Available online at www.sciencedirect.com
SCIENCE DIRECT
Image and Vision Computing 24 (2006) 515–524
image AND vision COMPUTING
www.elsevier.com/locate/imavis

A subspace method for projective reconstruction from multiple images with missing data

W.K. Tang *, Y.S. Hung

Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

Received 29 January 2004; received in revised form 10 December 2005; accepted 16 February 2006

Abstract

In this paper, we consider the problem of projective reconstruction based on the subspace method. Unlike existing subspace methods which require that all the points are visible in all views, we propose an algorithm to estimate projective shape, projective depths and missing data iteratively. All these estimation problems are formulated within a subspace framework in terms of the minimization of a single consistent objective function, hence ensuring the convergence of the iterative solution. Experimental results using both synthetic data and real images are provided to illustrate the performance of the proposed method.

Keywords: Multiple views; Subspace method; Factorization method; Structure from motion

1. Introduction

The reconstruction 3D Euclidean structure from multiple uncalibrated 2D images has been a long standing difficult problem in computer vision that has many important applications such as geometric modeling and virtual scene synthesis. Many different approaches have been proposed for 3D reconstruction from multiple 2D images, including direct methods [1] that impose metric constraints from the outset to estimate cameras and 3D scene directly in an Euclidean frame, and stratified methods [2] that perform the recovery of 3D structure in stages, first in a projective frame followed by an upgrade to an affine and then the Euclidean frame. The stratified approach has the advantage that the 3D Euclidean reconstruction is decomposed into independent and simpler steps. Furthermore, it is shown in [3] that there is less chance for optimization algorithms to be trapped in local minima in a projective frame than a Euclidean frame, which is not surprising because optimization for the projective reconstruction is free of metric constraints. In the stratified approach, projective reconstruction is a necessary step prior to Euclidean reconstruction.

The factorization approach to projective reconstruction has received considerable attention in recent years. Factorization-based methods have an inherent advantage of being able to handle any number of images simultaneously without special treatment for any subgroup of views. In the factorization-based approach, the projective reconstruction problem is formulated as one of factorizing a scaled measurement matrix containing unknown depth parameters into a product of the structure and shape matrices. A key issue in the factorization approach is the determination of the unknown projective depths. In [4], the depths are determined by means of epipolar constraints in a non-iterative manner, but the method requires the estimation of the fundamental matrices between pairs of views and is sensitive to noise. Most of the other recent approaches use iterative methods to estimate the projective depths by minimizing an algebraic error (e.g. [5–8]) or a subspace proximity measure (e.g. [9–11]). Alternatives to factorization methods include iterative eigen algorithm of [12,13] and bundle adjustment techniques [14,15] which performs reconstruction by minimizing the 2D reprojection error. However, bundle adjustment [14], being based on non-linear optimization algorithms, requires a good starting point to yield an acceptable solution. Simulation studies based on synthetic and real images sequences suggest that subspace-based methods are able to converge to a solution with reprojection errors close to that obtained by bundle adjustment, but do not require prior knowledge of a good initial solution. In particular, the method of [10] has the advantage of being independent of the

* Corresponding author. Tel.: +852 2859 2728, Fax: +852 2559 8738.
E-mail addresses: wktang@eee.hku.hk (W.K. Tang), yshung@eee.hku.hk (Y.S. Hung).

0262-8856/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.imavis.2006.02.003

coordinate system chosen for the image planes. However, there are two issues that have not been addressed in [10].

First, the algorithm of [10] is not guaranteed to converge. In the subspace method of [10], the iterative algorithm alternates between (i) performing an SVD of a measurement matrix to determine the best 4D subspace approximating that spanned by the image points scaled by some yet unknown depths, and (ii) determining the depths to re-scale the image points so that the subspace for each image is as close as possible to the 4D subspace obtained in (i). Both steps (i) and (ii) are posed as minimization problems. However, different measures are minimized in the two different steps as if they are distinct minimization problems. Because of this, the convergence of the algorithm cannot be established. Second, the method of [10] assumes that all object points are visible on all images. In practice, it is unlikely that this condition is satisfied due to occlusion. For the factorization method to be practically applicable, it is important that the method caters for object points, which are visible only on some of the images but are missing from the other images. Despite the merits of the subspace approach, we are not aware of any existing subspace-based factorization method that can handle missing points while ensuring convergence.

In this paper, we will use the subspace method of [10] as a basis for the factorization method. In order to establish convergence, we will formulate the factorization problem as a minimization problem with a single consistent objective function that is optimized for distinct purposes (with respect to different sets of parameters) throughout an iterative algorithm. Missing points will be estimated as part of the algorithm. With convergence of the algorithm in mind, the difficulty is that the estimation of the missing points must be performed in a way consistent with the measure being minimized in the solution of the factorization problem. A key point underlying our approach is that subspaces will always be represented by an orthonormal basis, and this condition is enforced as a constraint at all stages of optimization. The constraint that subspace basis should always be orthonormal may render a minimization problem non-linear. Necessary solutions to such problems will be developed in this paper.

The paper is organized as follows. A measure for subspace inclusion is introduced in Section 2 before we formulate the factorization problem. An algorithmic solution to the factorization problem incorporating missing point estimation is developed in Section 3. Experimental results using both synthetic data and real images are provided in Section 4 to illustrate the performance of the proposed method. Section 5 contains some concluding remarks.

We shall use the following notation: $span(M)$ denotes the subspace spanned by the rows of a matrix M; $\|M\|_F$ denotes the Frobenius norm of a matrix M.

2. The subspace method

Consider a set of 3D points $X_j = [\,x_j \quad y_j \quad z_j \quad 1\,]^{\mathrm{T}}$ $(j = 1,\ldots,n)$ viewed by m cameras with projection matrices $P_i \in \Re^{3\times4}$ $(i = 1,\ldots,m)$. Let the projection of X_j on the ith view be the image point $w_{ij} = [\,u_{ij} \quad v_{ij} \quad 1\,]^{\mathrm{T}}$ $(i = 1,\ldots,m;\ j = 1,\ldots,n)$ in normalized homogeneous coordinates. The 2D image points w_{ij} on the ith view can be assembled into an *unscaled measurement matrix* W_i (for the ith view) given by:

$$W_i = [\,w_{i1} \quad w_{i2} \quad \ldots \quad w_{in}\,] \in \Re^{3\times n}.$$

The image point w_{ij} is related to X_j by

$$\lambda_{ij} w_{ij} = P_i X_j \tag{1}$$

where λ_{ij} is the *depth* of the object point X_j seen by the ith camera (also referred to as the depth corresponding to w_{ij}). The *shape matrix* X combines all 3D points X_j as:

$$X = [\,X_1 \quad X_2 \quad \ldots \quad X_n\,] \in \Re^{4\times n}. \tag{2}$$

The product of P_i and X represents the projections of all the 3D points onto the ith view, giving rise to a *scaled measurement matrix* $\tilde{W}_i$ for the ith view:

$$\tilde{W}_i = [\,\lambda_{i1} w_{i1} \quad \lambda_{i2} w_{i2} \quad \ldots \quad \lambda_{in} w_{in}\,] = P_i X \in \Re^{3\times n}. \tag{3}$$

We will also write the scaled measurement matrix as

$$\tilde{W}_i = [\,\tilde{w}_{i1} \quad \tilde{w}_{i2} \quad \ldots \quad \tilde{w}_{in}\,]$$

where

$$\tilde{w}_{ij} = \lambda_{ij} w_{ij}.$$

The scaled measurement matrix is related to unscaled measurement matrix by

$$\tilde{W}_i = W_i \Lambda_i \tag{4}$$

where Λ_i is a diagonal scaling matrix:

$$\Lambda_i = diag(\lambda_{i1},\ldots,\lambda_{in}) \in \Re^{n\times n}. \tag{5}$$

From (3), the rows of the scaled measurement matrix $\tilde{W}_i$ span a 3D subspace in $\Re^n$ that lies in the 4D subspace spanned by the rows of X, i.e.

$$span(\tilde{W}_i) \subseteq span(X) \subset \Re^n. \tag{6}$$

Let $P = [P_1^{\mathrm{T}}, P_2^{\mathrm{T}},\ldots,P_m^{\mathrm{T}}]^{\mathrm{T}} \in \Re^{3m\times4}$ be the *joint projection matrix*. The projection of all the 3D points onto all the views gives rise to a *scaled measurement matrix* $\tilde{W}$:

$$\tilde{W} = \begin{bmatrix} \tilde{W}_1 \\ \vdots \\ \tilde{W}_m \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix} X = PX \in \Re^{3m\times n}. \tag{7}$$

It follows from (7) that the 4D subspace spanned by the rows of X can be written:

$$span(X) = span(\tilde{W}) = \bigcup_{i=1}^{m} span(\tilde{W}_i). \tag{8}$$

In general, the depths $\{\lambda_{ij}\}$ are unknown. In the factorization method for projective reconstruction, a set of consistent depths $\{\lambda_{ij}\}$ has to be estimated so that the scaled measurement matrix $\tilde{W}$ can be factorized into two rank-4 matrices P and X as in (7). This condition can be expressed in terms of the subspace inclusion conditions (6) and (8), which provide the basis for the

subspace method. In practice, these subspace conditions will not be satisfied exactly because of image noise and other uncertainties. The principle of the subspace method is to determine a rank-4 matrix X whose row space best approximates the union of all 3D subspaces spanned by the $\tilde{W}_i$'s, and to estimate for each view the depths Λ_i so that the subspace spanned by $\tilde{W}_i$ is as close as possible to being contained in the subspace spanned by X. To give a more precise statement of the factorization problem based on the subspace method, we need to introduce a measure for subspace inclusion to quantify how far the conditions (6) and (8) are satisfied.

2.1. Measure for subspace inclusion

Let $X_\perp$ be a matrix whose columns form an orthonormal basis for the right null space of X. Suppose, without loss of generality, that the rows of X has been normalized to give an orthonormal basis for *span*(X). That is, $[X^{\mathrm{T}} \quad X_\perp]$ is an orthogonal matrix satisfying:

$$X[X^{\mathrm{T}} \quad X_\perp] = [I \quad 0]. \tag{9}$$

For any $\tilde{W}_i$, let $O(\tilde{W}_i)$ denote an orthonormal basis for *span*($\tilde{W}_i$). If $\tilde{W}_i$ is of full row rank, $O(\tilde{W}_i)$ can be obtained as:

$$O(\tilde{W}_i) = (\tilde{W}_i \tilde{W}_i^{\mathrm{T}})^{-1/2} \tilde{W}_i. \tag{10}$$

Clearly, (6) is satisfied if and only if:

$$O(\tilde{W}_i) X_\perp = 0.$$

In the case where *span*($\tilde{W}_i$) is not contained in *span*(X), we will use $(1/\sqrt{3})\|O(\tilde{W}_i)X_\perp\|_F$ as a subspace inclusion measure (of $\tilde{W}_i$ in X) to indicate how far *span*($\tilde{W}_i$) deviates from being contained in *span*(X). The factor $1/\sqrt{3}$ is introduced to normalize the range of the subspace inclusion measure to [0,1]. Note that since $[X^{\mathrm{T}} \quad X_\perp]$ is orthogonal:

$$\frac{1}{\sqrt{3}}\|O(\tilde{W}_i)[X^{\mathrm{T}} \quad X_\perp]\|_F = \frac{1}{\sqrt{3}}\|O(\tilde{W}_i)\|_F = 1.$$

It follows that:

$$\frac{1}{3}\|O(\tilde{W}_i)X^{\mathrm{T}}\|_F^2 + \frac{1}{3}\|O(\tilde{W}_i)X_\perp\|_F^2 = 1. \tag{11}$$

Hence:

$$0 \le \frac{1}{\sqrt{3}}\|O(\tilde{W}_i)X_\perp\|_F \le 1.$$

To approximate the condition (6) as close as possible, one would minimize the subspace inclusion measure. Note that by (11), minimizing $(1/\sqrt{3})\|O(\tilde{W}_i)X_\perp\|_F$ is equivalent to maximizing $(1/\sqrt{3})\|O(\tilde{W}_i)X^{\mathrm{T}}\|_F$. The later may have some advantages since X may have a dimension $(=4\times n)$ considerably smaller than the dimension $(=n\times(n-4))$ of $X_\perp$, particularly if n is large.

2.2. Problem formulation

Given image coordinates w_{ij} $(i=1,\ldots,m;\ j=1,\ldots,n)$, the factorization problem is to determine projective depths λ_{ij} so that the scaled measurement matrix $\tilde{W}$ can be factorized into two rank-4 matrices P and X as in (7). If the image coordinates contain noise, the factorization can only be approximate and the results will depend on the criterion used in the factorization. In the subspace method for factorization, the criterion for the determination of λ_{ij} is to enforce (6) and (8) as far as possible, subject to the constraint that X has rank-4. Using the subspace inclusion measure defined above, the problem can be stated as

$$\min_{\lambda_{ij}} \min_{rank(X)=4} \frac{1}{3m} \sum_{i=1}^{m} \|O(W_i \Lambda_i) X_\perp\|_F^2 \tag{12}$$

where the subspace inclusion measures for all m views have been combined into a single objective function for minimization in (12).

The ability to handle missing points is essential for any multi-view techniques, as there are bound to be missing data due to occlusion. Let the available image points be indexed by ordered pairs of the set:

$$\mathcal{I} = \{(i,j) | w_{ij} \text{ is observed as point } j \text{ on view } i\}. \tag{13}$$

If there are image points missing from some of the views, the unscaled measurement matrix W_i will contain 'holes' that need to be filled in before the minimization problem (12) can be solved. In this case, (12) has to be extended to include the estimation of missing elements as:

$$\min_{\{w_{ij} | (i,j) \notin \mathcal{I}\}} \min_{\lambda_{ij}} \min_{rank(X)=4} \frac{1}{3m} \sum_{i=1}^{m} \|O(W_i \Lambda_i) X_\perp\|_F^2. \tag{14}$$

As far as the missing data points are concerned, since both λ_{ij} and w_{ij} are to be determined, we may estimate $\tilde{w}_{ij} (=\lambda_{ij} w_{ij})$ instead of w_{ij}. In this case, (14) becomes:

$$\min_{\{\tilde{w}_{ij} | (i,j) \notin \mathcal{I}\}} \min_{\lambda_{ij}} \min_{rank(X)=4} \frac{1}{3m} \sum_{i=1}^{m} \|O(W_i \Lambda_i) X_\perp\|_F^2. \tag{15}$$

3. Algorithmic solution

In view of the form of the minimization problem (15), it is natural to consider solving the variables of the three nested minimization problems by estimating X, λ_{ij} and $\tilde{w}_{ij}$ (missing points only) successively in an iterative loop, giving rise to the following algorithm (where superscript k denotes the variables in the kth iteration). In the algorithm, each of X, λ_{ij} and $\tilde{w}_{ij}$ is solved in turn as a free parameter of an optimization problem while the other two variables are fixed at their latest estimates.

Algorithm 1. (subspace algorithm)

(1) Put $k=0$ and assign initial values to $\{w_{ij}^0 | (i,j) \notin \mathcal{I}\}$ (in W_i^0) and λ_{ij}^0 (in Λ_i^0). (e.g. set w_{ij}^0 for all missing points at the mean value of the available points and $\lambda_{ij}^0 = 1 \,\forall\, i, j$).
(2) Put $k=k+1$. Regarding W_i^{k-1} and Λ_i^{k-1} as fixed, determine $X_\perp^k$ by solving

$$\varepsilon_1^k = \min_{rank(X^k)=4} \frac{1}{3m} \sum_{i=1}^{m} \|O(W_i^{k-1} \Lambda_i^{k-1}) X_\perp^k\|_F^2 \tag{16}$$

subject to (9).

(3) For each i ($=1,\ldots,m$), fix W_i^{k-1} and $X_\perp^k$ and determine λ_{ij}^k (in Λ_i^k) by solving:

$$\min_{\lambda_{ij}^k} \|O(W_i^{k-1} \Lambda_i^k) X_\perp^k\|_F^2. \tag{17}$$

Compute:

$$\varepsilon_2^k = \frac{1}{3m} \sum_{i=1}^{m} \|O(W_i^{k-1} \Lambda_i^k) X_\perp^k\|_F^2. \tag{18}$$

(4) Fix $X_\perp^k$ and determine $\{\tilde{w}_{ij}^k | (i,j) \notin \mathcal{I}\}$ (in $\tilde{W}_k^i$) by solving:

$$\varepsilon_3^k = \min_{\{\tilde{w}_{ij}^k | (i,j) \notin \mathcal{I}\}} \frac{1}{3m} \sum_{i=1}^{m} \|O(\tilde{W}_i^k) X_\perp^k\|_F^2. \tag{19}$$

(5) Repeat steps 2, 3 and 4 until ε_3^k converges.
(6) Output W_i^k, Λ_i^k and X^k and stop.

Since, the same objective function is being minimized (but with respect to different sets of parameters) in the three minimization problems (16), (17) and (19) in Algorithm 1, the cost ε_1^k, ε_2^k and ε_3^k are monotonic decreasing satisfying:

$$\cdots \geq \varepsilon_3^{k-1} \geq \varepsilon_1^k \geq \varepsilon_2^k \geq \varepsilon_3^k \geq \cdots \geq 0.$$

It follows that ε_3^k is guaranteed to converge. It is possible for Algorithm 1 to converge to a local minimum. (This is however in contrast to many existing factorization methods that alternate between inconsistent objective functions in the optimization process and are not guaranteed to converge even to a local minimum.) Extensive experiments (see [16]) show that the subspace algorithm has a better chance of converging to the global minimum as compared to methods like bundle adjustment. This is because the subspace algorithm is not purely local descent and therefore is less prone to be trapped at local minima.

It remains to solve the minimization problems (16), (17) and (19). In these problems, the objective function is linear in $X_\perp$, but non-linear in Λ_i (containing λ_{ij}) and $\tilde{W}_i$ (containing the missing points $\tilde{w}_{ij}$). We will next develop solutions to the minimization problems (16), (17) and (19). In the following, the superscript k with be dropped for notational simplicity.

3.1. Estimation of rank-4 shape matrix

We need to estimate a rank-4 shape matrix in step 2 of Algorithm 1. Regarding W_i and Λ_i ($i=1,\ldots,m$) as known, X is determined by solving

$$\min_{rank(X)=4} \frac{1}{3m} \sum_{i=1}^{m} \|O(W_i \Lambda_i) X_\perp\|_F^2$$

$$= \max_{rank(X)=4} \frac{1}{3m} \left\| \begin{bmatrix} O(W_1 \Lambda_1) \\ \vdots \\ O(W_m \Lambda_m) \end{bmatrix} X^{\mathrm{T}} \right\|_F^2 \tag{20}$$

subject to $XX^{\mathrm{T}} = I$. It is well established that (20) can be solved by means of an SVD:

$$\begin{bmatrix} O(W_1 \Lambda_1) \\ \vdots \\ O(W_m \Lambda_m) \end{bmatrix} = USV^{\mathrm{T}}.$$

With the singular values along the diagonal of S arranged in decreasing order of magnitude, the solution to (20) is obtained by setting X and $X_\perp$ to the submatrices consisting of the first 4 and the last $(n-4)$ columns of V, respectively.

3.2. Depth estimation

Given W_i and X, the depths λ_{ij} ($j=1,\ldots,n$) along the diagonal of Λ_i are determined from a minimization problem of the form (17), which is equivalent to:

$$\max_{\lambda_{ij}} \|(W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1/2} (W_i \Lambda_i) X^{\mathrm{T}}\|_F^2. \tag{21}$$

The objective function (21) can be written:

$$f(\Lambda_i) = trace[X \Lambda_i W_i^{\mathrm{T}} (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} W_i \Lambda_i X^{\mathrm{T}}].$$

To maximize $f(\Lambda_i)$ with respect to λ_{ij}, consider:

$$\frac{\partial f(\Lambda_i)}{\partial \lambda_{ij}} = trace\left\{ \frac{\partial}{\partial \lambda_{ij}} [X \Lambda_i W_i^{\mathrm{T}} (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} W_i \Lambda_i X^{\mathrm{T}}] \right\}.$$

Making use of the formula $d(Z^{-1}) = -Z^{-1}(\mathrm{d}Z)Z^{-1}$, we have:

$$\begin{aligned} \frac{\partial f(\Lambda_i)}{\partial \lambda_{ij}} = {} & trace\{X_j w_{ij}^{\mathrm{T}} (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} W_i \Lambda_i X^{\mathrm{T}}\} \\ & + trace\{X \Lambda_i W_i^{\mathrm{T}} (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} w_{ij} X_j^{\mathrm{T}}\} \\ & - trace\{X \Lambda_i W_i^{T} (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} (2\lambda_{ij} w_{ij} w_{ij}^{\mathrm{T}}) (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} W_i \Lambda_i X^{\mathrm{T}}\}. \end{aligned} \tag{22}$$

If we define

$$\bar{X}_{ij} = X \Lambda_i W_i^{\mathrm{T}} (W_i \Lambda_i^2 W_i^{\mathrm{T}})^{-1} w_{ij} \in \Re^{4 \times 1} \tag{23}$$

then (22) can be simplified as

$$\begin{aligned} \frac{\partial f(\Lambda_i)}{\partial \lambda_{ij}} &= trace\{X_j \bar{X}_{ij}^{\mathrm{T}}\} + trace\{\bar{X}_{ij} \bar{X}_j^{\mathrm{T}}\} - 2\lambda_{ij}\, trace\{\bar{X}_{ij} \bar{X}_{ij}^{\mathrm{T}}\} \\ &= 2\bar{X}_{ij}^{\mathrm{T}} X_j - 2\lambda_{ij} \bar{X}_{ij}^{\mathrm{T}} \bar{X}_{ij} \end{aligned}$$

where we have made use of $trace(AB) = trace(BA)$ when A has the same dimension as B^{T}. A necessary condition for λ_{ij} to

maximize $f(\Lambda_i)$ is:

$$\frac{\partial f(\Lambda_i)}{\partial \lambda_{ij}} = 2\bar{X}_{ij}^{\mathrm{T}} X_j - 2\lambda_{ij}\bar{X}_{ij}^{\mathrm{T}}\bar{X}_{ij} = 0. \tag{24}$$

As $\bar{X}_{ij}$ in (24) depends on λ_{ij}, the equation cannot be solved in closed form for λ_{ij}. However, an iterative solution to (24) can be obtained by setting

$$\lambda_{ij} = \frac{\bar{X}_{ij}^{\mathrm{T}} X_j}{\bar{X}_{ij}^{\mathrm{T}}\bar{X}_{ij}} \tag{25}$$

where the $\bar{X}_{ij}$ on the right hand side of (25) are based on old values of λ_{ij} obtained in the previous iteration. Our experience is that such a solution will typically converge in about five iterations. If we use suboptimal estimates of λ_{ij} in Algorithm 1, the algorithm will still converge, but at the expense of more iterations of the main loop. For example, if we compute λ_{ij} only once using (25) in step 3 of Algorithm 1, the number of iterations through the main loop may be increased by three to four times. As a result, the total computation time is increased. It is therefore preferable to iterated the sub-loop until λ_{ij} converges in step 3.

3.3. Missing data estimation

Given $X_\perp$, the problem stated in (19) is:

$$\begin{aligned}&\min_{\{\tilde{w}_{ij}|(i,j)\notin\mathcal{A}\}} \frac{1}{3m}\sum_{i=1}^{m} \|O(\tilde{W}_i)X_\perp\|_F^2 \\ &= \min_{\{\tilde{w}_{ij}|(i,j)\notin\mathcal{A}\}} \frac{1}{3m}\sum_{i=1}^{m} \|(\tilde{W}_i\tilde{W}_i^{\mathrm{T}})^{-1/2}\tilde{W}_i X_\perp\|_F^2.\end{aligned} \tag{26}$$

This is a non-linear optimization problem where some elements of the matrix $\tilde{W}_i$ (corresponding to missing data points with $(i,j)\notin\mathcal{A}$) contain free parameters. From the left hand side of (26), we can regard the problem as one of finding an orthonormal basis for a 3D subspace which is free up to choices in the entries $\{\tilde{w}_{ij}^k|(i,j)\notin\mathcal{A}\}$ so that the subspace is orthogonal to $X_\perp$ as far as possible. We will now show that by separating the fixed and the free parts of $\tilde{W}_i$, we can transform (26) into an equivalent problem solvable by means of a singular value decomposition technique.

Since, each column of $\tilde{W}_i$ is either free or fixed, we can reorder the columns of $\tilde{W}_i$ by a permutation matrix R_i so that

$$\tilde{W}_i R_i = [\,M \quad A\,] \tag{27}$$

where M contains all the free columns of $\tilde{W}_i$ corresponding to missing points and A contains the available part of $\tilde{W}_i$ corresponding to indices $(i,j)\in\mathcal{A}$. Note that R_i is orthogonal satisfying $R_i R_i^{\mathrm{T}} = I$. Let $R_i^{\mathrm{T}} X_\perp$ be partitioned in a way compatible with $[\,M \quad A\,]$:

$$R_i^{\mathrm{T}} X_\perp = \begin{bmatrix} X_{\perp 1} \\ X_{\perp 2}\end{bmatrix}.$$

Then, we have

$$\begin{aligned}(\tilde{W}_i\tilde{W}_i^{\mathrm{T}})^{-1/2}\tilde{W}_i X_\perp^{\mathrm{T}} &= ((\tilde{W}_i R_i)(\tilde{W}_i R_i)^{\mathrm{T}})^{-1/2}(\tilde{W}_i R_i)(R_i^{\mathrm{T}} X_\perp) \\ &= \left([\,M \quad A\,]\begin{bmatrix} M^{\mathrm{T}} \\ A^{\mathrm{T}}\end{bmatrix}\right)^{-1/2}[\,M \quad A\,]\begin{bmatrix} X_{\perp 1} \\ X_{\perp 2}\end{bmatrix} \\ &= (MM^{\mathrm{T}} + AA^{\mathrm{T}})^{-1/2}[\,M \quad (AA^{\mathrm{T}})^{-1/2}\,]\begin{bmatrix} X_{\perp 1} \\ (AA^{\mathrm{T}})^{-1/2}AX_{\perp 2}\end{bmatrix}\end{aligned} \tag{28}$$

where we have assumed that A is of full row rank ($=3$) so that $(AA^{\mathrm{T}})^{1/2}$ is invertible.

Proposition 1. *Given A of full row rank. For any $M\in\Re^{3\times p}$, let*:

$$\begin{aligned}U_1 &= (MM^{\mathrm{T}} + AA^{\mathrm{T}})^{-1/2}M \quad \text{and} \\ U_2 &= (MM^{\mathrm{T}} + AA^{\mathrm{T}})^{-1/2}(AA^{\mathrm{T}})^{1/2}\end{aligned} \tag{29}$$

Then, (29) *represents a one-to-one correspondence between the set of all $M\in\Re^{3\times p}$ and row orthonormal matrices $[\,U_1 \quad U_2\,]\in\Re^{3\times(p+3)}$ for which U_2 is non-singular.*

Proof. For any $M\in\Re^{3\times p}$, the matrix $[\,U_1 \quad U_2\,]$ defined by Eq. (29) satisfies

$$[\,U_1 \quad U_2\,][\,U_1 \quad U_2\,]^{\mathrm{T}} = I$$

and hence is a row orthonormal matrix. Since, A is of full row rank, U_2 is non-singular.

Conversely, suppose $[\,U_1 \quad U_2\,]$ is row orthonormal with U_2 non-singular. From (29), we can solve for a unique M given by

$$M = (MM^{\mathrm{T}} + AA^{\mathrm{T}})^{1/2}U_1 = (AA^{\mathrm{T}})^{1/2}U_2^{-1}U_1 \tag{30}$$

which completes the proof. □

Proposition 1 shows that to solve an optimization problem in which the free parameter M appears in the form given by the right hand side of (29) is equivalent to solving for a row orthonormal $[\,U_1 \quad U_2\,]$ under the condition that U_2 is non-singular. In view of (28), the problem (26) is equivalent to:

$$\begin{aligned}&\min_{[U_1 U_2]} \|[\,U_1 \quad U_2\,]\begin{bmatrix} X_{\perp 1} \\ (AA^{\mathrm{T}})^{-1/2}AX_{\perp 2}\end{bmatrix}\|_F^2 \quad \text{subject to} \\ &[\,U_1 \quad U_2\,][\,U_1 \quad U_2\,]^{\mathrm{T}} = I \quad \text{and} \quad \det U_2 \neq 0.\end{aligned} \tag{31}$$

The minimization problem (31) can be solved by means of a singular value decomposition:

$$\begin{bmatrix} X_1^{\mathrm{T}} \\ (AA^{\mathrm{T}})^{-1/2}AX_2^{\mathrm{T}}\end{bmatrix} = USV^{\mathrm{T}}.$$

$[\,U_1 \quad U_2\,]$ is then obtained as the last three rows of U^{T} corresponding to the largest three singular values of S. The condition that U_2 is non-singular needs to be verified. Since, A is of full row rank, we see from (29) that U_2 is non-singular unless M becomes unbounded. Noting that M is an estimate of the homogeneous image coordinates of the 2D projections of

the missing points represented by X_1, M must be finite since X_1 as obtained in step 2 of the algorithm is finite. Hence, the case where U_2 is singular should never arise. Under the condition that U_2 is non-singular, M can be determined using (30). $\tilde{W}_i$ can then be recovered using (27). With the solutions given in Sections 3.1–3.3, Algorithm 1 can be readily implemented.

4. Experimental results

In this section, the proposed method is firstly evaluated using synthetic data and compared with Sturm–Triggs' method [4] and Heyden's subspace method [10]. Then, we consider real images with missing data, in which case only the proposed method is evaluated as the methods of [4,10] are not applicable.

4.1. Synthetic data

A total of 40 points are generated at random spatial locations within a sphere C of radius 0.2 m. Eleven cameras with fixed intrinsic parameters are then placed randomly inside a box B of dimensions 0.4 m×1 m×1 m in front of the 3D points, with the centre of the box located at 1 m from the centre of the sphere C, as shown in Fig. 1(a). The cameras are pointed towards the set of points and the intrinsic parameters are chosen so that all points are visible on all the image planes within an image area of 800×800 pixels.

The image points are contaminated by different Gaussian noise levels (with standard deviation ranging from 0 to 4 pixels in increments of 0.5 pixel). For each noise level, the algorithm of Section 3 is applied to 20 sets of randomly generated noisy data and mean values of various error measures are computed. The figures to be given below are plots of such mean values.

4.1.1. Performance on 2D reprojection error

The 2D reprojection errors (2DRPE) of the reconstruction results using our algorithm, Sturm–Triggs' method [4] and Heyden's method [10] for different noise levels are given in Fig. 1(b). Since, the subspace algorithms do not directly minimize 2D reprojection error, the results are obtained while the root mean square (RMS) values of the 2D reprojection errors are smallest. Fig. 1(b) shows that our method produces only marginally better 2D reprojection errors compared with the method of [10]. The RMS 2DRPE of the method of [4] is almost equal to the added noise. Since, the proposed method and Heyden's method are based on similar subspace criterion, it is not surprising that the results are similar when both methods are applicable. Note that the RMS 2DRPE are almost linearly related to and are less than the magnitude of the added noise levels. It is also notable that both subspace methods are superior than the factorization method [4] in term of 2DRPE.

4.1.2. Performance on projective depth estimation

Since, the projective depths cannot be uniquely recovered, to assess how good the estimated projective depths are, we make use of the cross ratio of estimated projective depths defined as ($i=1,\ldots,m-1$; $j=1,\ldots,n-1$)

$$\hat{c}_{ij} = \frac{\hat{\lambda}_{ij}\hat{\lambda}_{i+1,j+1}}{\hat{\lambda}_{i,j+1}\hat{\lambda}_{i+1,j}} \tag{32}$$

where $\hat{\lambda}_{ij}$ represents the estimated projective depths. The mean-squared cross-ratio error (MSCRE) is then defined as

$$\frac{1}{(m-1)(n-1)}\sum_{i,j}(c_{ij}-\hat{c}_{ij})^2 \tag{33}$$

where c_{ij} are the cross ratios computed using the ground truth depths. Fig. 1(c) shows that the MSCRE of the proposed method is slightly better than that obtained using the methods of [4,10] and both methods of [4,10] are overlapped.

4.1.3. Performance on 3D error

The evaluation on 3D error is performed by upgrading the reconstructed 3D points X_j^* in the projective space to an Euclidean space by means of a collineation $T \in \Re^{4\times4}$ which is obtained by minimizing

$$e_{3D} = \sqrt{\min_T \sum_j \|X_j - \alpha_j T X_j^*\|^2} \tag{34}$$

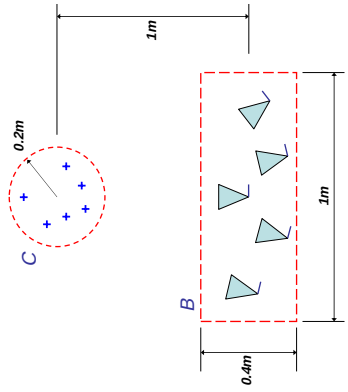
where X_j is the ground truth 3D point and α_j is a scaling factor for normalizing the fourth component of TX_j^* to 1. The *RMS* 3D error is defined as $1/\sqrt{n}e_{3D}$. Fig. 1(d) shows the RMS 3D errors plotted against different noise levels. The RMS 3D errors for both subspace methods are very small compared with the diameter (=400 mm) of the cloud of 3D points.

4.1.4. Convergence of algorithm

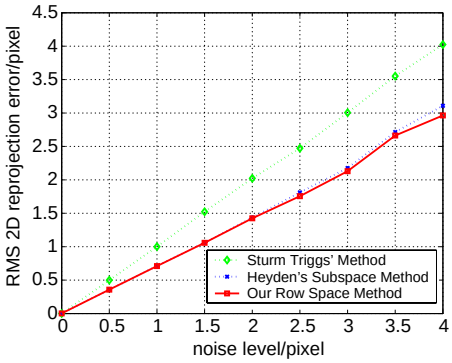
Fig. 1(e) shows the convergence of the objective function being minimized in the proposed method for different noise levels. For the synthetic scene, it takes less than 10 iterations to converge to an acceptable solution. The number of iterations for convergence is not significantly affected by the added noise levels. Computational experiments show that the computation time required for each iteration of the main loop in Algorithm 1 is of the order of $O(nm)$, i.e. proportional to the size of the measurement matrix.

4.1.5. Missing data estimation

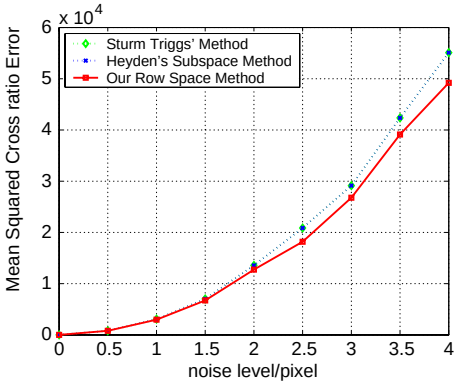
To evaluate the performance of our algorithm for missing data estimation, we use the same synthetic data sets as before but remove 10% of 2D points in the measurement matrix randomly to simulate missing data. Since, Sturm–Triggs' and Heyden's methods do not cater for missing points, only the proposed method is applied in this experiment. The RMS 2D reprojection errors for both visible points and missing points are plotted against different noise levels in Fig. 1(f). For visible points, the RMS 2D reprojection errors are computed with respect to the noisy data as well as the ground truth, whereas for missing points, the reprojection errors are computed with respect to ground truth only as measured data is supposedly not available. Fig. 1(f) shows that the errors of the estimated missing points are comparable with the levels of the added noise.



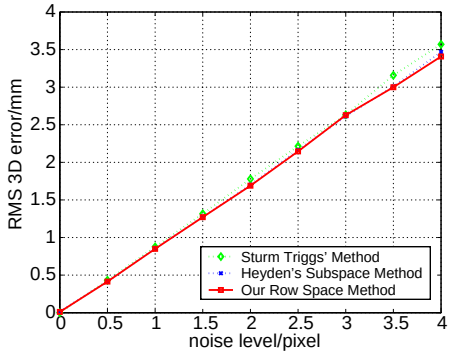
(a) Synthetic scene configuration



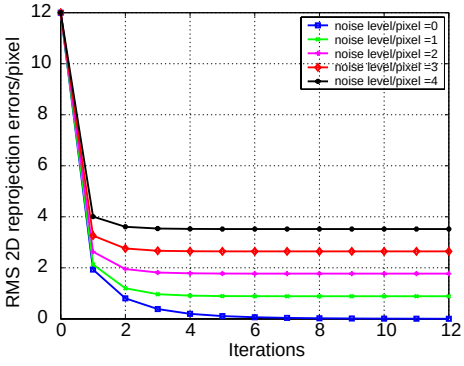
(b) 2D reprojection error



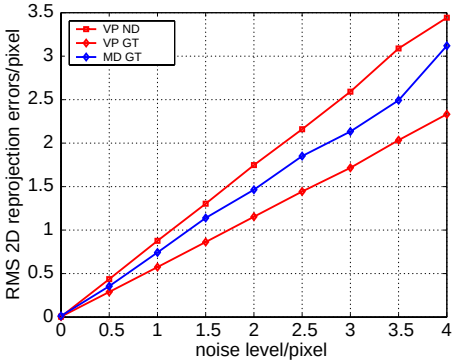
(c) Projective-depth cross-ratio error



(d) 3D error



(e) Convergence for different noise levels

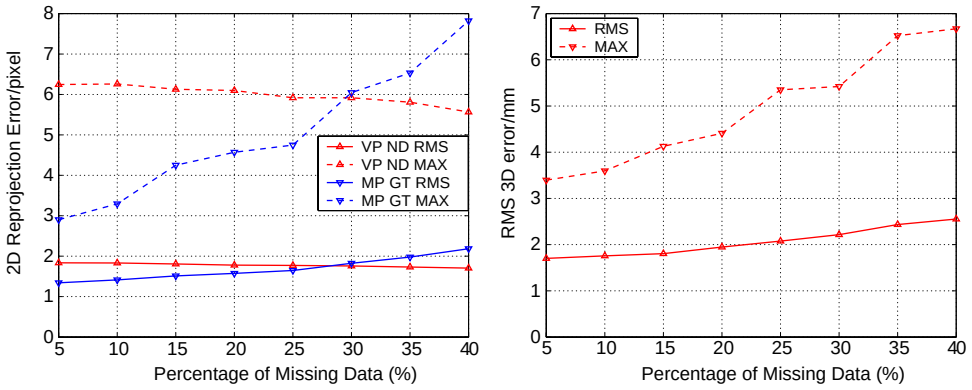


(f) Missing data estimation (VP-ND and VP-GT denote the 2DRPE of the reprojected visible points compared with noisy data and ground truth respectively; MD-GT denotes the 2DRPE of estimated missing data compared with the ground truth)

Fig. 1. Results of the synthetic data example.

In order to assess how much missing data the algorithm can handle, we remove from the synthetic images varying percentages of data ranging from 5 to 40% in 5% increments. The Gaussian noise level is set at $\sigma = 2$ pixels. We perform 30 trials for each percentage of missing data with 2D points randomly marked as missing. To ensure that the data is not degenerate, we require that at least seven points remain visible to every three consecutive views. This condition becomes increasingly hard to satisfy when the percentage of missing points exceeds 40%, as the scene contains only 40 feature points.

Fig. 2(a) and (b) show that both the 2D reprojection error and 3D error remain acceptable for all visible points, but the RMS and maximum error for the missing points increase with increasing percentages of missing data. The increase is gradual and somewhat linear as opposed to an abrupt breakdown of the algorithm. The amount of missing data that can be tolerated therefore depends on the level of error acceptable for the reconstruction. In the real experiments to follow, we will see that reasonable reconstruction of shape can be achieved from real images with more than 50% missing data.



(a) 2D reprojection error (VP-ND-RMS and VP-ND-MAX denote the RMS and maximum 2DRPE of the reprojected visible points compared with noisy data respectively; MP-GT-RMS and MP-GT-MAX denote the RMS and maximum 2DRPE of estimated missing data compared with the ground truth respectively)

(b) 3D error (RMS and MAX denote the RMS and maximum 3DE compared with ground truth respectively)

Fig. 2. Results of the synthetic data with missing data.

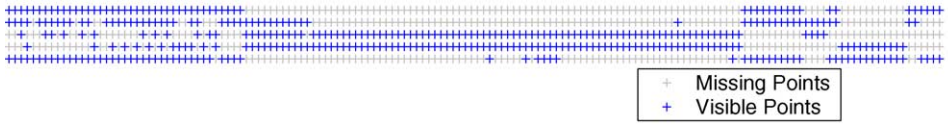
4.2. Real images with missing data

The proposed method is further evaluated using two real image sequences with large percentages of missing points. The methods of [4,10] are not applicable to these examples because of the missing data.

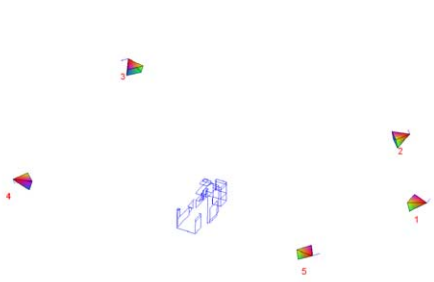
4.2.1. Castle model image sequence

In this example, images of a toy castle model are captured using a Canon D-30 digital camera with all camera intrinsic parameters fixed. The castle model has dimensions measuring less than a rectangular box of size 25 cm×9 cm×16 cm ($W\times D\times H$) and is imaged with a checker-board calibration pattern

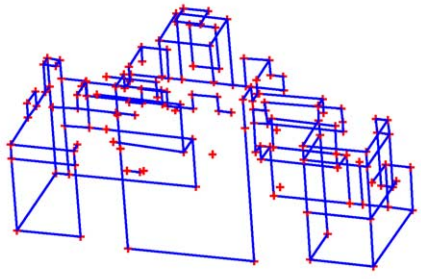
(a) An image from the castle model image sequence



(b)The distribution of 2D missing points



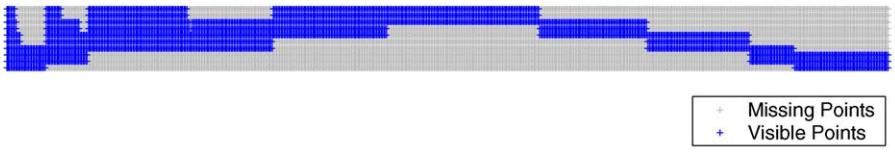
(c) A view of the Euclidean reconstruction of the 3D scene



(d) A close-up of the reconstructed castle model

Fig. 3. Castle model image sequence.

(a) The first view of the Model House image sequence



(b) The distribution of 2D missing points

(c) A view of the Euclidean reconstruction of the 3D scene

(d) Side view of `Model House'

Fig. 4. Model house image sequence.

(which however is not used) in the background. The image size is 2160×1440 (pixels). Five images (including three front and two rear views) are taken around the castle at an elevation angle of around 30°. The rear views are captured after turning the castle 180° on the calibration background. Fig. 3(a) shows the first image of the sequence. A total of 154 corresponding points are matched manually across the five images. Manual matching enables us to select the specific points for reconstruction, so that the reconstructed castle can be represented by a wire-frame for easy visualization. None of the points are visible in all five images and most of them are visible only in two views. Fig. 3(b) shows the missing-data map of the measurement matrix where each element w_{ij} is marked at the (i, j)th position as the symbol '+' in blue if it corresponds to a visible point and as a light grey '+' if it corresponds to missing data. There are 378 missing points among the five views (i.e. the percentage of missing points is 49.1%).

Our implementation of the algorithm of Section 3 takes 533 iterations to converge (requiring 12.4 s on a 2.4 GHz PC Pentium-4). The maximum 2D reprojection error (for visible points only) among all the views is 3.53 pixels and the RMS 2D reprojection error (also for visible points only) is 1.05 pixels.

To visualize the reconstruction, we use the normalization method of Han and Kanade [17,18] to upgrade the projective space to the Euclidean space, assuming that the principal point is fixed and the skew ratio is zero across all views. A view of the scene of the reconstructed Euclidean space is shown in Fig. 3(c) and a close-up in Fig. 3(d).

4.2.2. Model house

In this example, we make use of the image sequence and correspondences of a 'Model House' (courtesy of the Visual Geometry Group at University of Oxford). There are 10 images of size 768×576. Fig. 4(a) shows the first image in the sequence and Fig. 4(b) shows the missing-data map of the image sequence. There are a total of 672 3D object points matched automatically, and the percentage of missing data is 57.6%. Our proposed method takes 2000 iterations to converge in 225 s. The RMS 2D reprojection error for visible points is 0.72 pixels and the maximum 2D reprojection error is 9.07 pixels. Using the normalization method of [17] to upgrade the projective space to the Euclidean space, a view of the upgraded Euclidean space is shown in Fig. 4(c). It can be seen that the reconstructed cameras are almost in circular motion. A side view of the reconstructed 3D points is given in Fig. 4(d), which shows that the angle between the front wall of the house and the

ground is approximately 90°. Despite some mis-matched corresponding points in this image sequence, the proposed method is sufficiently robust to converge to a reasonable solution.

5. Conclusion

In this paper, we have used the subspace approach to consider the factorization problem for projective reconstruction from multiple images. The novelty of our method lies in a formulation of the factorization problem in terms of the optimization of a single objective function that is minimized in a consistent manner with respect to three distinct sets of parameters for different purposes. As a result, we are able to estimate missing data as part of the algorithm while preserving guaranteed convergence of the iterative solution. The ability to handle missing points greatly enhances the applicability of the factorization method to real images. Simulation results (including examples (see [16]) not shown here because of reasons of space) and comparisons with other methods shows that the subspace method produces reconstructions that are superior in terms of 2D reprojection errors. The solution obtained using the proposed method can however be further refined using bundle adjustment [8,14,15] to minimize the 2D reprojection error.

Acknowledgement

The work described in this paper was supported by grants from the Research Grants Council of Hong Kong Special Administrative Region, China (Project Nos. HKU7058/02E and HKU7135/05E).

References

[1] R.I. Hartley, Euclidean reconstruction from uncalibrated views, in J. Mundy, A. Zisserman (Eds.), Applications of Invariance in Computer Vision vol. LNCS 825, Springer, Berlin, 1993, pp. 237–256.

[2] O.D. Faugeras, Stratification of three-dimensional vision: projective, affine, and metric representations, Journal of the Optical Society of America-A 12 (3) (1995) 465–484.

[3] J. Oliensis, V. Govindu, An experimental study of projective structure from motion, IEEE Transactions Pattern Analysis and Machine Intelligence 21 (7) (1999) 665–671.

[4] P. Sturm, B. Triggs, A factorization based algorithm for multi-image projective structure and motion, European Conference on Computer Vision, Cambridge, UK, 1996, pp. 709–720.

[5] B. Triggs, Factorization methods for projective structure and motion, IEEE International Conference on Computer Vision and Pattern Recognition, IEEE, San Francisco, 1996, pp. 845–851.

[6] S. Mahamud, M. Hebert, Y. Omori, J. Ponce, Provably-convergent iterative methods for projective structure from motion, IEEE International Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii 2001, pp. 1018–1025.

[7] S. Mahamud, M. Hebert, Iterative projective reconstruction from multiple views, IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 2000, pp. 430–437.

[8] W.K. Tang, Y.S. Hung, A factorization-based method for projective reconstruction with minimization of 2D reprojection errors, in: L.V. Gool (Ed.), Proceedings of 24th Annual Pattern Recognition Symposium DAGM 2002, vol. LNCS-2449, Springer, Zurich, Switzerland, 2002, pp. 387–394. URL available from: http://link.springer.de/link/service/series/0558/papers/2449/24490387.pdf.

[9] W.K. Tang, Y.S. Hung, A column-space approach to projective reconstruction, Computer Vision and Image Understanding 101 (3) (2006) 166–176.

[10] A. Heyden, R. Berthilsson, G. Sparr, An iterative factorization method for projective structure and motion from image sequences, Image and Vision Computing 1713 (1999) 981–991.

[11] G. Sparr, Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences, IEEE International Conference Pattern Recognition 1996.

[12] G.Q. Chen, G.G. Medioni, Efficient iterative solutions to m-view projective reconstruction problem, IEEE International Conference on Computer Vision and Pattern Recognition, vol. II, 1999, pp. 55–61.

[13] G.Q. Chen, G.G. Medioni, Practical algorithms for stratified structure-from-motion, Image and Vision Computing 20 (2002) 103–123.

[14] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment—A modern synthesis, in: W. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice, Vol. LNCS 1883, Springer, Berlin, 2000, pp. 298–375.

[15] Y.S. Hung and W.K. Tang, Projective Reconstruction from Multiple Views with Minimization of 2D Reprojection Error, International Journal of Computer Vision, 66(3) (2006) 305–317 doi:10.1007/s11263-005-3675-0.

[16] W.K. Tang, A factorization-based approach to 3D reconstruction from multiple uncalibrated images, PhD thesis, The University of Hong Kong, Department of Electrical and Electronic Engineering, 2004.

[17] M. Han, T. Kanade, Scene reconstruction from multiple uncalibrated views, Tech. Rep. CMU-RI-TR-00-09, Robotics Institute, Carnegie Mellon University (January 2000).

[18] M. Han, T. Kanade, Multiple motion scene reconstruction from uncalibrated views, IEEE International Conference Computer Vision, vol. 1, 2001, pp. 163–170.