

Reconstructing Camera Projection Matrices from Multiple Pairwise Overlapping Views

Jacob Goldberger

The Weizmann Institute of Science, Rehovot, 76100 Israel

my current address is:

Department of Computer Science, University of Toronto,

Toronto M5S 3G4, Canada

jacob@cs.toronto.edu

Abstract

In this work we address the problem of projective reconstruction from multiple views with missing data. Factorization based algorithms require point correspondences across all the views. In many applications this is an unrealistic assumption. Current methods that solve the problem of projective reconstruction with missing data require correspondence information across triplets of images. We propose a projective reconstruction method that yields a consistent camera set given the fundamental matrices between pairs of views without directly using the image correspondences. The algorithm is based on breaking the reconstruction problem into small steps. In each step, we eliminate as much uncertainty as possible.

Key words: structure from motion, projective reconstruction, multiple view geometry, linear reconstruction techniques

1 Introduction

$3D$ reconstruction from multiple views is a central problem in computer vision. Applications range from a precise measurement system with several fixed cameras to approximate structure and motion from real-time video for active robot navigation. To accomplish these tasks, we can ignore the issue of camera calibration and reconstruct the scene only up to an unknown global projective transformation. Later, if necessary for the task, we can use metric information about the $3D$ object to obtain Euclidean reconstruction. In this paper we concentrate on the projective reconstruction step. We address the problem of obtaining a consistent camera set from available fundamental matrices between pairs of views.

Traditional projective reconstruction methods are applied to two three or four views. $3D$ structure of a scene can be recovered up to an unknown projective transformation, where the camera geometry can be represented by the fundamental matrix, the trifocal or the quadrifocal tensor respectively [7]. Generalizations of tensorial constraints to multiple views have been treated by many researchers e.g. [20,4,10]. Tomasi and Kanade [19] and Ullman and Basri [23] considered the case of multiple affine camera matrices. They utilized the singular value decomposition to factor the image point matrix and thus to obtain the structure and the affine cameras. Kahl and Heyden [12] proposed a more general approach that uses closure constraints. To apply the factorization method in a projective setup, first the relative depth of each image point must be found [18]. Several variants of the factorization method that can be applied to projective cameras have recently been suggested [18,17,9,21,15]. In the case of affine cameras, the factorization method yields the maximum

likelihood (ML) estimation (assuming Gaussian image error). In the case of perspective cameras, however, the solution obtained from factorization-like methods has no geometrical meaning [5]. Another major drawback of the factorization method is the requirement of point correspondences across all the views, i.e. each object point must be visible in every image. This is an unrealistic assumption both in the case of an object viewed from several directions and in a video sequence of a dynamic scene.

A commonly used method (with remarkable success) is bundle adjustment. The bundle adjustment method modifies both the $3D$ structure and the cameras in order to minimize the reprojection error. If the image error is Gaussian, the bundle adjustment yields the ML estimation. There is no closed-form formula for the ML structure and motion. Instead, iterative methods should be utilized to perform the non-linear optimization. Hence a good initial solution is required to ensure convergence to the global maximum. Current methods that find an initial solution for the problem of projective reconstruction with missing data (in the sense that not every point is seen in every image) require correspondence information across triplets of images. One approach is to perform reconstruction sequentially. The trifocal tensor, computed from matching points across triplet of views, integrates each new image into the current reconstruction [2,5,1]. Another approach is to sequentially collect constraints from image triplets. Then the camera set is obtained as a null space of the constraint matrix [22,11,16].

In many common situations we cannot obtain point correspondences across three (or more) images. For example, consider four cameras arranged in a circle with a $3D$ object located in the middle [13] (see Figure 1). The four pictures taken by the cameras only overlap in a pairwise manner. A more complicated

example is the following (see Figure 2). Consider a $3D$ object placed in the middle of the room and surrounded by eight cameras. Each camera is located at a distinct corner of the room and views a different portion of the object. Each part the object is only viewed by two cameras. Each camera produces an image that overlaps with the images produced by the three neighboring cameras (see Figure 2). In these two examples, the only available information is point correspondences across pairs of views. The method described in [16] can be modified to handle image sets with only pairwise correspondences and therefore it can in principle solve both cases of Figures 1 and 2. The method we propose, however, does not use the correspondences but only the fundamental matrices computed from the correspondences. A similar problem was recently addressed by Levi and Werman [14]. They proposed a method to compute missing fundamental matrices from known fundamental matrices for up to six views.

In this paper we address the problem of structure from motion, given multiple views in which point correspondences are only available between (some of the) pairs of views. The paper presents a procedure that yields a consistent camera set given a subset of the fundamental matrices between pairs of views. Our method differs from [9–11,22] and others in that given the fundamental matrices, we do not need to have any image correspondences in order to construct the projection matrices. Our method is related to Triggs’ closure relations [20,22]. In both cases the basic constraint on the projections is the closure relations between the projection matrices and their matching tensor. Triggs proposed that the closure relations should first be found, and then all the equations should be gathered into a big linear system and finally all the projections should be solved at once. This offers the advantage of having ev-

everything handled uniformly and the disadvantage of yielding a system matrix that is large (but sparse). However, in the reconstruction problems analyzed in this paper, we can not find all the closure relations in advance. Instead, we take a more local approach and eliminate as much as possible between each pair of images. We solve many small systems rather than one large one. Another related difference is that Triggs computed the scales by following input points around each loop (so for each loop, we have to see at least one input point in all its images). We use just the fundamental matrices and epipoles, hence there is no need for the image points to be visible around the whole loop.

The paper is organized as follows. Section 2 presents a basic relation between two camera matrices that are consistent with the fundamental matrix. Section 3 applies the results of Section 2 to obtain projective reconstruction entities that are independent of the cameras' coordinate system. Section 4 utilizes the invariants defined in Section 3 to obtain a linear projective reconstruction algorithm given the fundamental matrices between all pairs of views. In Section 5 we address the problem of missing data and suggest a reconstruction method that can be applied even for cases where point matchings across image triplets are not available. Experimental results are presented in Section 6.

2 The Geometry of Two Views

In this section we summarize the geometric relation between two views. Given an arbitrary 3×4 matrix for the first camera and the fundamental matrix, we derive a parametric expression for all the consistent second camera matrices. Assume a scene is viewed by two cameras. Using a pin hole camera model, we

consider the operation of each camera as a projective transformation from \mathbf{P}^3 to \mathbf{P}^2 that is represented by a 3×4 matrix. Denote the two camera matrices by $P_i = (A_i \ v_i)$, $i = 1, 2$ such that A_i is a 3×3 matrix and v_i is a 3×1 vector. The fundamental matrix F between the two views is the following matrix: $F = [v_2 - A_2 A_1^{-1} v_1] A_2 A_1^{-1} ([x]$ is the skew-symmetric matrix representing the cross product by x). The epipole $v = v_2 - A_2 A_1^{-1} v_1$ is the left eigenvector of F that corresponds to the eigen value 0 (i.e. $v^\top F = 0$). The fundamental matrix F encapsulates the epipolar geometry between the two views. Each pair of corresponding image points p_1 and p_2 satisfies the equation $p_2^\top F p_1 = 0$.

Suppose that the two camera matrices are unknown and the only available information is point correspondences across the two views. F can be computed in a linear manner from eight (or more) pairs of corresponding points [8]. F is uniquely determined up to a scalar factor. The equation $p_2^\top F p_1 = 0$ implies that two camera matrices P_1 and P_2 can be the cameras that created the two images if $X^\top P_2^\top F P_1 X = 0$ holds for every $X \in \mathbf{P}^3$. In this case, we say that P_1 and P_2 are consistent with F .

Definition: Let F be the fundamental matrix between two views, and let v be the epipole (i.e. $v^\top F = 0$). Assume that F and v are normalized such that $\|F\| = 1$ and $\|v\| = 1$. The relative camera matrix between the first image and the second one is the matrix $P_{12} = ([v]F \ v)$. Note that the canonic camera matrix $(I \ 0)$ and the relative camera matrix P_{12} form a consistent pair of camera matrices.

Theorem 1: Let F be the fundamental matrix between two views, let v be the epipole and let P_{12} be the relative camera matrix defined by F and v . Two 3×4 matrices P_1 and P_2 are consistent with F if and only if there are a scalar

$\beta_{12} \neq 0$ and a four dimensional vector r_{12} such that:

$$P_2 = P_{12} \begin{pmatrix} \beta_{12}P_1 \\ r_{12}^\top \end{pmatrix} \quad (1)$$

Proof: Assume that P_1 and P_2 satisfy relation (1). To prove that P_1 and P_2 are consistent with F we have to show that $X^\top P_2^\top F P_1 X = 0$ for every $X \in \mathbf{P}^3$.

$$\begin{aligned} X^\top P_2^\top F P_1 X &= X^\top (\beta_{12}P_1^\top \ r_{12}) ([v]F \ v)^\top F P_1 X \\ &= \beta_{12}(F P_1 X)^\top [v]^\top F P_1 X + X^\top r_{12} v^\top F P_1 X \end{aligned}$$

utilizing the assumption the $v^\top F = 0$ and the fact that the matrix $[v]$ is skew-symmetric, we obtain the desired result.

Assume now that P_1 and P_2 are consistent with F . We shall show that we can find a scalar β_{12} and a vector r_{12} such that (1) is fulfilled. We noted that $(I \ 0)$ and P_{12} form a consistent pair of camera matrices. A projective reconstruction is unique up to an overall projective transformation [3]. Hence, there exists a 4×4 non-singular matrix H such that $\beta_{12}P_1 = (I \ 0)H$ and $P_2 = P_{12}H$. A matrix H that satisfies the equation $\beta_{12}P_1 = (I \ 0)H$ must be in the form:

$$H = \begin{pmatrix} \beta_{12}P_1 \\ r_{12}^\top \end{pmatrix} \quad (2)$$

such that r_{12} is a four dimensional vector that is determined by the equation $P_2 = P_{12}H$. Substituting (2) in the equation $P_2 = P_{12}H$ yields relation (1). \square

Observe that if P_1 and P_2 satisfy equation (1), then β_{12} and r_{12} are uniquely

determined by P_1 , P_2 and P_{12} . For example, in the case of the camera pair $(I \ 0)$ and P_{12} , these values are $\beta_{12} = 1$ and $r_{12}^\top = (0001)$.

From relation (1) we can deduce the bilinear closure relation between the fundamental matrix and the cameras matrices defined by Triggs [22]. Assume that P_1 and P_2 are consistent with F . Multiplying both sides of equation (1) by the skew-symmetric matrix $[v]$, yields: $[v]P_2 = \beta_{12}[v][v]FP_1$. Note that $[v][v]F$ is equal to F up to a scale factor. Hence we derive the bilinear closure relation: $[v]P_2 = \gamma_{12}FP_1$ such that γ_{12} is a scale factor that depends on the normalization of F and v .

3 The β -Coefficients Matrix of n Views

Assume we are given n views of a scene without any cameras calibration information. For each pair of views we can extract the relative camera matrix from the image data. Denote the relative camera matrix between images i and j by $P_{ij} = (A_{ij}, v_{ij}) = ([v_{ij}]F_{ij}, v_{ij})$. P_{ij} is determined from the corresponding points between the images (and by the normalization convention $\|v_{ij}\| = 1, \|F_{ij}\| = 1$).

Given a consistent camera set $P = \{P_1, \dots, P_n\}$, the following relation exists between each pair of views:

$$P_j = P_{ij} \begin{pmatrix} \beta_{ij} P_i \\ r_{ij}^\top \end{pmatrix} \quad i, j = 1, \dots, n \quad (3)$$

Given the camera matrices P_i, P_j, P_{ij} , the scale factor β_{ij} and the four dimen-

sional vector r_{ij} are uniquely determined. To include the case $i = j$, define $P_{ii} = (I \ 0)$. This implies that $\beta_{ii} = 1$ and $r_{ii}^\top = (0001)$. Note that if all the scale factors β_{ij} are known, then the equations (3) form a linear system with the unknown camera matrices.

We can capture all the scale factors between the pairs of camera matrices by defining the following $n \times n$ matrix:

$$\beta(P) = \begin{pmatrix} \beta_{11} & \dots & \beta_{1n} \\ \vdots & & \vdots \\ \beta_{n1} & \dots & \beta_{nn} \end{pmatrix} \quad (4)$$

The notation $\beta(P)$ emphasizes that the matrix β depends on the camera set P . We shall show, however, that in some sense the matrix β does not depend on a specific consistent cameras set.

Suppose that $\tilde{P} = \{\tilde{P}_1, \dots, \tilde{P}_n\}$ is another camera matrix set that is consistent with the fundamental matrices extracted from the image data. Since reconstructing a scene can be done uniquely up to a global projective transformation [3], \tilde{P} can be written in the following form:

$$\tilde{P} = \{\alpha_1 P_1 H, \dots, \alpha_n P_n H\} \quad (5)$$

such that H is a non-singular 4×4 matrix (representing a projective transformation of \mathbf{P}^3) and $\alpha_1, \dots, \alpha_n$ are non-zero scalars.

Lemma: Let P and \tilde{P} be two consistent camera sets. The elements of the two

matrices $\beta(P)$ and $\beta(\tilde{P})$ satisfy the following equation:

$$\beta_{12}(P)\beta_{23}(P) \cdot \dots \cdot \beta_{n-1,n}(P)\beta_{n1}(P) = \beta_{12}(\tilde{P})\beta_{23}(\tilde{P}) \cdot \dots \cdot \beta_{n-1,n}(\tilde{P})\beta_{n1}(\tilde{P})$$

Proof: Multiplying both sides of equation (3) by $\alpha_j H$ and substituting the representation of \tilde{P} according to equation (5) yields the relation: $\beta_{ij}(\tilde{P}) = \alpha_j \beta_{ij}(P) \alpha_i^{-1}$. Hence:

$$\begin{aligned} \beta_{12}(\tilde{P})\beta_{23}(\tilde{P}) \cdot \dots \cdot \beta_{n-1,n}(\tilde{P})\beta_{n1}(\tilde{P}) &= \frac{\alpha_2}{\alpha_1} \beta_{12}(P) \cdot \dots \cdot \frac{\alpha_n}{\alpha_{n-1}} \beta_{n-1,n}(P) \frac{\alpha_1}{\alpha_n} \beta_{n1}(P) \\ &= \beta_{12}(P)\beta_{23}(P) \cdot \dots \cdot \beta_{n-1,n}(P)\beta_{n1}(P) \end{aligned}$$

The lemma trivially implies that for each ordered subset $\{j_1, \dots, j_k\}$ of the index set $\{1, \dots, n\}$, the expression $\beta_{j_1, j_2} \beta_{j_2, j_3} \cdot \dots \cdot \beta_{j_{k-1}, j_k} \beta_{j_k, j_1}$ is independent of the cameras coordinate system. In other words, the product of scale factors along a closed circle of views is independent of the cameras' coordinate system. Note that the determinant of the matrix β can be written as a sum of $n!$ elements in the form of $\beta_{12}\beta_{23} \cdot \dots \cdot \beta_{n-1,n}\beta_{n1}$. Hence also $\det(\beta)$ is independent of the coordinate system.

As an example, we show how the entity $\beta_{ij}\beta_{ji}$ can be computed from the fundamental matrix between the images i and j . The camera pair $P_i = (I \ 0)$ and $P_j = P_{ij}$ is a consistent camera set (with $\beta_{ij} = 1$). Hence:

$$P_j = P_{ij} \begin{pmatrix} I & 0 \\ & \\ & \\ 0001 \end{pmatrix}, \quad (I \ 0) = P_{ji} \begin{pmatrix} \beta_{ji} P_j \\ & \\ & \\ r_{ji}^\top \end{pmatrix} \quad (6)$$

Substituting the first equation of (6) in the second one and multiplying both

sides by $[v_{ji}]$ yields the bilinear closure relation :

$$[v_{ji}] = \beta_{ji}\beta_{ij}[v_{ij}]A_{ji}A_{ij} \quad (7)$$

where $A_{ij} = [v_{ij}]F_{ij}$. The scalar $\beta_{ji}\beta_{ij}$ is obtained as a total least squares [6] solution of equation (7).

As another example, we show how the invariant $\beta_{12}\beta_{23}\beta_{31}$ can be computed from the fundamental matrices between the three images. We can assume that $P_1 = (I \ 0)$ and $P_2 = P_{12}$. This assumption implies that $\beta_{12} = 1$. A matrix P_3 that is consistent with P_1 and P_2 must satisfy the following two relations (P_3 can be scaled such that $\beta_{31} = 1$).

$$P_3 = P_{23} \begin{pmatrix} \beta_{23}P_{12} \\ r_{23}^\top \end{pmatrix}, \quad (I \ 0) = P_{31} \begin{pmatrix} P_3 \\ r_{31}^\top \end{pmatrix} \quad (8)$$

Substituting the first equation of (8) in the second one yields:

$$(I \ 0) = P_{31} \begin{pmatrix} P_{23} \begin{pmatrix} \beta_{23}P_{12} \\ r_{23}^\top \end{pmatrix} \\ r_{31}^\top \end{pmatrix} \quad (9)$$

We can derive, in a manner similar to equation (7), the following linear equations with the unknown scalar $\beta_{31}\beta_{23}\beta_{12}$:

$$[[v_{31}]A_{31}v_{23}][v_{31}] = \beta_{31}\beta_{23}\beta_{12}[[v_{31}]A_{31}v_{23}][v_{31}]A_{31}A_{23}A_{12} \quad (10)$$

4 A Factorization-Like Reconstruction Method

Before delving into the problem of missing data, we first demonstrate in this section how the invariants, developed in the previous section, can be used to obtain a linear reconstruction algorithm given the relative camera matrices. The algorithm for the case of no missing data is similar to the reconstruction method proposed by Triggs [22]. Theorem 1 implies that a consistent camera set $P = \{P_1, \dots, P_n\}$, should satisfy the following equations:

$$P_j = P_{ij} \begin{pmatrix} \beta_{ij} P_i \\ r_{ij}^\top \end{pmatrix} \quad i, j = 1, \dots, n \quad (11)$$

or alternatively:

$$[v_{ij}]P_j = \beta_{ij}[v_{ij}]A_{ij}P_i \quad (12)$$

such that P_i , β_{ij} and r_{ij} are unknown and $P_{ij} = (A_{ij}, v_{ij}) = ([v_{ij}]F_{ij}, v_{ij})$ is the relative camera matrix derived from the corresponding points between the i and j images (and by the normalization convention $\|v_{ij}\| = 1, \|F_{ij}\| = 1$). If all the scale factors β_{ij} are known, the system (11) is linear. Hence the scale factors β_{ij} play the same role that projective depth plays in factorization methods. Given the projective depth, the image points matrix become bilinear in the structure and the cameras. The scale factors can be computed in the following manner. Given a consistent camera set $P = \{P_1, \dots, P_n\}$, we can assume, without any loss of generality, that

$$\beta_{12} = \beta_{13} = \dots = \beta_{1n} = 1 \quad (13)$$

The assumption $\beta_{1i} = \beta_{1j} = 1$ trivially implies the relation:

$$\beta_{ij} = \frac{\beta_{1i}\beta_{ij}\beta_{j1}}{\beta_{1j}\beta_{j1}} \quad (14)$$

It was proved in the previous section that both the numerator and the denominator are independent of the cameras' coordinate system. The scalar $\beta_{1j}\beta_{j1}$ can be linearly extracted, using equation (7) and the scalar $\beta_{1i}\beta_{ij}\beta_{j1}$ can be linearly computed from equation (10).

The terms $\beta_{1i}\beta_{ij}\beta_{j1}$ and $\beta_{1j}\beta_{j1}$ were computed for different consistent cameras sets. However since they are invariants, they can be plugged into equation (14) to obtain β_{ij} that is consistent with equation (13). Given all the scale factors β , the system (11) (or (12)) can be linearly solved. The camera set solution is the null space of equation (12). The solution obtained from the algorithm presented here has no geometric interpretation. However, it can be used as a starting point for iterative methods (such as bundle adjustment) that maximize the likelihood function. Note that we did not make full use of the assumption that no data is missing. We have only assumed that there is a reference image that the fundamental matrices between one reference view and all other views can be estimated from image correspondences. A similar setup was used in [24] to perform a projective reconstruction from trifocal tensors extracted from triplets of views. The advantage of our method is that it can solve even those situations when there are only pairwise correspondences between the views and no triples of correspondences are available. Moreover we do not need to return to the data once having the fundamental matrices.

5 Reconstruction with Missing Data

In this section we address the case of a scene observed by multiple cameras with missing data in the sense that there are not enough points between all views to determine fundamental matrices between all pairs of views. We first analyze the relatively simple case where each image is related to two (or more) previous images. The reconstruction is performed in a sequential manner. Assume the camera matrices P_1, \dots, P_{t-1} were already found and we want to find a consistent camera matrix for the t -th image. Given point correspondences across the image triplet indexed $t-2, t-1, t$ we can use the trifocal tensor that was extracted from the data to compute P_t in the following manner. If P_{t-2} is the canonic matrix $(I \ 0)$, then the trifocal tensor has the form : $T_i^{jk} = (v_t)^j (A_{t-1})_i^k - (v_{t-1})^k (A_t)_i^j$ such that $P_{t-1} = (A_{t-1}, v_{t-1})$ and $P_t = (A_t, v_t)$. Since P_{t-1} is known, P_t can be extracted from the trifocal tensor formula by solving a linear system. For a general matrix P_{t-2} we first find a 4×4 matrix H such that $P_{t-2}H = (I \ 0)$. The trifocal tensor expression gives a relation between $P_{t-1}H$ and P_tH and, as before, P_tH is obtained as a solution of a linear system. Multiplying the solution by H^{-1} yields the desired camera matrix P_t . In the presence of noise, since there are more equations than variables, a least-squares approach should be used. After each time a new view is integrated into the reconstruction, a bundle adjustment step can be performed to distribute the reconstruction error uniformly over all the views [5].

We shall now discuss the more complicated case where we cannot arrange the images in a manner such that each image overlaps with two or more previous images. We first demonstrate the algorithm on the example, illustrated in

Figure 2, of eight cameras placed in eight corners of a cube. The camera numbers that appear in the sequel are related to the numbering in Figure 2. To initialize the reconstruction we set $P_1 = (I \ 0)$ and $P_2 = P_{12}$. P_3 overlaps only with P_2 . Therefore, all we can know is that there exists a four dimensional vector r_{23} such that:

$$P_3 = P_{23} \begin{pmatrix} P_2 \\ r_{23}^\top \end{pmatrix} \quad (15)$$

but r_{23} is still undetermined. P_4 is related both to P_1 and P_3 . Hence

$$P_4 = P_{34} \begin{pmatrix} P_3 \\ r_{34}^\top \end{pmatrix} = P_{14} \begin{pmatrix} \beta_{14} P_1 \\ r_{14}^\top \end{pmatrix} \quad (16)$$

Substituting (15) in (16) yields:

$$P_{34} \begin{pmatrix} P_{23} \begin{pmatrix} P_2 \\ r_{23}^\top \end{pmatrix} \\ r_{34}^\top \end{pmatrix} = P_{14} \begin{pmatrix} \beta_{14} P_1 \\ r_{14}^\top \end{pmatrix} \quad (17)$$

The linear system (17) consists of twelve equations with thirteen unknowns. Therefore (unless the four cameras are coplanar) the camera matrices P_3 and P_4 can be uniquely defined up to an unknown scalar denoted by x . Hence P_3 and P_4 can be written in the form $P_3 = P_3^0 + P_3^1 x$ and $P_4 = P_4^0 + P_4^1 x$ such that $P_3^0, P_3^1, P_4^0, P_4^1$ are 3×4 coefficients matrices obtained from solving equation (17). In a similar manner, given P_1 and P_2 we can find P_5 and P_6 up

to a scalar denoted by y , i.e. $P_5 = P_5^0 + P_5^1 y$ and $P_6 = P_6^0 + P_6^1 y$. The camera P_7 is related both to P_3 and P_5 . Hence

$$P_7 = P_{57} \begin{pmatrix} P_5^0 + yP_5^1 \\ r_{57}^\top \end{pmatrix} = P_{37} \begin{pmatrix} \beta_{37}P_3^0 + \beta_{37}xP_3^1 \\ r_{37}^\top \end{pmatrix} \quad (18)$$

Considering the product $\beta_{37}x$ as a new variable, the system (18) consists of twelve linear equations with eleven unknowns. While the solution of (18) reveals P_7 , it also eliminates the uncertainty that exists regarding the camera matrices P_3, P_4, P_5, P_6 . The last camera P_8 is related to P_4, P_6 and P_7 which were already found. P_8 can, therefore, be extracted from the following equation :

$$P_8 = P_{78} \begin{pmatrix} P_7 \\ r_{78}^\top \end{pmatrix} = P_{68} \begin{pmatrix} \beta_{68}P_6 \\ r_{68}^\top \end{pmatrix} = P_{48} \begin{pmatrix} \beta_{48}P_4 \\ r_{48}^\top \end{pmatrix}$$

This completes the procedure for obtaining a consistent camera set for the camera configuration illustrated in Figure 2.

The algorithm to obtain a consistent camera set for the general case is the following. The reconstruction algorithm is initialized by setting $P_1 = (I \ 0)$ and $P_2 = P_{12}$. There is a degree of freedom in the reconstruction of an overall projective transformation. This setting fixes one of the equivalent solutions. For $t = 3, \dots, n$ we merge the t -th view into the current reconstruction in the following way. Assume there is (pairwise) overlapping between the t -th view and m previous views indexed by s_1, \dots, s_m . If $m = 1$ all we can know is that

there exists a four dimensional vector $r_{s_1,t}$ such that:

$$P_t = P_{s_1,t} \begin{pmatrix} P_{s_1} \\ r_{s_1,t}^\top \end{pmatrix} \quad (19)$$

but $r_{s_1,t}$ remains undetermined. At this step it is possible that P_{s_1} is already completely determined. It is also possible that P_{s_1} is represented by a linear expression that includes undetermined variables. In both cases, from (19) we derive a linear expression for P_t . If $m > 1$, i.e. the t -th view is related to two or more previous views, we can solve some of unknown variables. The following relations exist between the m previous views and the current one:

$$P_t = P_{s_j,t} \begin{pmatrix} \beta_{s_j,t} P_{s_j} \\ r_{s_j,t}^\top \end{pmatrix} \quad j = 1, \dots, m \quad (20)$$

We can assume that $\beta_{s_1,t} = 1$. From relations (20) we can extract $m - 1$ equations:

$$P_{s_1,t} \begin{pmatrix} P_{s_1} \\ r_{s_1,t}^\top \end{pmatrix} = P_{s_j,t} \begin{pmatrix} \beta_{s_j,t} P_{s_j} \\ r_{s_j,t}^\top \end{pmatrix} \quad j = 2, \dots, m \quad (21)$$

If P_{s_j} is not completely known, then $\beta_{s_j,t} P_{s_j}$ is not a linear expression. Suppose that the linear expression for P_{s_j} has the form: $P_{s_j} = P_{s_j}^0 + P_{s_j}^1 x_1 \dots + P_{s_j}^k x_k$. The term $\beta_{s_j,t} P_{s_j}$ can be linearized by replacing the $k + 1$ variables $\beta_{s_j,t}, x_1, \dots, x_k$ by another set of $k + 1$ variables $\beta_{s_j,t}, x_1 \beta_{s_j,t}, \dots, x_k \beta_{s_j,t}$.

The solution of the linear system (21) either reveals P_t completely or constrains P_t to belong to an affine subspace. In cases where P_t is not completely

determined, we derive a linear expression for P_t that can be solved using the overlapping between the t-th view and views that will be processed later. The algorithm operates given the following assumption. If at least one of the camera matrices that appears in (21) (i.e. P_{s_1}, \dots, P_{s_m}) is not completely determined then there are enough equations in (21) to obtain a unique solution for P_t and also to eliminate the uncertainty about the camera matrices P_{s_1}, \dots, P_{s_m} . Given this assumption there is no need to relate variables across distinct cameras. The example of eight cameras satisfies this assumption. The only relevant situation occurs during the computation of P_7 where the cameras P_3 and P_5 that appear in equation (18) are not completely determined. The solution of (18) eliminates the uncertainty about P_3 and P_5 .

There may be degrees of freedom in the reconstruction (apart from the overall projective reconstruction). In this case even at the end of the process some of the unknowns are not resolved. For example, in the case of a circle of four pairwise overlapping views there is one extra degree of freedom [13].

6 Experimental Results

In the first experiment, we tested the accuracy and stability of the reconstruction algorithm with respect to noise in the image data. The camera set is composed of eight cameras placed in the corners of a 2 meter cube (see Figure 2). The artificial scene corresponds to a set of 50 3D points distributed uniformly inside a cube of size 0.4 meter that is placed in the center of the cameras cube. The image size is 1000×1000 pixels. We are not aware of any reconstruction algorithm that can successfully handle the cube structured cameras assuming that no points in images are used to help the construction

of the camera matrices from the fundamental known matrices. For comparison, two alternative reconstruction algorithms (that can be applied only if more information is given) were also implemented. In the first alternative case we assumed that we can extract the fundamental matrices between all the camera pairs. In the second situation we assumed the "parallel F-e chain" [22] camera structure. In this structure each view is related to the first two views (see figure 3). For these two cases we first computed the scale factors using the fundamental matrices and then gathered all the equations with the eight unknown camera matrices into a one linear system. We performed 100 experiments for a given value of the deviation of noise. The result of each experiment is the mean squares reprojection error. Since the noise is Gaussian, this quantity is also the likelihood of the reconstruction. The results are summarized in Figure 4. The best results are obtained, as expected, in the case where the views are fully overlapped and all the fundamental matrices can be computed from the data. From Figure 4 we can conclude that in the case of a cube shaped camera set, in spite of the fact that we are computing the scale factors from longer loops of views, a projective reconstruction can still be performed on a noisy set without much degradation when compared to the two cases where redundant information is available. Hence the proposed algorithm can be used as an initial solution for the iterative bundle adjustment algorithm.

Real data experiments were conducted with an image sequence consisting of four frames that are shown in Figure 5. Each frame is a picture, taken from a distinct corner, of two toy cars placed in the middle of the room. It can be seen from Figure 5 that most point correspondences can be found only in consecutive views. For each pair of consecutive views, 15 point correspondences were

manually selected. In the first step, the fundamental matrices and the epipoles were computed [8]. Next, equation (16) is used to obtain a one dimensional space of possible consistent four cameras (apart from the degree of freedom of a 3D projective transformation). The average 2D distance between the data points and the reprojection of the reconstructed points is 1.72 pixels. After applying the bundle adjustment to one of the possible solutions, the average distance decreased to 0.63 pixels.

7 Discussion

In this paper we proposed a method for projective reconstruction from multiple views with missing data. Compared to previous methods, this method can handle severe occlusion problems where the overlapping between cameras is minimal, namely point correspondence can only be found across two views. The proposed method yields camera matrices which are consistent with the set of computed fundamental matrices between pairs of views. The method is based on the closure relations originally defined by Triggs. The paper presents a different way to solve the problem. In particular a new way to tackle the computation of the projective depths. The new method avoids the need for any image points to be visible around the whole loop of views (Triggs' method required that at least one point be visible in all the views). Our method uses only fundamental matrices to construct the set of consistent projection matrices and does not use point reconstructed from different image pairs to find the relationships between coordinate systems of the cameras that took the pairs. Therefore, in our method no correspondences need to be existing among more than two views. It should be stated, however that if correspondences among

more views is available than other methods (e.g. [1,22,24]) allow for an efficient and more robust integration of overlapping pairs. The proposed method also breaks the problem into small steps instead of imposing the closure relations for all the sequence. This is particularly useful when the closure relations are not all known from the start. In this case our method enables constructing the closure relations on-line. Another feature is that the projective reconstruction is performed sequentially. The method is particularly well suited to the case of fixed cameras that are located far apart where the view points are very distinct and there is not enough overlap between images to use trilinear matching tensor.

References

- [1] S. Avidan and A. Shashua, Threading fundamental matrices, *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol 23(1), 2001, pp 73-77.
- [2] P. Beardsley, A. Zisserman and D. Murray, Sequential updating of projective and affine structure from motion, *Int. Journal of Computer Vision* 23, 1997, pp. 235-260.
- [3] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig?, *Proc. of the European Conference on Computer Vision*, 1992, pp 563-578.
- [4] O. Faugeras and B. Mourrain, On the geometry and algebra of the point and line correspondences between n images, *Proc. of the Fifth Int. Conference on Computer Vision Boston, MA*, 1995, pp 951-956.
- [5] A. Fitzgibbon and A. Zisserman, Automatic camera recovery for closed or open image sequences, *Proc. of the European Conference on Computer Vision*, 1998, pp 311-326.
- [6] G. Golub and C. Van Loan, *Matrix Computation*, Johns Hopkins University Press, 1989.
- [7] R. Hartley and A. Zissermann, *Multiple view geometry in computer vision*, Cambridge University Press, Cambridge UK, 2000.
- [8] R. Hartley, In defense of the 8-point algorithm, *Proc. of the Fifth Int. Conference on Computer Vision Boston, MA*, 1995, pp 1064-1070.
- [9] A. Heyden, Projective structure and motion from image sequences using subspace Methods, *Scandinavian Conf on image Analysis*, 1997, pp 963-8.

- [10] A. Heyden, A common framework for multiple views tensors, *Proc. of the European Conference on Computer Vision*, 1998, pp 3-19.
- [11] D. Jacobs, Linear fitting with missing data: Applications to structure from motion and to characterizing intensity images, *Proc of IEEE conf Computer Vision and Pattern Recognition*, 1997, pp 206-212.
- [12] F. Kahl and A. Heyden, Affine structure and motion from points, lines and conics, *International Journal of Computer Vision*, 1999, pp. 163-180.
- [13] S. Laveau, Geometric d'un systeme de N cameras. Theorie, estimation et applications, Ph.D. Thesis, INRIA, 1996.
- [14] N. Levi and M. Werman, The viewing graph, *Proc of IEEE conf Computer Vision and Pattern Recognition*, 2003, pp 599-606.
- [15] S. Mahamud and M. eber, Iterative projective reconstruction from multiple views, *Proc of IEEE conf Computer Vision and Pattern Recognition*, 2000.
- [16] D. Martinec and T. Pajdla, Structure from many perspective images with occlusions, *Proc. of the European Conference on Computer Vision*, 2002, pp 355-369.
- [17] G. Sparr, Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences, *Proc of Int. conf. on Pattern recognition*, 1996, pp 328-333.
- [18] P. Sturm and B. Triggs, A factorization based algorithm for multi-image projective structure and motion, *Proc. of the European Conference on Computer Vision*, 1996.
- [19] C. Tomasi and T. Kanade, Shape and motion from image streams under Orthography: A factorization method, *Int. J. Computer Vision* 9(2), 1992, pp 137-154.
- [20] B. Triggs, Matching constraints and the joint image, *Proc. of the Fifth Int. Conference on Computer Vision*, 1995, pp 338-343.
- [21] B. Triggs, Factorization methods for projective structure and motion, *Proc. of Conf on Computer Vision and Pattern Recognition*, 1996, pp 845-851.
- [22] B. Triggs, Linear projective reconstruction from matching tensor, *Image & Vision Computing* vol 15, 1997, pp. 617-626.
- [23] S. Ullman and R. Basri, Recognition by linear combination of models, *Trans. on Pattern Anal. and Machine Intell.* 13, 1991, pp. 992-1006.
- [24] M. Urban, T. Pajdla and V. Hlavac, Projective reconstruction from N views having one view in common, *Workshop on vision Algorithms: Theory and Practice*, pringer LNCS 1883, 1999, pp 116-131.

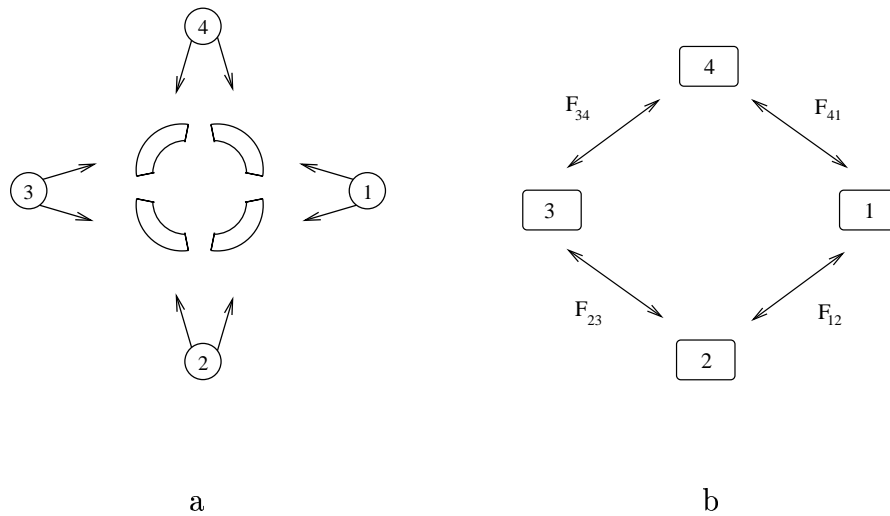


Fig. 1. (a) Four cameras are placed around a 3D object. (b) The pictures taken by the cameras overlap in a pairwise manner. The fundamental matrices, which can be extracted from the measurements, encapsulate all the pairwise geometric information.

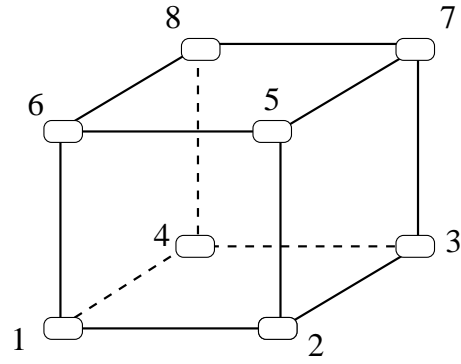


Fig. 2. Eight cameras are located in eight corners of a room. The 3D object is placed in the middle. A straight line between two cameras denotes view overlapping. Matching points can be extracted only from connected pairs of views.

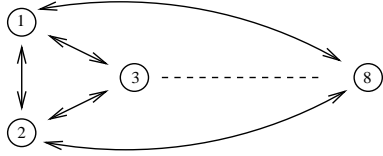


Fig. 3. A parallel configuration of eight views. Each view is related to the first two views

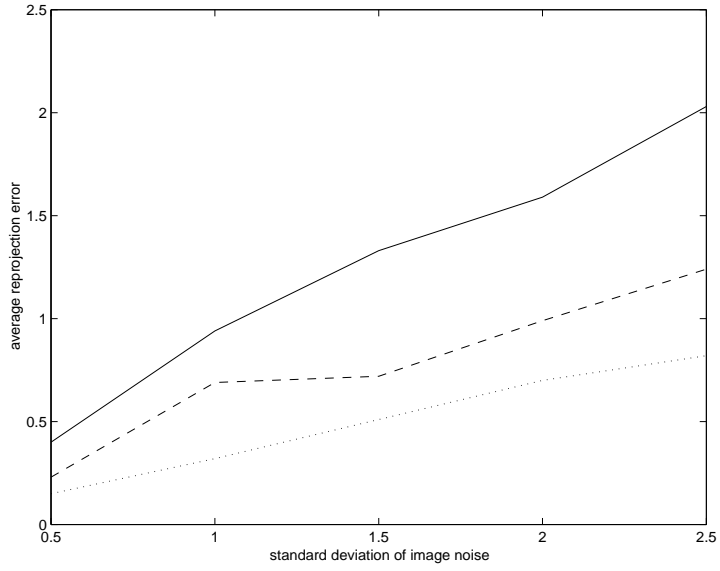


Fig. 4. Average reprojection error vs. image noise. Results of the cube shape are in a solid line, results of the parallel shape are in a dashed line and results where all the fundamental matrices are known is in a dotted line.



a



d



b



c

Fig. 5. Four images of two cars taken from four directions. In the view sequence a,b,c,d, point correspondences can be found only in consecutive views.