Trajectory Reconstruction for Affine Structure-from-Motion by Global and Local Constraints

Hanno Ackermann and Bodo Rosenhahn
Leibniz University Hannover
`firstname.secondname@tnt.uni-hannover.de`

Abstract

The problem of reconstructing a 3D scene from a moving camera can be solved by means of the so-called Factorization method. It directly computes a global solution without the need to merge several partial reconstructions. However, if the trajectories are not complete, i.e. not every feature point could be observed in all the images, this method cannot be used. We use a Factorization-style algorithm for recovering the unobserved feature positions in a non-incremental way. This method uniformly utilizes all data and finds a global solution without any need of sequential or hierarchical merging. Two contributions are made in this work: Firstly, partially known trajectories are completed by minimizing the distance between the subspace and the trajectory within an affine subspace associated with the trajectory. This amounts to imposing a global constraint on the data. Secondly, we propose to further include local constraints derived from epipolar geometry into the estimation. It is shown how to simultaneously optimize both constraints. By using simulated and real image sequences we show the improvements achieved with our algorithm.

1. Introduction

One of the oldest problems in computer vision is to estimate the three-dimensional structure of a rigid object from images taken by cameras moving around it. This is sometimes labeled as the *structure-from-motion* problem. The classic approach is to first establish correspondences between feature points in the images and then to compute the structure of the object and the motion of the camera by these correspondences.

By means of the so-called *Factorization* method, it is possible to compute the 3D reconstruction of the shape in a single step. Its advantage is that it uniformly utilizes all available data and directly computes a global solution, i.e. reconstructions from several sub-sequences need not be merged. However, the use of this method requires that the correspondences of all feature points must be known across all images. If not all points could be observed in every image, this method cannot be used. Many algorithms have been introduced to complete partially observed trajectories.

In the case of an affine camera, Tomasi and Kanade [19] used the fact that the trajectories of a rigid body moving arbitrarily in space are constrained to be in a 3-dimensional affine subspace. Their idea was to obtain an initial estimate of this subspace by starting with a completely known subset of the measurement matrix. A subspace is fitted to this matrix, and incomplete rows and columns of the total matrix are incrementally recovered subsequently. The first problem with this method is that finding the best initial submatrix is known to be NP-hard, as noticed by Jacobs in [10]. Secondly, the initially computed subspace may be inaccurate, and the error is uncontrollably propagated which might result in a wrong solution.

Jacobs [10] extended this idea and fitted subspaces to several submatrices given by triples of columns. The author combines these subspaces to obtain an estimate of the global subspace. This estimate is then used to fill in the missing values. However, this algorithm does not uniformly utilize the data, and it merges partial reconstructions.

Brand [3] generally showed how to compute the SVD of a matrix having missing elements. Starting with a completely known submatrix, the missing row or column elements are incrementally chosen such that the rank constraint is minimized. Since this approach is similar to the one of Tomasi and Kanade, it shares the disadvantages: finding the best initial submatrix to start with is NP-hard and the error made at the initial factorization might cause the algorithm to fail. Chen and Suter [5] extended Brand's method by using only certain, reliable trajectories for the subspace estimation. They proposed how to determine these trajectories and introduced a second stage of their algorithm to find a global optimum. Sugaya and Kanatani [17] select reliable trajectories at each iteration and iteratively estimate the unknown parts of trajectories.

Hartley and Schaffalitzky [8] proposed an algorithm to factorize a matrix having unknown entries without using any estimates of the unknown values. Marquez and Costeira [13] introduced a non-incremental method for estimating

the missing data. They iteratively estimate the unobserved feature points. Additionally, they utilize a constraint on the structure of the affine camera matrices.

In the case of a perspective camera an often used technique is to initialize a statistical optimization technique such as *bundle adjustment* with solutions computed from multi-linear constraints [6, 9, 16, 18, 20]. Alternatively, it is possible to use multi-linear constraints and sequentially [1, 2] or hierarchically [7] built up a conjoint reconstruction of the 3D shape.

Martinec and Pajdla [14] combine local and global constraints: in their iterative algorithm, they alternate between estimating fundamental matrices and then fitting a subspace to obtain a projective estimate of the 3D shape. Epipolar geometry is used for recovering the projective depths, and the subspace constraint to estimate the missing data. In [15] the same authors propose to combine multiple epipolar constraints to estimate the projective depths.

Algorithms using multi-linear or closure constraints use only observed data, but need to choose particular combinations of images, and combine the partial reconstructions by some means. Conversely, factorization algorithms, also called subspace algorithms, utilize all data uniformly and do not require to sequentially or hierarchically merge partial reconstructions. However, they usually use estimates of the unobserved data. The latter might cause them to reconstruct a heavily distorted shape.

We propose to use a subspace algorithm for completing the trajectories. Uniformly utilizing all available data, we are independent of the image ordering. Our algorithm requires neither sequential nor hierarchical merging of partial reconstructions, and it is not incremental.

The contributions in this work can be summarized as follows:

1. We utilize the global subspace constraint, and minimize a distance between the trajectory and the estimated subspace. Since only unknown elements of the trajectory may be changed, an affine subspace is associated with the trajectory, and a simple expression is derived for minimizing the distance to the subspace within the affine space.

2. We integrate local constraints derived from epipolar geometry into the algorithm. It is shown how to combine the subspace constraint and the epipolar constraints to estimate the missing data simultaneously.

To the best knowledge of the authors, neither one of these two ideas has been used in the context of completing partially known trajectories.

It has been stated before that subspace algorithms fail to converge even if few missing data are present since estimates are used during the optimization [3]. Conversely,

utilizing only epipolar geometry may fail in the case of degenerate camera configurations [9]. However, if both types of constraints are combined, these problems can be compensated.

The paper has the following structure. In Sec. 2, we shortly summarize the subspace constraint. Our result on missing data estimation using the subspace constraint is derived in Sec. 3.1. The geometric constraints are presented in Sec. 3.2. We evaluate our algorithm in Sec. 4 using synthetic and real image sequences. The paper concludes with a brief summary in Sec. 5.

2. Affine Subspace Constraint

Suppose N rigidly moving points are tracked through M consecutive frames taken by an affine camera. The projection of the αth point onto the κth image can be modeled by

$$\mathbf{x}_{\kappa\alpha} = \begin{pmatrix} u_{\kappa\alpha} \\ v_{\kappa\alpha} \\ 1 \end{pmatrix} = P_\kappa \mathbf{X}_\alpha + \mathbf{t}_\kappa \qquad (1)$$

where $u_{\kappa\alpha}$ and $v_{\kappa\alpha}$ denote the measured x and y coordinates in the image positions [9, 17]. The 2×3 matrices P_κ define the affine projection in the κth image, and $\mathbf{X}_\alpha$ are the 3-dimensional positions of the αth point. The 2-vectors $\mathbf{t}$ define a translation for each camera.

Let the $2M \times N$ matrix W be

$$W = \begin{pmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{M1} & \cdots & \mathbf{x}_{MN} \end{pmatrix}, \qquad (2)$$

and the $2M \times 3$ matrix P, the $2M$ vector $\mathbf{t}$ and the $3 \times N$ matrix X be

$$P = \begin{pmatrix} P_1 \\ \vdots \\ P_M \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_M \end{pmatrix},$$

$$X = \begin{pmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_N \end{pmatrix}. \qquad (3)$$

Eq. (1) may now be written as

$$W = PX + \mathbf{t}. \qquad (4)$$

Locating the coordinate origin at $\mathbf{t}$, Eq. (4) implies that the columns of W are constrained to be in an 3-dimensional affine subspace spanned by the columns of P.

If the origin of the coordinate system is not fixed to the centroid $\mathbf{t}$, the trajectories are constraint to be in a 4-dimensional subspace. We utilize this fact and directly fit a 4D subspace to the trajectories. Under the assumption of independent and identically distributed zero-mean Gaussian noise, a maximum-likelihood estimate of the subspace can be computed using singular value decomposition of the matrix defined in Eq. (2).

3. Recovery of Partial Trajectories

The Factorization method computes a global and unique solution to the structure-from-motion problem. However, if the feature points could not be observed in all images, this method cannot be used. This happens if there is occlusion or if problems occur during tracking the feature points. In this section, an algorithm is described, which can recover partially observed trajectories.

In the first part we utilize the global subspace constraint, and derive how to minimize the distance between a trajectory and the estimated subspace along an associated affine space. In the second part it is shown how the local epipolar constraints can be exploited. The last part of this Section explains how to combine both types of constraints into a single equation system so that all constraints may be simultaneously optimized.

3.1. Subspace Constraint

Consider the following matrix:

$$W = \begin{pmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdot & \mathbf{x}_{14} & \mathbf{x}_{15} \\ \cdot & \mathbf{x}_{22} & \cdot & \mathbf{x}_{24} & \mathbf{x}_{25} \\ \mathbf{x}_{31} & \mathbf{x}_{32} & \mathbf{x}_{33} & \cdot & \mathbf{x}_{35} \\ \cdot & \cdot & \mathbf{x}_{43} & \mathbf{x}_{44} & \cdot \\ \mathbf{x}_{51} & \mathbf{x}_{52} & \cdot & \mathbf{x}_{54} & \cdot \end{pmatrix}. \quad (5)$$

The $\cdot$ symbol denotes missing elements. Let k be the number of missing elements in one column vector $\mathbf{w}_\alpha$ of W.

The idea is to separate the known from the unknown entries of each particular column vector. The known entries need be kept fixed, but we may choose the missing ones as we like. Therefore, each vector $\mathbf{w}_\alpha$ having unknown entries gives rise to an affine subspace as follows (cf. Fig. 1):

$$\mathbf{w}_\alpha = A\mathbf{y} + \mathbf{v}. \quad (6)$$

The origin of this affine space is given by the $2M$-dimensional vector $\mathbf{v}$. Each two of its entries equal $\mathbf{x}_{\kappa\alpha}$ if the observation is known and zero otherwise. The vector $\mathbf{y}$ corresponds to the positions $\mathbf{x}_{\kappa\alpha}$ which were not observed, and its length is twice the number of unknown vectors $\mathbf{x}_{\kappa\alpha}$ in $\mathbf{w}_\alpha$. The basis A of this affine space consists of $2k$ $2M$-dimensional basis vectors $\mathbf{e}_i$. Each two basis vectors correspond to one unknown observation $\mathbf{x}_{\kappa\alpha} = (u \ v)^\top$ in the vector $\mathbf{w}_\alpha$. The first of the two basis vectors equals 1 at the coordinate corresponding to the coordinate u of $\mathbf{x}_{\kappa\alpha}$ in the vector $\mathbf{w}_\alpha$; the second basis vector equals 1 at the coordinate corresponding to v, respectively. They are zero elsewhere.

The idea is to assign the unknown entries in $\mathbf{w}_\alpha$ values which minimize the distance to the subspace. In other words, we want to minimize the distance between $\mathbf{w}_\alpha$ and its orthogonal projection onto the subspace. Since only val-
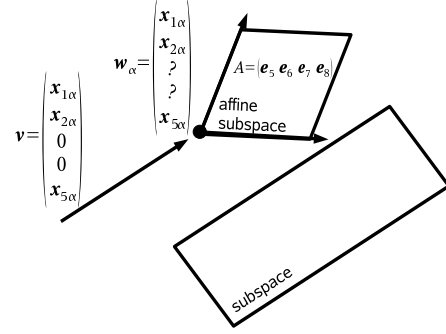

Figure 1. Each vector $\mathbf{w}_\alpha$ having unknown entries is separated into the known and the unknown components. Since the unknown components may be chosen arbitrarily, they give rise to an affine subspace with the origin being the vector consisting of the known entries of $\mathbf{w}_\alpha$ and being zero otherwise. Here, $\mathbf{e}_i$ denotes a 10-dimensional basis vector with its ith coordinate set to 1. The 2-vectors $\mathbf{0}$ are identically zero.

ues for the unknown observations may be chosen, the distance minimization is performed *within* the affine subspace spanned by the missing entries.

Assume that we are given an orthonormal basis U for the subspace which spans W, i.e. the column vectors of U have unit length and are mutually orthogonal. The missing entries of $\mathbf{w}_\alpha$ are now chosen such that the distance of $\mathbf{w}_\alpha$ to the subspace becomes 0, i.e. we want the distance between $\mathbf{w}_\alpha$ and its orthogonal projection $\hat{\mathbf{w}}_\alpha$ onto U to become zero:

$$\begin{aligned} 0 &= \|\mathbf{w}_\alpha - \hat{\mathbf{w}}_\alpha\|^2 \\ \Leftrightarrow 0 &= \|\left(I - UU^\top\right)\mathbf{w}_\alpha\|^2. \end{aligned} \quad (7)$$

Here, I denotes the identity matrix. Since U is an orthonormal matrix the expression for orthogonal projection is simplified.

Inserting Eq. (6) into Eq. (7), we obtain

$$0 = \|\left(I - UU^\top\right)\left(A\mathbf{y} + \mathbf{v}\right)\|^2. \quad (8)$$

Using normal equations, the solution to Eq. (8) is given by

$$\mathbf{y} = -\left(\left(I - UU^\top\right)A\right)^+ \left(I - UU^\top\right)\mathbf{v}, \quad (9)$$

where $(\cdot)^+$ denotes the general inverse.

If the subspace were known, $\mathbf{y}$ could be directly determined. Conversely, if the complete matrix W were known, the subspace could be computed. This suggests the following algorithm: First, the points in W which could not be observed are initialized to 1. Then, an 4D-subspace is fitted to W and the unknown observations are estimated by Eq. (9). The two steps of subspace fitting and missing data estimation are repeated until convergence.

3.2. Including Epipolar Constraints

The iterative subspace estimation algorithm presented in the previous section uses estimated trajectories. Since these

estimates can significantly differ from the true position, an iterative algorithm alternating between subspace fitting and missing data estimation may converge to a wrong solution. The more unknown feature positions, the worse a solution, computed by using only the subspace constraint. In order to prevent this, we include local constraints derived from epipolar geometry.

Utilizing epipolar geometry does not imply that any particular images are selected. In general, it is difficult to analytically determine the images which are best suited to compute the epipolar geometry, so we integrate *all* epipolar constraints we are able to use.

Given two images i and j of the same rigid scene, corresponding points $\mathbf{x}_i = (u \ v \ 1)_i^\top$ and $\mathbf{x}_j = (u \ v \ 1)_j^\top$ between the two images satisfy the affine epipolar equation [9]

$$\mathbf{x}_j^\top F_{ji} \mathbf{x}_i = 0, \tag{10}$$

The affine Fundamental matrix F is a 3×3 matrix of rank 2 consisting of five parameters

$$F = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix}. \tag{11}$$

If there are at least four correspondences between any two images, it is possible to compute the affine Fundamental matrix by a linear method. Due to the parametrization in Eq. (11), the matrix F automatically satisfies the rank-2 constraint [9].

If some point $\mathbf{x}$ could not be observed in image $\mathcal{I}'$ but it was observed in other images $\mathcal{I}$, we can compute the epipolar lines

$$\mathbf{l}_{ji} = F_{ji} \mathbf{x}_i, \quad j \in \mathcal{I}', \ \ i \in \mathcal{I} \tag{12}$$

of this point in image $\mathcal{I}'$ provided that we know the Fundamental matrices between the images.

In the absence of noise and assuming that the cameras are in general positions, the epipolar lines corresponding to other cameras which observed the feature point intersect in a single point $\mathbf{x}$, i.e.

$$\sum_{i \in \mathcal{I}} \mathbf{l}_i^\top \mathbf{x} = 0 \tag{13}$$

(cf. Fig. 2). The solution to Eq. (13) is the best estimation of the unobserved feature point $\mathbf{x}$ in terms of epipolar geometry.

In the presence of noise, the epipolar lines do not intersect in a single point. However, as long as the noise in the observed points of the images is not too large, the intersections will be close to the true position. Utilizing Eq. (13) for estimating the unobserved positions implies that an algebraic distance is minimized, not the correct yet hard to optimize geometrical distance.
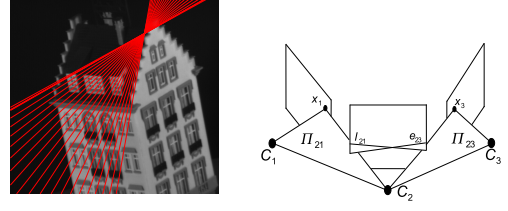

Figure 2. Left Image: The affine epipolar lines (red lines) intersect in a single point. Confer to Sec. 4 for detailed information about the sequence. Right Image: The camera centers are denoted by C_1, C_2 and C_3. The feature point observed in image 1 and 3 is denoted by x_1 and x_3, respectively. Here, l_{21} and l_{23} denote the epipolar lines induced by planar homographies Π_{21} and Π_{23} between images 1 and 2, and 3 and 2, respectively. The intersection is the best estimation of the unobserved feature in image 2 in terms of epipolar geometry.

Since we do not know which epipolar lines intersect closest to the true position of the unobserved point, the unknown position of $\mathbf{x}_j$ in image $\mathcal{I}'$ may be determined by least-squares minimization of all the Eqs. (13). Imagine a sequence of five images. A particular feature point was not observed in images one and two, and the Fundamental matrices could not be computed between images one and five. Denote by $(u \ v \ 1)_i^\top$ the feature in the first or second image, respectively. The joint system of Eqs. (13) then is as follows

$$\begin{pmatrix} \mathbf{l}_{13}^\top & 0 \\ \mathbf{l}_{14}^\top & \\ & \mathbf{l}_{23}^\top \\ 0 & \mathbf{l}_{24}^\top \\ & \mathbf{l}_{25}^\top \end{pmatrix} \begin{pmatrix} \begin{Bmatrix} u \\ v \\ 1 \end{Bmatrix}_1 \\ \begin{Bmatrix} u \\ v \\ 1 \end{Bmatrix}_2 \end{pmatrix} = \mathbf{0}. \tag{14}$$

Here, $\mathbf{0}$ denotes a vector solely consisting of zeros. The upper right and lower left blocks of the matrix on the left side are identically zero.

3.3. Joint Optimization

We will now describe how to jointly optimize both types of constraints. In order to do so, both the Eqs. (8) and (13) need be merged into a single equation system.

However, Eq. (8) is inhomogeneous while Eqs. (13) seem to be homogeneous. This is due to the definition of the vectors $\mathbf{x}_j = (u \ v \ 1)_j^\top$ which contain a 1 as third coordinate in the context of Eqs. (13). Since only the first two coordinates of $\mathbf{x}_j$ may be varied, we subtract all third coefficients of the epipolar lines $\mathbf{l}_{ji}$ from both sides of Eqs. (13)

$$\sum_{j \in \mathcal{I}'} (l_1)_{ji} u_j + (l_2)_{ji} v_j = -(l_3)_{ji}. \tag{15}$$

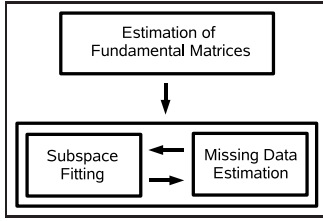Here, $(l_k)_{ji}$ denotes the kth coordinate of the epipolar line $\mathbf{l}_{ji}$.

Figure 3. A schematic of the method introduced in this work. First, all Fundamental matrices are estimated. Subsequently, the algorithm alternates between subspace fitting and missing data estimation.

For instance, consider a sequence of five images. For some trajectory α, let two observations in the first and second images be unknown. The following Fundamental matrices can be computed: F_{13} (between the third and first image), F_{14}, F_{23}, F_{24} and F_{25}. Let $P = \left(I - UU^\top \right)$, $\mathbf{0}_{6\times 4}$ be a 6×4 matrix solely consisting of zeros, and $\mathbf{0}_2$ be a 2-vector being identically 0. Then, the joint equation system arising for this case is as shown in Eq. 16:

$$
\left(
\begin{array}{c}
P \left(
\begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\multicolumn{4}{c}{\mathbf{0}_{6\times 4}}
\end{array}
\right) \\
\begin{array}{cccc}
(l_1)_{13} & (l_2)_{13} & 0 & 0 \\
(l_1)_{14} & (l_2)_{14} & 0 & 0 \\
0 & 0 & (l_1)_{23} & (l_2)_{23} \\
0 & 0 & (l_1)_{24} & (l_2)_{24} \\
0 & 0 & (l_1)_{25} & (l_2)_{25}
\end{array}
\end{array}
\right)
\left(
\begin{array}{c}
\left(\begin{array}{c} u \\ v \end{array}\right)_{1\alpha} \\
\left(\begin{array}{c} u \\ v \end{array}\right)_{2\alpha}
\end{array}
\right)
$$

$$
= \left(
\begin{array}{c}
-P \left(
\begin{array}{c}
\mathbf{0}_2 \\
\mathbf{0}_2 \\
\mathbf{x}_{2\alpha} \\
\mathbf{x}_{4\alpha} \\
\mathbf{x}_{5\alpha}
\end{array}
\right) \\
\left(
\begin{array}{c}
-(l_3)_{13} \\
-(l_3)_{14} \\
-(l_3)_{23} \\
-(l_3)_{24} \\
-(l_3)_{25}
\end{array}
\right)
\end{array}
\right) . \quad (16)
$$

The scale of the two equation systems (8) and (15) is different, so they need be normalized. Since any scaling of the rows of Eq. (8) changes the geometrical meaning, we adjust Eqs. (15). We determine the scale factor for Eqs. (15) so that the vector on the right side has the same length as the vector on the right side of Eq. (8).

Whenever it is possible to compute a Fundamental matrix we include the corresponding epipolar constraints into the estimation. All epipolar lines are indiscriminately utilized without considering possible degenerate camera positions. For some images, however, there are no correspondences between images in which a particular point was observed and the images where the point could not be observed. In this case, we complete the trajectories by only minimizing the distance to the subspace, i.e. only the subspace constraint is utilized then. A scheme of the final algorithm is shown in Fig. 3.

4. Experimental Results

In this Section we evaluate the algorithm introduced in this paper. Firstly, the performance is measured using a synthetic image sequence. The accuracy is calculated for different ratios of missing data and different levels of noise. Secondly, the method is tested on real image sequences.

4.1. Synthetic Sequence

We created a synthetic image sequence of 200 points on a cylindrical surface shown in Fig. 4(a). The points were projected into images of size 512×512 pixels by a perspective camera with focal length 600 pixels. The sequence consists of 20 images.

Using this data, our algorithm is compared with Power-Factorization [8]. This algorithm estimates the left and right subspace of the input matrix by only considering the known entries. We used the implementation kindly provided by Buchanon [4][1].

Incomplete trajectories were simulated by randomly removing a certain percentage of points. This simulates trajectories as might be created using a SIFT tracker [12], where feature points are missing in one frame, but may reappear in subsequent images. The accuracy of the algorithm was measured by computing the Frobenius norm between the matrix consisting of recovered trajectories and the matrix consisting of the ground-truth trajectories. Furthermore the 3D error was computed by computing the homography which optimally projects the reconstructed shape to the ground-truth shape. The average Euclidean distance between the 3D points was taken as measure. Both algorithms were iterated until convergence. PowerFactorization was randomly started 100 times and the best result was used.

To determine how robust the proposed algorithm is to missing points, we sampled the sequence in Fig. 4(a) ten times for sampling ratios of $r = 10\%, 20\%, \ldots, 70\%$. Results are shown in the two leftmost plots of Fig. 4(b). The left of the two plots correspond to the Frobenius norm between the reconstructed matrix and the ground-truth one, the right plot the 3D error. The solid line indicates the mean error of the ten trials of our algorithm, the dashed line the maximum error; the dotted line the mean error of Power-Factorization, the dash-dotted line the maximum error. Both algorithms perform almost identically for low sampling ratios. For a large amount of unobserved points, the proposed algorithm converges more reliably to a good solution. For 70%, our method failed only once to compute a very accurate solution while PowerFactorization computed a satisfactory solution only twice and even failed to converge in two more trials.

To determine how susceptible the proposed algorithm is to noise, we fixed the number of missing data to 60% and
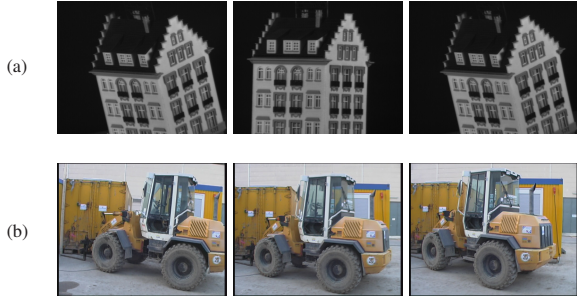
---

[1]www.robots.ox.ac.uk/ amb

Figure 5. (a) The three images show the first, an intermediate, and the last frame of the 182-image hotel sequence. It consists of 816 trajectories. (b) The first, an intermediate, and the last image of a sequence of an industrial dredger are shown. The sequence consists of 50 frames and 464 points.

increasingly added normally distributed noise with standard deviation $\sigma = 0, 0.5, \ldots, 3.0$. The results are shown in the two rightmost plots of Fig. 4(b). Our method recovers the trajectories same as good as PowerFactorization while the accuracy of the reconstructed 3D shape is somewhat better. Although we tried PowerFactorization 1500 times, it failed to compute an accurate solution in one trial for $\sigma = 0$.

For both experiments, the proposed algorithm was five to fifty times faster than PowerFactorization.

4.2. Real Image Sequences

The performance of the algorithm was evaluated using two different real image sequences.

First, we used the sequence of the model of a hotel[2]. It consists of 182 images. Three of them are shown in Fig. 5(a). The publicly available *voodoo camera tracker*[3] was utilized for establishing a total of 816 correspondences. Due to brightness changes, the tracker looses almost all points around frame 55, so the amount of missing data is approximately 55%. The top diagram in Fig. 6 shows which feature was found in which image (black indicates that the feature was observed in the respective frame). By visual inspection we conclude that outliers are not present in this data set.

After completing the trajectories using the approach from Sec. 3, we reconstructed the affine 3D shape and motion with the algorithm from [11]. Four images of the reconstructed shape are shown in Fig. 7. The overall quality of the shape is satisfactory. However, some angles are not exactly rectangular as they ought be. This is a well-known problem due to the use of the affine camera model whereas the images were created using a perspective camera [6, 9].

The second sequence shows an industrial dredger. In this sequence, the camera moves freely around the object. A total of 464 points were tracked over 50 frames. The

first, middle and last images of the sequence are shown in Fig. 5(b). The correspondences were also established using the voodoo camera tracker. Although the overall quality of the correspondences is good, some outliers are present in the tracking data. However, we did not employ *any* outlier rejection, neither for estimating the subspace nor for computing the Fundamental matrices. In this sequence approximately 20% of the feature points were not be observed due to occlusion and lost trajectories. For the recovery of the trajectories, we used all feature points tracked over at least twenty frames. The observation matrix which shows in which frames each feature could be observed is shown in the bottom Fig. 6.

Although the amount of missing data seems to be low, this sequence is especially challenging: the depth variation within the sequence amounts to the distance between camera and dredger, at least, and thus perspective distortion is strong. For many images the depth variation within the scene is multiple times larger than the distance between camera and scene. This violates the affine camera model assumption which requires that the depth variation in the scene is negligible compared with the distance to the camera.

After having recovered the trajectories, we used an affine 3D-reconstruction algorithm [11]. Two images of the reconstructed 3D shape are shown Fig. 8(a). We assigned each 3D point the average color over all images it was observed in. As can be seen the reconstruction looks reasonable, and the distortion due to the affine camera model is low. For the purpose of better visualization, the reconstructed shape was manually augmented by adding texture patches obtained from the images. The resulting images are shown in Fig. 8(b).

5. Conclusions

In this paper, we introduced a method to recover partial trajectories for affine 3D-reconstruction. It uniformly utilizes all available data and non-incrementally computes a global solution.

Local and global constraints on rigid scenes are jointly optimized. The cost function utilized to enforce the subspace constraint has not been used in the context of recovering partial trajectories. An expression was derived how to minimize the distance to the subspace in the direction of the unknown data. The use of the global subspace constraint was complemented by including as many local epipolar constraints as possible. We showed how to simultaneously optimize both constraints.

The effectiveness of the proposed method was demonstrated by simulated and real image sequences without using any outlier rejection strategy. Epipolar lines arising from cameras being in degenerate positions were not excluded from the estimation. For the simulated sequence, it
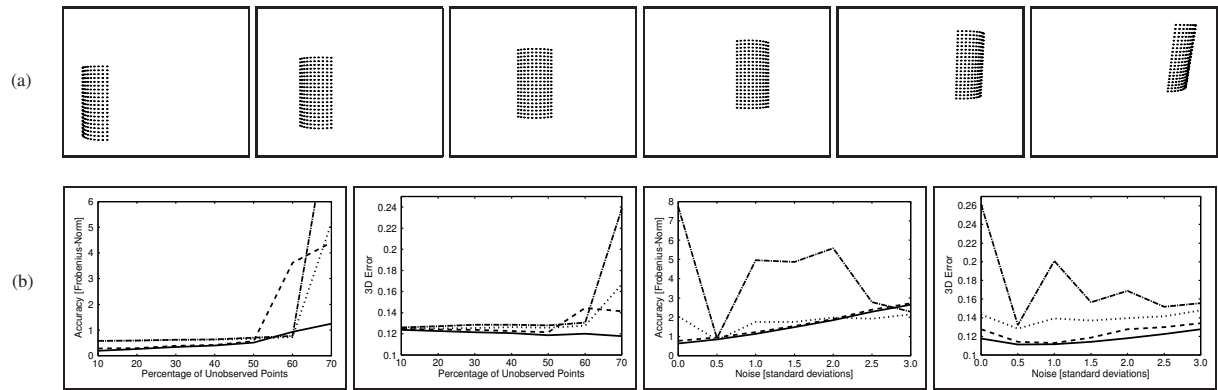
Figure 4. (a) Simulated image sequence of 200 points on a cylindrical surface through 20 frames (six images are shown here). (b) Left left two plots correspond to a noise-free experiment in which the amount of unobserved points was gradually increased. The right two plots correspond to an experiment in which the noise was gradually increased (the amount of missing data was fixed to 60%). The solid line indicates the mean error of ten trials of the proposed algorithm, the dashed line the maximum error. The dotted line indicates the mean error achieved using PowerFactorization, the dash-dotted line indicates the maximum error.

Figure 7. Reconstructed shape corresponding to the 182 image sequence shown in Fig. 5(a). The sequence consists of 816 trajectories and the the missing data amounts to approximately 55%. As can be seen, the shape looks satisfactory. Due to the use of the affine camera model some angles are distorted which is shown in the fourth image (top-view).

was shown that the trajectories were accurately recovered. For the two real image sequences, affine 3D shapes computed from the recovered trajectories were satisfactory. The solution of this algorithm can be used to initialize a statistical optimization technique such as bundle adjustment.

References

[1] S. Avidan and A. Shashua. Threading fundamental matrices. In *5th European Conference on Computer Vision (ECCV)*, pages 124–140, Freiburg, Germany, June 1998.

[2] P. A. Beardsley, A. Zisserman, and D. W. Murray. Sequential updating of projective and affine structure from motion. *International Journal of Computer Vision*, 23(3):235–259, 1997.

[3] M. Brand. Incremental singular value decomposition of uncertain data with missing values. In *7th European Conference on Computer Vision (ECCV)*, pages 707–720, Copenhagen, Denmark, June 2002.

[4] A. Buchanan and A. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 316–322, San Diego, CA, USA, 2005.

[5] P. Chen and D. Suter. Recovering the missing components in a large noisy low-rank matrix: Application to sfm. *IEEE Transactions on Pattern Analyis and Machine Intelligence*, 26(8):1051–1063, August 2004.

[6] O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. The MIT Press, Cambridge MA, USA, 2001.

[7] A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *5th European Conference on Computer Vision (ECCV)*, pages 124–140, Freiburg, Germany, June 1998.

[8] R. Hartley and F. Schaffalizky. Powerfactorization: an approach to affine reconstruction with missing and uncertain data. In *Australia-Japan Advanced Workshop on Computer Vision*, pages 707–720, Copenhagen, Denmark, June 2002.

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

[10] D. W. Jacobs. Linear fitting with missing data for structure-from-motion. *Computer Vision and Image Understanding*, 82(1):57–81, April 2001.

[11] K. Kanatani. Latest progress of 3d-reconstruction from multiple camera images. *Robotics Research Trends*, pages 33–75, 2008.

[12] D. G. Lowe. Object recognition from local scale-invariant features. In *7th International Conference on Computer Vision (ICCV)*, pages 1150–1157, Corfu, Greece, 1999.

[13] M. Marquez and J. Costeira. Optimal shape from motion estimation with missing and degenerate data. In *IEEE Workshop on Application of Computer Vision (WACV)*, Copper Mountain, CO, USA, January 2008.

Figure 8. Reconstructed shape corresponding to the sequence shown in Fig. 5(b). 464 trajectories were tracked over 50 images. The amount of missing data was approximately 20%. Although the depth variation within the scene is larger than the distance between camera and dredger, the reconstructed shape looks reasonable. For the upper two images we computed the average colors of the features over all images which the points could be observed in. The right image in Fig. 8(a) shows the scene viewed from above. We manually augmented the shapes shown in Fig. 8(b) by texture patches obtained from the original images.
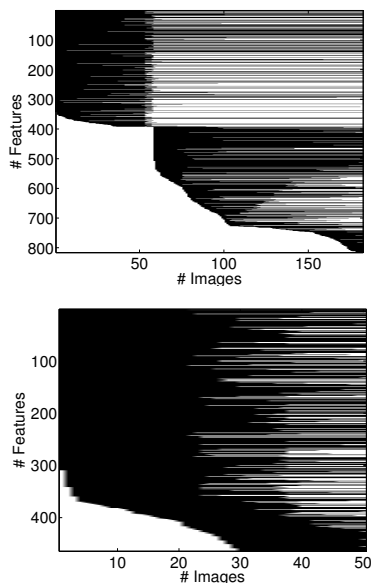

Figure 6. Visibility of the tracked feature points, i.e. which point (y-axis) was observed in which image (x-axis). The black color indicates that the feature could be observed, white that it was not observed. The upper figure corresponds to the hotel sequence shown in Fig. 5(a) (missing data ratio of approximately 55%), the lower to the sequence shown in Fig. 5(b) (missing data ratio of approximately 20%).

[14] D. Martinec and T. Pajdla. Structure from many perspective images with occlusions. In *7th European Conference on Computer Vision (ECCV)*, pages 542–544, Copenhagen, Denmark, June 2002.

[15] D. Martinec and T. Pajdla. 3d reconstruction by fitting low-rank matrices with missing data. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 198–205, San Diego, CA, USA, 2005.

[16] M. Pollefeys, M. Vergauwen, F. Verbiest, K. Cornelis, and L. V. Gool. From image sequence to 3d models. In E. Baltsavias, A. Gruen, and L. V. Gool, editors, *Automatic Extraction of Man-Made Objects From Aerial and Space Images*, pages 403–410, 2001.

[17] Y. Sugaya and K. Kanatani. Extending interrupted feature point tracking for 3-d affine reconstruction. In *8th European Conference on Computer Vision (ECCV)*, pages 310–321, Prague, Czech Republic, May 2004.

[18] T. Thormählen, H. Broszio, and A. Weissenfeld. Keyframe selection for camera motion and structure estimation from multiple views. In *8th European Conference on Computer Vision (ECCV)*, pages 523–535, Prague, Czech Republic, May 2004.

[19] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[20] B. Triggs, P. McLauchlan, R. I. Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.