m & \cdots & \omega_{mk}\mathbf{A}_m & \mathbf{c}_m \end{bmatrix}}_{\mathbf{M}_{2m\times(3k+1)}} \underbrace{\begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \\ \mathbf{t}_i^T \end{bmatrix}}_{\mathbf{B}_{(3k+1)\times n}} \tag{40}$$

where the motion matrix is augmented by one additional column, while the shape matrix is augmented by one row, which can be taken as a homogeneous term. The above equation can be written in short as $\mathbf{W}_{2m\times n} = \mathbf{M}_{2m\times(3k+1)}\mathbf{B}_{(3k+1)\times n}$. Different to the expression of (39), the rank of the tracking matrix is $3k + 1$, instead of $3k$, in this case. Given the tracking data, the structure bases and motion matrix can be easily obtained via SVD decomposition or the alternative factorization with the rank constraint. Since the expression (40) does not depend on data registration, it can naturally handle outliers and missed features.

Based on the new formulation, the preceding proposed augmented factorization, alternative factorization, and weighted factorization algorithms can be directly extended to the nonrigid scenario. The only difference here with respect to the rigid case lies in the rank constraint applied to the tracking matrix. Thus, a set of motion and structure matrices can be easily decomposed as shown in (40). Obviously, the decomposition is not unique and we need to find a metric upgrading matrix $\mathbf{H} \in \mathbb{R}^{(3k+1)\times(3k+1)}$ to upgrade the solution to the metric space. Then, the nonrigid structure and camera motion parameters can be factorized from $\mathbf{M}$ and $\mathbf{S}$, respectively. A detailed discussion about this approach can be found in our early study [47].

VII. EVALUATIONS ON SYNTHETIC IMAGES

The proposed algorithm was validated and evaluated extensively using simulated image data. During the simulation, we generated 100 random space points within a cube of $40\times40\times40$. Then, these 3D points are projected to a sequence of 50 frames using affine camera models. The image and camera parameters are set as follows: The image size is set as 800×800 pixel; the camera is set at 600 to the object with the center varying randomly within ±40 in each direction; the focal lengths are chosen randomly between 500 to 550; and the rotations are set randomly within $\pm60°$.

A. INFLUENCE OF REGISTRATION

We compared the proposed augmented factorization scheme with its rank-3 counterpart with respect to various image centroid displacements. During the test, different levels of Gaussian noise was added to every simulated image features. To simulate the scenario that the centroid cannot be reliably estimated due to outliers and missing features, we deviate the centroid of image features with certain amount of displacement, and register all image points with respect to the deviated centroid.

Using the corrupted data, we recover the motion and shape matrices using the SVD factorization with the rank-4 and the rank-3 constraints, respectively; then, reproject the solution back onto the images and calculate the reprojection residuals. In order to evaluate the performance of different algorithms, we calculate the difference between the ground truth and the corresponding back-projected features, we call the average of all these errors as the mean reprojection variance, which is defined as follows.

$$E_{rv} = \frac{1}{mn}\|\mathbf{W}_0 - \hat{\mathbf{M}}\hat{\mathbf{S}}\|_F^2 \tag{41}$$

where $\mathbf{W}_0$ is the tracking matrix without noise; and $\hat{\mathbf{S}}$ and $\hat{\mathbf{M}}$ are the estimated structure and motion matrices, respectively. In order to obtain a statistically meaningful result, we perform 100 independent tests at each noise level. Fig. 5 shows the mean reprojection variance with respect to different centroid displacements and noise levels, which are defined as the STD of the Gaussian noise.

As shown in Fig. 5, it is evident that the misaligned centroid has no influence to the proposed augmented factorization, however, the error of centroid has a significant impact to the performance of the rank-3 based approach. As demonstrated in the experiment, the error caused by the centroid greatly outperforms that by the image noise. Therefore, the augmented affine factorization is a wise choice in practice, especially in the presence of outliers, missing points, and/or large measurement errors.

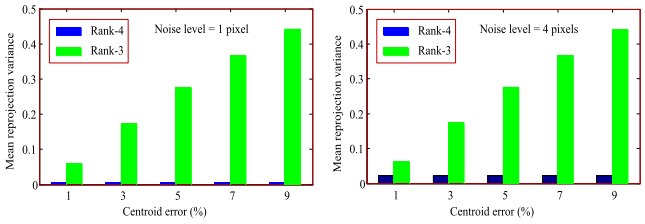

FIGURE 5. The mean reprojection variance with respect to different centroid displacements at the noise level of (a) 1 pixel and (b) 4 pixels.

B. PERFORMANCE EVALUATION

In this test, we evaluated and compared the performance of the proposed approach with respect to other robust factorization algorithms in terms of accuracy and computational complexity.

We add Gaussian noise to the above generated image sequence and vary the noise level from 1 to 5 pixels in steps of 1 pixel. In the mean time, 5% and 20% outliers were added to the tracking matrix, respectively. Using the corrupted data, we recover the motion and shape matrices using the propose technique. As a comparison, three competing approaches were implemented as well. The first one is based on an outlier correction scheme [18], the second one is based on $L1$ minimization [21], and the last one is based on mixture of Gaussian model [24].

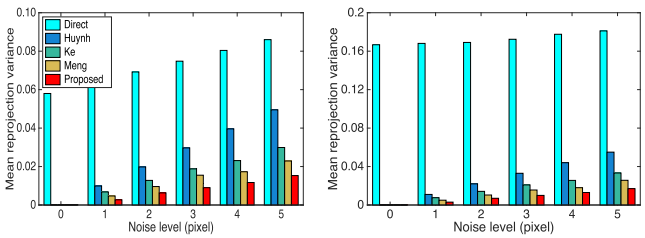

FIGURE 6. The mean reprojection variance at different noise levels with (a) 5% outliers and (b) 20% outliers.

Fig. 6 shows the mean reprojection variance at different noise levels and outliers ratios, where 'Direct' stands for the regular augmented factorization without outlier removal, 'Huynh' stands for [18], 'Ke' denotes [21], and 'Meng' stands for [24]. Here the reprojection variance was estimated only from the original inlying data by eliminating the added outliers so as to provide a fair comparison of different approaches, and the results were evaluated by 100 independent tests. It is obvious all three robust algorithms are resilient to outliers, as can be seen in Fig. 6, the ratio of outliers has little influence to the reprojection variance of the three robust algorithms. However, the proposed scheme outperforms other approaches in terms of accuracy, while the direct factorization yields significant error due to the influence of outliers.

We also compared the complexity of different approaches in terms of real computation time. All algorithms were implemented using Matlab on a Lenovo T500 laptop with 2,26GHz CPU. To generate different sizes of the tracking data, we vary the frame number 50 to 300 in steps of 50, and add 10% outliers to the tracking data. Table 1 tabulates the real computation time of different approaches. We can see from the table that the complexity of the proposed algorithm is much less than [21] and [24]. This is because [21] and [24] are based on the minimization of $L1$ norm, which is computationally more intensive than others. The proposed algorithm is slower than [18] since [18] does not incorporates the weighted factorization scheme, which leads to a much higher reprojection error as demonstrated in Fig. 6.

TABLE 1. Real computation time of different algorithms (unit: second).

Frame no.	50	100	150	200	250	300
Huynh [18]	1.19	2.35	3.68	6.41	10.45	12.69
Ke [21]	1.81	6.27	14.32	26.94	44.87	67.53
Meng [24]	2.01	7.25	16.28	30.14	49.86	76.69
Proposed	1.27	3.93	8.12	14.28	22.40	32.13

C. EVALUATIONS ON NONRIGID FACTORIZATION

We evaluated the performance of the augmented nonrigid factorization algorithm. In this experiment, we simulated a deformable space cube with 21 evenly distributed rigid points on each side. At the same time, we generated three sets of dynamic points (33×3 points) on the adjacent surfaces of the cube that were moving outward, as shown in Fig. 7. Using the synthetic cube, we simulate 100 continues images by the affine projection with random camera parameters with each image corresponds to a different 3D structure, and the image size is set at 800×800.

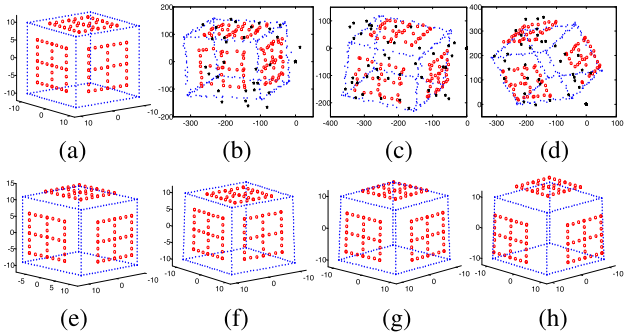

FIGURE 7. (a) (e) Two simulated space cubes with three sets of moving points; (b) (c) (d) three synthetic images with noise and outliers (black stars); and (f) (g) (h) the reconstructed 3D structures corresponding to the three images.

For the above simulated image sequence, we add 3 pixels Gaussian noise, as well as 10% outliers, to the tracking matrix. Fig. 7(b)–(d) show three noise and outlier corrupted images. Using the proposed robust algorithm, all outliers were successfully removed from the contaminated data, and

the motion and shape matrices were recovered. The corresponding 3D dynamic structures reconstructed by the proposed approach are shown in Fig. 7(f)–(h), respectively. From the results we can find that the dynamic cubic structures are correctly recovered by the proposed robust strategy.

VIII. EVALUATIONS ON REAL SEQUENCES

The method was evaluated extensively on a number of real image datasets. We will report the experimental results and evaluations on four real sequences in the paper.

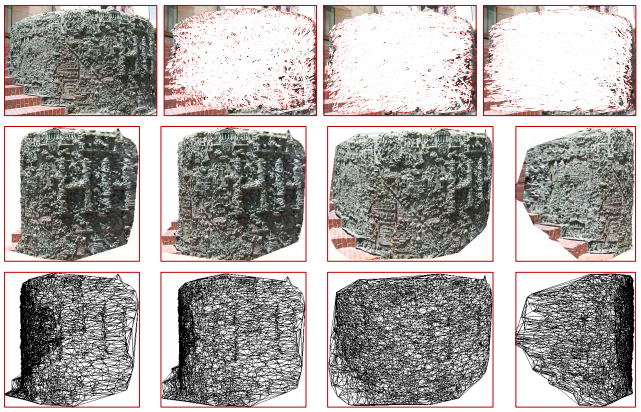

FIGURE 8. Reconstruction results of the fountain base sequence. (First row) four frames from the sequence, where the first one is a texture image, the other three images are overlaid with the the tracked features and outliers with disparities to the first image; (mid row) the reconstructed VRML model of the scene shown from different viewpoints with texture mapping; and (last row) the corresponding triangulated wireframe of the VRML model.

The first experiment is on the sequence of a fountain base captured at downtown San Francisco. The sequence consists of 10 images and on average 5648 features were tracked across the sequence using the feature tracking system [43]. It should be noted that feature tracking for this type of scene is hard since the texture of the images is homogeneous and repetitive. The tracking results contain many mismatches, in addition, we add an additional 5% false matching points in order to test the robustness of the proposed strategy. Fig. 8 shows all these features overlapped with disparities to the reference frame. Using the proposed algorithm, we successfully detect and remove the outliers. Then, we employ the weighted alternative algorithm to recover the 3D structure and camera motion parameters. Finally, the solution was upgraded to the metric space. As shown in Fig. 8, the reconstructed 3D structure looks realistic and most details are correctly identified.

The histogram distribution of the reprojection residual matrix (30) with outliers is shown in Fig. 9(a). The residuals are largely conform to the Gaussian assumption. It can be seen from the distribution that the outliers can be explicitly distinguished from the inliers by the estimated threshold, as shown in the figure. The histogram distribution of the residuals of the detected inlying data is shown in Fig. 9(b). Obviously, the residual error is reduced significantly after

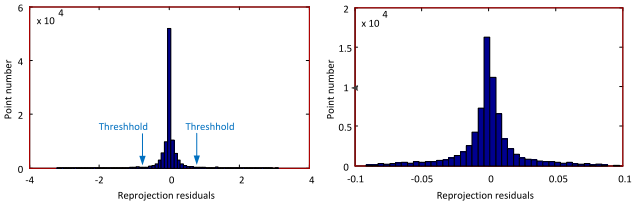

FIGURE 9. The histogram distribution of the reprojection residuals of the fountain base sequence before (a) and after (b) outliers rejection.

rejecting the outliers. As a quantitative evaluation, the final reprojection errors by different approaches are tabulated in Table 2, from which we can see that the proposed scheme yields the lowest reprojection error.

TABLE 2. Reprojection errors (pixel) for different datasets.

Dataset	Fountain	Hearst	Dinosaur	Face
Huynh [18]	0.736	0.742	0.926	0.697
Ke [21]	0.579	0.635	0.733	0.581
Meng [24]	0.512	0.578	0.645	0.525
Proposed	0.426	0.508	0.597	0.453

The second sequence is a corner of the Hearst Gym at UC Berkeley. There are 12 images in the sequence, and on average 1890 features were tracked in total. The correctly detected inlying features, together with about 5% outliers are shown in Fig. 10. Using the proposed robust algorithm, we successfully recovered the Euclidean structure of the scene, as shown in Fig. 10, all outliers are correctly detected and removed. As a comparison, the reprojection errors obtained using different algorithms are listed in Table 2, which shows that the proposed approach outperforms other robust algorithms.

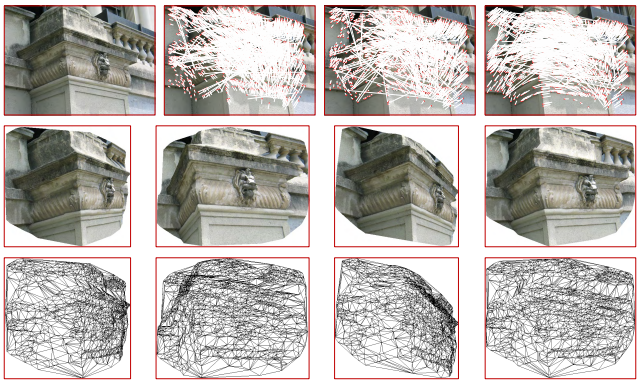

FIGURE 10. Reconstruction results of the Hearst Gym sequence. (First row) four frames from the sequence, where the first one is a texture image, the other three images are overlaid with the the tracked features and outliers with disparities to the first image; (mid row) the reconstructed VRML model of the scene shown from different viewpoints with texture mapping; and (last row) the corresponding triangulated wireframe of the VRML model.

The third test is on a deformable dinosaur sequence [5]. The image sequence consists of 231 frames with deformable structure of a dinosaur model. The image size is 570×338 pixel, and in total 49 features were tracked across the sequence. The initial tracking data are not very reliable, as shown in Fig. 11. We also add an additional 8% mismatches to the data so as to evaluate the robustness of the algorithm.

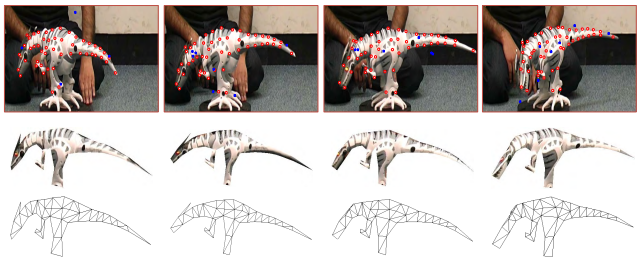

FIGURE 11. (First row) four frames from the dinosaur sequence superimposed with the tracked features (red circles) and added outliers (blue stars); (mid row) the reconstructed VRML models associated with each frame; and (last row) the corresponding triangulated wireframes of the VRML models.

Using the proposed approach, all outliers were successfully rejected, however, a few inliers were also removed due to large tracking errors. We then utilize the proposed nonrigid factorization algorithm to recover the structure and motion matrices, and upgrade the solution to the metric space. Fig. 11 shows the recovered deformable structures and the associated wireframes. The VRML model is visually plausible and the deformation of the model is correctly reconstructed.

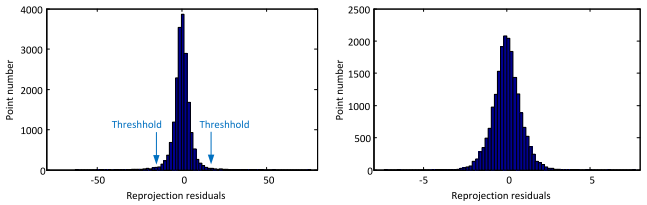

FIGURE 12. The histogram distribution of the reprojection residuals of the dinosaur sequence before (a) and after (b) outliers rejection.

The histogram distribution of the reprojection residual matrix (30) with outliers is shown in Fig. 12(a). 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. A threshold is computed from the mean and STD of the distribution, and the histogram of the residuals produced by the final solution after rejecting outliers is shown in Fig. 12(b), which shows a significant decrease on the residual errors. For comparison, we also extend the algorithms of 'Huynh,' 'Ke,' and 'Meng' to the nonrigid scenarios, and the reprojection errors by different algorithms are shown in Table 2. The proposed scheme yields the lowest reprojection error in this test.

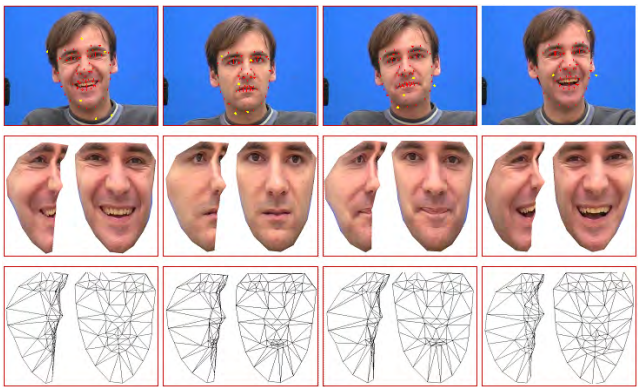

FIGURE 13. Reconstruction results of a human face with different facial expressions. (First row) four frames from the face sequence superimposed with the tracked features (red circles) and added outliers (yellow stars); (mid row) the reconstructed VRML models associated with each frame; and (last row) the corresponding triangulated wireframes of the VRML models.

The last experiment is on the Franck face sequence downloaded from FGnet,[1] as shown in Fig. 13. We selected 200 images with different facial expressions from the sequence. The image resolution is 720×576, and 68 features, which are automatically tracked using the active appearance model, are provided by the dataset. For test purpose, 8% outliers are added to the tracking data.

We apply the proposed robust scheme to remove the outliers and recover the metric structure of the face. The reconstructed VRML model of the face with texture and the corresponding wireframes are shown in Fig. 13. From the results we can see that different facial expressions have been correctly recovered by the proposed approach. The reprojection errors by different approaches are tabulated in Table 2. Like in other experiments, the proposed approach also yields the best performance in this experiment.

IX. CONCLUSION

In this paper, we first proposed a new augmented factorization framework which has been proved to be more accurate than the classical affine factorization, especially in the situation when the centroid of the imaged features could not be reliably recovered due to the presence of missing and outlying data. Then, we presented an alternatively weighted factorization algorithm to handle incomplete tracking data and alleviate the influence of large image noise. Finally, a robust factorization scheme was designed to deal with contaminated data with outliers and missing points. The proposed technique requires no prior information of the error distribution in the tracking data, and it can be directly extended to nonrigid factorization, which was rarely discussed in the literature. Extensive evaluations on both synthetic and real datasets demonstrated the advantages of the proposed scheme over previous methods.

[1] http://www-prima.inrialpes.fr/FGnet/html/home.html

REFERENCES

[1] A. Agrawal and R. Chellappa, "Robust ego-motion estimation and 3-D model refinement using surface parallax," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1215–1225, May 2006.

[2] 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. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1558–1565.

[3] A. Agudo, B. Calvo, and J. M. M. Montiel, "Finite element based sequential Bayesian non-rigid structure from motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 1418–1425.

[4] P. M. Q. Aguiar and J. M. F. Moura, "Rank 1 weighted factorization for 3D structure recovery: Algorithms and performance analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1134–1149, Sep. 2003.

[5] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Trajectory space: A dual representation for nonrigid structure from motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1442–1456, Jul. 2011.

[6] J.-C. Bazin, Y. Seo, R. Hartley, and M. Pollefeys, "Globally optimal inlier set maximization with unknown rotation and focal length," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 803–817.

[7] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3D shape from image streams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2000, pp. 690–696.

[8] S. Choi, T. Kim, and W. Yu, "Performance evaluation of RANSAC family," *J. Comput. Vis.*, vol. 24, no. 3, pp. 271–300, 1997.

[9] S. Christy and R. Horaud, "Euclidean shape and motion from multiple perspective views by affine iterations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 11, pp. 1098–1104, Nov. 1996.

[10] Y. Dai, H. Li, and M. He, "Projective multiview structure and motion from element-wise factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2238–2251, Sep. 2013.

[11] Y. Dai, H. Li, and M. He, "A simple prior-free method for non-rigid structure-from-motion factorization," *Int. J. Comput. Vis.*, vol. 107, no. 2, pp. 101–122, 2014.

[12] A. P. Eriksson and A. van den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L_1 norm," in *Proc. CVPR*, Jun. 2010, pp. 771–778.

[13] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[14] A. Gruber and Y. Weiss, "Multibody factorization with uncertainty and missing data using the em algorithm," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2004, pp. 1–8.

[15] R. Hartley and F. Schaffalitzky, "PowerFactorization: 3D reconstruction with missing or uncertain data," in *Proc. Austral.-Jpn. Adv. Workshop Comput. Vis.*, vol. 74. 2003, pp. 76–85.

[16] R. Hartley and R. Vidal, "Perspective nonrigid shape and motion recovery," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 276–289.

[17] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[18] D. Q. Huynh, R. Hartley, and A. Heyden, "Outlier correction in image sequences for the affine camera," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 585–590.

[19] M. Irani and P. Anandan, "Factorization with uncertainty," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 539–553.

[20] H. H. Je and A. Fitzgibbon, "Secrets of matrix factorization: Approximations, numerics, manifold optimization and random restarts," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4130–4138.

[21] Q. Ke and T. Kanade, "Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2005, pp. 739–746.

[22] R. Kennedy, L. Balzano, S. J. Wright, and C. J. Taylor, "Online algorithms for factorization-based structure from motion," *Comput. Vis. Image Understand.*, vol. 150, pp. 139–152, Sep. 2016.

[23] Z. Liu, P. Monasse, and R. Marlet, "Match selection and refinement for highly accurate two-view structure from motion," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[24] D. Meng and F. De La Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1337–1344.

[25] D. D. Morris and T. Kanade, "A unified factorization algorithm for points, line segments and planes with uncertainty models," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 696–702.

[26] I. Nurutdinova and A. Fitzgibbon, "Towards pointless structure from motion: 3D reconstruction and camera parameters from general 3D curves," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2363–2371.

[27] T. Okatani, T. Yoshida, and K. Deguchi, "Efficient algorithm for low-rank matrix factorization with missing components and performance comparison of latest algorithms," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 842–849.

[28] J. Oliensis and R. Hartley, "Iterative extensions of the sturm/triggs algorithm: Convergence and nonconvergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2217–2233, Dec. 2007.

[29] K. E. Ozden, K. Schindler, and L. Van Gool, "Multibody structure-from-motion in practice," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1134–1141, Jun. 2010.

[30] M. Paladini, A. Del Bue, J. Xavier, L. Agapito, M. Stošić, and M. Dodig, "Optimal metric projections for deformable and articulated structure-from-motion," *Int. J. Comput. Vis.*, vol. 96, no. 2, pp. 252–276, 2012.

[31] G. Qian, R. Chellappa, and Q. Zheng, "Bayesian algorithms for simultaneous structure from motion estimation of multiple independently moving objects," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 94–109, Jan. 2005.

[32] L. Quan, "Self-calibration of an affine camera from multiple views," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 93–105, 1996.

[33] V. Rabaud and S. Belongie, "Re-thinking non-rigid structure from motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[34] B. Resch, H. Lensch, O. Wang, M. Pollefeys, and A. Sorkine-Hornung, "Scalable structure from motion for densely sampled videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3936–3944.

[35] D. Scaramuzza, "1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," *Int. J. Comput. Vis.*, vol. 95, no. 1, pp. 74–85, 2011.

[36] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. CVPR*, Jun. 2016, pp. 4104–4113.

[37] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in *Proc. Eur. Conf. Comput. Vis.*, 1996, pp. 709–720.

[38] J. Taylor, A. D. Jepson, and K. N. Kutulakos, "Non-rigid structure from locally-rigid motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2761–2768.

[39] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, Nov. 1992.

[40] L. Torresani, A. Hertzmann, and C. Bregler, "Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 878–892, May 2008.

[41] B. Triggs, "Factorization methods for projective structure and motion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1996, pp. 845–851.

[42] G. Wang, "Robust structure and motion factorization of non-rigid objects," *Front. Robot. AI*, vol. 2, p. 30, Nov. 2015.

[43] G. Wang, "A hybrid system for feature matching based on SIFT and epipolar constraints," Dept. Elect. Commun. Eng., Univ. Windsor, Windsor, ON, USA, Tech. Rep. ECE201601, 2006.

[44] G. Wang and Q. M. J. Wu, "Perspective 3-D Euclidean reconstruction with varying camera parameters," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1793–1803, Dec. 2009.

[45] G. Wang and Q. M. J. Wu, "Quasi-perspective projection model: Theory and application to structure and motion factorization from uncalibrated image sequences," *Int. J. Comput. Vis.*, vol. 87, no. 3, pp. 213–234, 2010.

[46] G. Wang, J. S. Zelek, and Q. M. J. Wu, "Structure and motion recovery based on spatial-and-temporal-weighted factorization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 1590–1603, Nov. 2012.

[47] G. Wang, J. S. Zelek, and Q. M. J. Wu, "Robust structure from motion of nonrigid objects in the presence of outlying and missing data," in *Proc. Int. Conf. Comput. Robot Vis. (CRV)*, May 2013, pp. 159–166.

[48] G. Wang, J. S. Zelek, Q. J. Wu, and R. Bajcsy, "Robust rank-4 affine factorization for structure from motion," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 180–185.

[49] H. Wang, T.-J. Chin, and D. Suter, "Simultaneously fitting and segmenting multiple-structure data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1177–1192, Jun. 2012.

[50] J. Yan and M. Pollefeys, "A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 865–877, May 2008.

[51] J. Yu, T.-J. Chin, and D. Suter, "A global optimization approach to robust multi-model fitting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2041–2048.
[52] A. Zaharescu and R. Horaud, "Robust factorization methods using a Gaussian/uniform mixture model," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 240–258, 2009.
[53] L. Zelnik-Manor, M. Machline, and M. Irani, "Multi-body factorization with uncertainty: Revisiting motion consistency," *Int. J. Comput. Vis.*, vol. 68, no. 1, pp. 27–41, 2006.
[54] J. Zhang, M. Boutin, and D. G. Aliaga, "Pose-free structure from motion using depth from motion constraints," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2937–2953, Oct. 2011.
[55] E. Zheng and C. Wu, "Structure from motion using structure-less resection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2075–2083.

GUANGHUI WANG (M'10) received the Ph.D. degree in computer vision from the University of Waterloo, Canada, in 2014. From 2003 to 2005, he was a Research Fellow and Visiting Scholar with the Department of Electronic Engineering, The Chinese University of Hong Kong. From 2005 to 2007, he was a Professor with the Department of Control Engineering, Changchun Aviation University, China. From 2006 to 2010, he was a Research Fellow with the Department of Electrical and Computer Engineering, University of Windsor, Canada.

He is currently an Assistant Professor with the University of Kansas, USA. He is also with the Institute of Automation, Chinese Academy of Sciences, China, as an Adjunct Professor. He has authored one book *Guide to Three Dimensional Structure and Motion Factorization* (Springer-Verlag). He has published over 90 papers in peer-reviewed journals and conferences. His research interests include computer vision, structure from motion, object detection and tracking, artificial intelligence, and robot localization, and navigation. He has served as an Associate Editor and on the editorial board of two journals, as an Area Chair or TPC member of over 20 conferences, and as a reviewer of over 20 journals.