Projective Reconstruction from Incomplete Trajectories by Global and Local Constraints

H. Ackermann[1], B. Rosenhahn[1]

[1] Leibniz University Hanover, Germany, {*last name*}@*tnt.uni-hannover.de*

Abstract

The paper deals with projective shape and motion reconstruction by subspace iterations. A prerequisite of factorization-style algorithms is that all feature points need be observed in all images, a condition which is hardly realistic in real videos. We therefore address the problem of estimating structure and motion considering missing features. The proposed algorithm does not require initialization and uniformly handles all available data. The computed solution is global in the sense that it does not merge partial solutions incrementally or hierarchically. The global cost due to the factorization is further amended by local constraints to regularize and stabilize the estimations. It is shown how both costs can be jointly minimized in the presence of unobserved points. By synthetic and real image sequences with up to 60% *missing data we demonstrate that our algorithm is accurate and reliable.*

1 Introduction

We consider the estimation of 3D structure and motion of a rigid object from multiple images given only point correspondences if both intrinsic and extrinsic parameters of the cameras are unknown. This is referred to as the *projective structure-from-motion* problem (SfM). Previous works on this subject are discussed in 1.1.

Assuming an *affine* camera model, the so-called *factorization* algorithm can solve the SfM-problem by a single singular value decomposition. Mathematically equivalent is the problem of fitting a linear subspace. In the following, the terms subspace fitting and matrix factorization will both be used to refer to affine SfM.

As the singular value decomposition cannot be computed if some entries of the matrix are missing, algorithms to solve this challenging task are briefly reviewed in 1.2. Please notice that *all* algorithms for incomplete matrix factorization imply an affine camera model. Assuming a projective camera model significantly increases difficulty. The differences between both tasks will be discussed in 1.2.

1.1 3D-Reconstruction

The epipolar constraint can be used for this purpose. However, it only exists for two, three or four images. If there are more than four images, the sequence can be split into pieces to compute *partial reconstructions* for each piece. These partial reconstructions then need be merged sequentially.

The most widely used method to compute structure and motion for sequences of arbitrary length is *bundle adjustment* [28, 30, 13]. Although bundle adjustment algorithms can give accurate solutions, they are susceptible to local minima, and need be initialized with a good solution. A popular implementation of bundle adjustment is the *Bundler*-tool [23] which jointly solves tracking, outlier rejection, and Euclidean 3D-reconstruction. The latter is solved sequentially by interleaving several heuristics. To increase convergence probability, intrinsic parameters are retrieved from exif data of photos. While this algorithm has been shown to reconstruct even large-scale scenes, the complexity of the tool prohibits modifications, e.g. to integrate further constraints, if the tool fails. Furthermore, video sequences usually do not provide information on intrinsic parameters.

An algorithm which estimates a globally optimal solution was introduced by Tomasi and Kanade [29] for affine SFM. It uses the constraint that trajectory vectors reside in a 4-dimensional subspace. This method uniformly handles an arbitrary number of images and points.

This *subspace constraint* also holds true for a *projective camera* if the points are rescaled by scalar variables, called *projective depths*. Christy and Horaud [6] iteratively estimated the subspace by explicitly considering the error of an affine camera model. Sturm and Triggs [24] used closure constraints to estimate the projective depths. Tomita and Ueshiba [31] proposed to nonlinearly optimize the projective depths such that the subspace constraint is satisfied. Heyden et al. [10] and Mahamud and Hebert [15] optimized for the rank constraint by alternating subspace fitting and projective depth estimation. Separating both steps allows to linearly solve each sub-problem. It will be later shown that this is a variant of gradient descent.

1.2 Factorization From Incomplete Data

Subspace approaches are based on the assumption that all points can be observed in every image. This condition is often incorrect for complex scenes in which parts of the objects are occluded by other objects in some images. It can also be violated if parts of the scene are behind a moving camera in some images, or because of failure of the tracking software. The necessity that all points are visible in all views seriously

limits factorization approaches.

The earliest approaches for factorizing a partially known matrix, i.e. *affine* SfM, stem from the mathematical community. In the late 60s, a simple and fast algorithm was introduced which alternates least squares estimation of left and right subspaces [33]. It was later re-introduced to the computer vision community in [3, 9]. In the 70s, Newton-, and Gauss-Newton algorithms were developed [32, 21]. Shum et al. [22] applied the algorithm in [32] to affine SfM. Buchanon and Fitzgibbon [4] switched between alternating least squares approaches and Newton variants. Recently, algorithms from convex optimization were adopted to incomplete matrix factorization [7, 5]. Unfortunately, such algorithms currently do not guarantee that the solution satisfies the required rank constraint, hence are not applicable for SfM.

The earliest approach to incomplete matrix factorization developed by the computer vision community was done by Tomasi and Kanade [29]. They proposed to obtain an initial estimate of the subspace from a completely observed *subset* of the trajectories, and use this solution to estimate unknown parts of trajectories. The main disadvantage of this approach is its dependency on the initial solution. Typically, complete subsets correspond to some degenerate camera motion which significantly simplifies tracking. According to our experience, this algorithm works if the data is noise-free. Jacobs [11] proposed to combine several subspace estimates, each of them fitted to trajectories of sufficient number and length. This idea is not supported by the geometry induced by the problem. Sugaya and Kanatani [25] proposed an EM-style algorithm. They choose only certain, reliable trajectories by a statistical criterion in each iteration. Marques and Costeira [17] regularized subspace estimation by exploiting orthogonality constraints on individual projection matrices. Ackermann and Rosenhahn [2] used epipolar geometry for regularizing missing data estimation.

For the *projective* subspace method, only few solutions exist which consider missing points. Methods for an affine camera cannot be used directly as the projective depths are unknown. Martinec and Pajdla [18] generalized [11] to the projective case. They integrated their idea into the algorithm of Sturm and Triggs [24]. In [19], the same authors propose a solution how to conjoin piece-wise reconstructions. The solution is used to initialize a bundle adjustment. Tang and Hung [27] generalized the method of Berthilsson and Heyden [10] to incomplete trajectories. Their estimation of the projective depths requires nonlinear optimization. The algorithm is evaluated with sequences consisting of not more than ten images.

1.3 Contributions

Incremental algorithms based on the approach in [29] are not very stable and easily converge to incorrect solutions. Algorithms optimizing a global error term – alternating least squares, Newton- and EM-style methods – are more robust. Still, for a projective camera model, the only proposed method [27] does not manage many missing data. Using a five image sequence, this algorithm managed 40% unobserved 2D-points. Generally, EM-style algorithms are unstable if the amount of unobserved data is large as the number of unknowns grows with the unobserved 2D-points: up to our experience they can cope with up to 30% missing points for an affine camera, and up to 20% for a projective camera.

For an affine camera we demonstrated in [2] that the stability of an EM-approach can be greatly improved by incorporating further constraints derived from epipolar geometry. Since they apply regularization to individual unobserved 2D-features, they can be thought of as local constraints on the global factorization error term. On the other hand, epipolar geometry does not restrict the type of motion, i.e. priors such as smoothness of motion or a particular type of motion are not applied.

The contributions made in this work can be summarized as follows: our approach for affine structure-from-motion [2] is (1) generalized to the projective camera model. As opposed to the traditional geometrical derivation [10], we favor an algebraic derivation which (2) reveals that it is a variation of gradient descent. Using epipolar geometry usually implies that particular images are selected. It will be shown that (3) all epipolar constraints can be used simultaneously, and jointly optimized with the global subspace constraint. It is demonstrated that (4) the proposed algorithm converges with up to 60% points being unobserved, and it is robust to noise and varying initialization. Using epipolar geometry implies that no constraints are imposed on the motion of the camera except rigidity. If further prior-knowledge is available, it is (6) not difficult to integrate it into the presented scheme. We will further (7) make the source code available to the scientific community.

Although not explicitly considered in this work, it is possible to embed the proposed method into a RANSAC-scheme. Since such "higher level algorithms" depend on the quality of underlying SfM-methods, we think it worthwhile to improve projective SfM. Furthermore, the solutions computed by this algorithm can be used to initialize a bundle adjustment.

The paper has the following structure. In Sec. 2, we summarize projective reconstruction by subspace iterations. Our solution to 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 discussion in Sec. 5.

2 Projective Reconstruction

In this section, the projective reconstruction algorithms introduced in [10, 15] are shortly summarized.

Suppose N rigidly moving points are tracked through M images observed by a perspective camera with unknown and possibly varying intrinsic parameters. The projection of the

αth point onto the κth image can be modeled by

$$\lambda_{\kappa\alpha}\boldsymbol{x}_{\kappa\alpha} = \lambda_{\kappa\alpha}\begin{pmatrix} u_{\kappa\alpha} \\ v_{\kappa\alpha} \\ 1 \end{pmatrix} = P_\kappa \boldsymbol{X}_\alpha \tag{1}$$

where $u_{\kappa\alpha}$ and $v_{\kappa\alpha}$ denote the measured x and y coordinates in the image positions. For numerical stability [8], we scale the vectors $\boldsymbol{x}_{\kappa\alpha}$ with $s_0 = 1/600$. The 3×4 camera matrices P_κ consist of the intrinsic and extrinsic camera parameters and define the projection into the κth image. The 4-vectors $\boldsymbol{X}_\alpha$ denote the homogeneous position of the αth 3D point. The *projective depths* $\lambda_{\kappa\alpha}$ are scalars due to the perspective projection.

For the projections of all points into every image Eq. (1) may be written as Let the $3M \times N$ matrix W be

$$\begin{aligned} W^{3M\times N} &= \begin{bmatrix} \lambda_{11}\boldsymbol{x}_{11} & \cdots & \lambda_{1N}\boldsymbol{x}_{1N} \\ \vdots & \ddots & \vdots \\ \lambda_{M1}\boldsymbol{x}_{M1} & \cdots & \lambda_{MN}\boldsymbol{x}_{MN} \end{bmatrix} \\ &= \begin{bmatrix} P_1 \\ \vdots \\ P_M \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_N \end{bmatrix} = PX. \end{aligned} \tag{2}$$

Equation (2) implies that the columns of W are constrained to be in a 4-dimensional subspace spanned by the columns of P if the projective depths $\lambda_{\kappa\alpha}$ are correctly chosen.

Since the absolute scale of the projective depths is indeterminate we impose the constraint that $\|\boldsymbol{\lambda}_\alpha\| = 1$ where $\boldsymbol{\lambda}_\alpha = (\lambda_{1\alpha} \quad \cdots \quad \lambda_{M\alpha})^\top$. If the observed features $\boldsymbol{x}_{\kappa\alpha}$ are normalized to unit length it can be seen that the column vectors of W have length 1. Let D_α be the $3M \times M$ matrix consisting of the unit length length features $\boldsymbol{x}_{i\alpha}$, $i = 1, \ldots, M$, on its diagonal. Then we may write $\boldsymbol{w}_\alpha = D_\alpha \boldsymbol{\lambda}_\alpha$.

Utilizing $X = P^+ W$ where $(\cdot)^+$ denotes the generalized inverse it is possible to substitute X in Eq. (2)

$$0 = \left(I - PP^+\right) W. \tag{3}$$

Since $(I - PP^+)$ is the orthogonal projection onto the orthogonal complement of the subspace defined by P, Eq. (3) implies that the column vectors of W must be within P.

If P is orthonormal, and defining $S_\perp = \left(I - PP^\top\right)$, a quadratic cost function is given by

$$f\left(S_\perp, \lambda_{\kappa\alpha}\right) = \frac{1}{2}\|S_\perp W\|^2. \tag{4}$$

For determining the variables $S_\perp$ and all $\lambda_{\kappa\alpha}$, we need to minimize the function defined by Eq. (4) with respect to the two sets of variables by taking the partial derivatives

$$\frac{\partial}{\partial \boldsymbol{\lambda}_\alpha} f = D_\alpha^\top S_\perp D_\alpha \boldsymbol{\lambda}_\alpha \quad \text{and} \tag{5a}$$

$$\frac{\partial}{\partial S_\perp} f = W W^\top S_\perp. \tag{5b}$$

Here we used that $S_\perp^\top S_\perp = S_\perp$.

A necessary condition for minimizing Eq. (4) is that the partial derivatives defined in Eqs. (5) equal zero. Fixing one set of variables, it is possible to solve for the other by means of a linear homogeneous linear equation system. Iterating both steps gives the algorithms introduced in [10, 16].

To avoid computing the $(3M-4)$-dimensional nullspace of W at each iteration, it is more efficient to *maximize* the dual cost

$$f\left(P, \lambda_{\kappa\alpha}\right) = \frac{1}{2}\|PP^\top W\|^2. \tag{6}$$

Its derivatives are given by

$$\frac{\partial}{\partial \boldsymbol{\lambda}_\alpha} f = D_\alpha^\top P P^\top D_\alpha \boldsymbol{\lambda}_\alpha \quad \text{and} \tag{7a}$$

$$\frac{\partial}{\partial P} f = W W^\top P. \tag{7b}$$

Here, we used that $P^\top P = I$ where I denotes the identity matrix. The solution to Eq. (7a) is given by the right singular vector corresponding to the smallest singular value of $P^\top D_\alpha$, while the solution to Eq. (7b) is given by the left singular vectors corresponding to the four largest singular values of W.

3 Recovery of Partial Trajectories

The iterative subspace method presented in Sec. 2 computes a single solution to the projective structure-from-motion problem for any number of points and images. However, if some points cannot be observed in every image, this method cannot be used. That can happen if parts of the scene are occluded in some images, the software for establishing the correspondences fails, or if the scene parts of the scene appear or disappear. In this section, an algorithm is described which generalizes the projective subspace algorithm to the case of partially observed trajectories.

In the first part we utilize the global subspace constraint, and derive a formula for estimating missing points based on this global constraint. The idea is that the missing points are chosen such that each trajectory becomes as close to the subspace as possible. We will show that this distance minimization problem results in a simple formula which can be easily solved. In the second part it is shown how additional constraints can be integrated into this algorithm. Since it is readily available and can be easily computed, we chose epipolar geometry to locally regularize the global subspace constraint. The last part of this Section explains how both types of constraints can be combined into a single system of equations so that all constraints are optimized simultaneously.

3.1 Subspace Constraint

If some entries in $\boldsymbol{w}_\alpha$ are unknown, the cost function defined by Eq. (4) needs to include the unknown entries $\tilde{\boldsymbol{x}}_{\kappa\alpha}$ in W as variables

$$f\left(S_\perp, \tilde{\boldsymbol{x}}_{\kappa\alpha}, \lambda_{\kappa\alpha}\right) = \frac{1}{2}\|S_\perp W\|^2, \tag{8}$$

For the maximization of Eq. (8), these unknown entries have to be considered, i.e. a third partial derivative $\frac{\partial}{\partial \tilde{\boldsymbol{x}}_\alpha} f$ need be defined.

The idea is to choose the missing entries of one particular $\boldsymbol{w}_\alpha$ so that the distance between $\boldsymbol{w}_\alpha$ and the subspace is minimized. The unknown entries may be freely chosen, yet the known entries must stay fixed. This implies that each point $\boldsymbol{w}_\alpha$ with unknown entries induces an affine space. Its origin is given by the vector consisting of the known entries of $\boldsymbol{w}_\alpha$ and being zero elsewhere. The affine space is then spanned along the dimensions of the unknown entries of $\boldsymbol{w}_\alpha$. Estimating missing points of one trajectory then amounts to simply minimizing the distance between this affine space and the subspace.

The origin of this affine space is given by a $3M$-dimensional vector $\boldsymbol{v}_\alpha$. Its entries are the vectors $\boldsymbol{x}_{\kappa\alpha}$ in $\boldsymbol{w}_\alpha$ if they are known. If a point $\boldsymbol{x}_{\kappa\alpha}$ is unknown, the corresponding entries in $\boldsymbol{v}_\alpha$ is zero. The basis of the affine space is defined by a $3M \times 3k$ matrix A, where k denotes the number of unobserved entries in $\boldsymbol{w}_\alpha$. It consists of triples of basis vectors $\boldsymbol{e}_i$ which corresponds to one unknown observation $\boldsymbol{x}_{\kappa\alpha} = \begin{bmatrix} u & v & w \end{bmatrix}^\top$ in the vector $\boldsymbol{w}_\alpha$. The first vector of each triple of basis vectors equals 1 at the coordinate corresponding to the coordinate u of $\boldsymbol{x}_{\kappa\alpha}$ and is zero elsewhere, the second basis vector equals 1 at the coordinate corresponding to v, and the third basis vector equals 1 at the position of the third coordinate w.

Let Λ_α be the $3M \times 3M$ matrix with the vectors $\begin{bmatrix} \lambda_{i\alpha} & \lambda_{i\alpha} & \lambda_{i\alpha} \end{bmatrix}^\top$, $i = 1, \ldots, M$ on its diagonal, and $\boldsymbol{d}_\alpha = \begin{bmatrix} \boldsymbol{x}_{1\alpha}^\top & \cdots & \boldsymbol{x}_{M\alpha}^\top \end{bmatrix}^\top$. Then we may write

$$\boldsymbol{w}_\alpha = \Lambda_\alpha \boldsymbol{d}_\alpha = \Lambda_\alpha \underbrace{\left(A_\alpha \boldsymbol{y}_\alpha + \boldsymbol{v}_\alpha\right)}_{\text{affine space}}. \tag{9}$$

The vector $\boldsymbol{y}_\alpha$ consists of all the unknown points $\boldsymbol{x}_{\kappa\alpha}$. For instance, consider a vector $\boldsymbol{w}_\alpha$ with its first entry known but the second unknown. Then Eq. (9) is as follows

$$\boldsymbol{w}_\alpha = \Lambda_\alpha \left(\begin{bmatrix} 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \\ & \vdots & & \end{bmatrix} \boldsymbol{y}_\alpha + \begin{bmatrix} \boldsymbol{x}_{1\alpha} \\ \boldsymbol{0}_3 \\ \vdots \end{bmatrix} \right) \tag{10}$$

where $\boldsymbol{0}_3$ denotes a 3-vector with all its entries being 0.

The missing entries of $\boldsymbol{d}_\alpha$ are now chosen such that the distance of $\boldsymbol{w}_\alpha$ to the subspace becomes 0, i.e. we want the distance between $\boldsymbol{w}_\alpha$ and its orthogonal projection $\hat{\boldsymbol{w}}_\alpha = PP^+\boldsymbol{w}_\alpha = PP^\top \boldsymbol{w}_\alpha$ onto P to become zero. Inserting Eq. (9) into Eq. (3), we obtain

$$0 = \frac{1}{2} \| S_\perp \Lambda_\alpha \left(A_\alpha \boldsymbol{y}_\alpha + \boldsymbol{v}_\alpha\right) \|^2. \tag{11}$$

Differentiating it by the unknown observations $\boldsymbol{y}_\alpha$ yields the partial derivative $\frac{\partial}{\partial \boldsymbol{y}_\alpha} f$. Equation (8) can now be solved by alternatingly estimating the three partial derivatives

$$\frac{\partial}{\partial S_\perp} f = W W^\top S_\perp, \tag{12a}$$

$$\frac{\partial}{\partial \boldsymbol{y}_\alpha} f = A_\alpha^\top \Lambda_\alpha^\top S_\perp \Lambda_\alpha \left(A \boldsymbol{y}_\alpha + \boldsymbol{v}_\alpha\right), \quad \text{and} \tag{12b}$$

$$\frac{\partial}{\partial \boldsymbol{\lambda}_\alpha} f = D_\alpha^\top S_\perp D_\alpha \boldsymbol{\lambda}_\alpha \quad \text{for } \alpha = 1, \ldots, M. \tag{12c}$$

for zero.

Although the unit length constraints on the vectors $\boldsymbol{x}_{\kappa\alpha}$ seem to render Eq. (12b) into a complicated nonlinear and non-convex optimization problem, they can be ignored: any solution to $\boldsymbol{x}_{\kappa\alpha}$ with $\|\boldsymbol{x}_{\kappa\alpha}\|^2 \neq 1$ is projectively invariant to the solution with length 1. Therefore, the solution is given by

$$\boldsymbol{y}_\alpha = -\left(S_\perp \Lambda_\alpha A_\alpha\right)^+ S_\perp \Lambda_\alpha \boldsymbol{v}_\alpha \tag{13}$$

if the computed vectors $\boldsymbol{x}_{\kappa\alpha}$ are subsequently normalized to length 1.

Note that instead of minimizing Eqs. (12a) and (12c), we maximize Eqs. (7). Minimizing only Eq. (12b) allows to easily include further constraints as will be seen in the following.

3.2 Including Epipolar Constraints

The iterative subspace estimation algorithm presented in the previous section iteratively determines unknown points by using estimates from the previous iteration. This causes this approach to become unstable, if many points could not be observed. To stabilize and regularize the algorithm introduced in the previous section, further constraints are necessary. While it is convenient to use knowledge on the motion, such information is not generally available. On the other hand, generally available constraints are provided by epipolar geometry. They can be easily computed and assert a locally stabilizing effect.

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 for epipolar geometry, so we integrate all epipolar constraints we are able to use. In the following, we derive how they can be integrated into the proposed subspace algorithm.

Given two images i and j of the same rigid scene, corresponding points $\boldsymbol{x}_i = \begin{bmatrix} u & v & w \end{bmatrix}_i^\top$ and $\boldsymbol{x}_j = \begin{bmatrix} u & v & w \end{bmatrix}_j^\top$ between the two images satisfy the epipolar equation $\boldsymbol{x}_j^\top F_{ji} \boldsymbol{x}_i = 0$, The fundamental matrix F is a 3×3 matrix of rank 2 relating two points $\boldsymbol{x}_j$ and $\boldsymbol{x}_i$ in different images i and j. If there are at least eight correspondences between any two images, it is possible to compute the fundamental matrix by a linear method.

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

$$\boldsymbol{l}_{ji} = F_{ji} \boldsymbol{x}_i, \quad j \in \mathcal{I}', \quad i \in \mathcal{I} \tag{14}$$

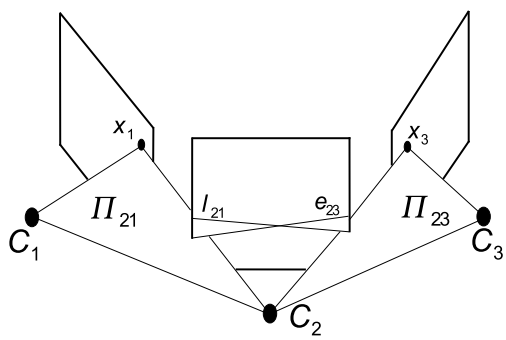


Figure 1: Epipolar lines corresponding to the same point observed in different images intersect in the same point. The camera centers are denoted by C_1, C_2 and C_3. The feature point observed in images 1 and 3 is denoted by x_1 and x_3, respectively. Here, l_{21} and l_{23} denote the epipolar lines 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.

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 $\boldsymbol{x}$, i.e.

$$\sum_{i \in \mathcal{I}} \boldsymbol{l}_i^\top \boldsymbol{x} = 0 \tag{15}$$

(cf. Fig. 1). The solution to Eq. (15) is the best estimation of the unobserved feature point $\boldsymbol{x}$ 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.

Since we do not know which epipolar lines intersect closest to the true position of the unobserved point, the unknown position of $\boldsymbol{x}_j$ in image $\mathcal{I}'$ may be determined by least-squares minimization of all the Eqs. (15). 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 $(u\ v\ w)_i^\top$ the feature in the first or second image, respectively. The joint system of Eqs. (15) then is as follows

$$\underbrace{\begin{bmatrix} \boldsymbol{l}_{13}^\top & \mathbf{0}_{2\times 3} \\ \boldsymbol{l}_{14}^\top & \\ & \boldsymbol{l}_{23}^\top \\ \mathbf{0}_{3\times 3} & \boldsymbol{l}_{24}^\top \\ & \boldsymbol{l}_{25}^\top \end{bmatrix}}_{L} \begin{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}_1 \\ \begin{bmatrix} u \\ v \\ w \end{bmatrix}_2 \end{bmatrix} = \mathbf{0}. \tag{16}$$

Algorithm 1 Constrained Projective Reconstruction

Input: Data vectors $\boldsymbol{x}_{\kappa\alpha}$, the minimal reprojection error $\epsilon_{\min}$, and the minimum difference between two consecutive reprojection error $\delta\epsilon_{\min}$
Output: Projection matrices P_κ and 3D points X_α.
Initialize the projective depths to $\lambda_{\kappa\alpha} = 1$, and the unobserved points $\tilde{\boldsymbol{x}}_{\kappa\alpha}$ to the centroid of all observed points.
Compute all columns $\boldsymbol{w}_\alpha$ of matrix W.
Compute a basis P of the subspace $\mathcal{L}$ consisting of the first four *left* singular vectors of W. The orthogonal complement of the subspace is given by $S_\perp = I - PP^\top$.
Let ϵ be the reprojection error defined by Eq. (18) and $\delta\epsilon$ be the difference between two consecutive reprojection errors.
while $\epsilon > \epsilon_{\min}$ or $\delta\epsilon > \delta\epsilon_{\min}$ **do**
  for $\alpha = 1$ to N **do**
    Determine the unknown measurements $\boldsymbol{x}_{\kappa\alpha}$ by solving the joint system defined by Eq. (17).
    Create the matrix D_α.
    Compute the projective depths $\boldsymbol{\lambda}_\alpha$ by minimizing Eq. (12c).
  end for
  Update the vectors $\boldsymbol{w}_\alpha$ in matrix W.
  Compute the projection matrices P_κ given by the row-triplets of P.
  Compute the 3D positions X_α given by the columns of $X = P^\top W$.
end while

Here, $\mathbf{0}_{i\times j}$ denotes a $i \times j$ matrix solely consisting of zeros. The upper right and lower left blocks of the matrix on the left side are identically zero. Denote by L_α the matrix containing all epipolar constraints corresponding to the αth trajectory on the left side of Eq. (16). It is possible to simultaneously optimize both the global subspace constraint and the local epipolar constraints by stacking the left and right sides of Eq. (13) and the ones due to the epipolar constraints:

$$\begin{pmatrix} S_\perp \Lambda_\alpha A_\alpha \\ L_\alpha \end{pmatrix} \boldsymbol{y}_\alpha = \begin{pmatrix} -S_\perp \Lambda_\alpha \boldsymbol{v}_\alpha \\ \mathbf{0} \end{pmatrix}. \tag{17}$$

We solve Eq. (17) by means of normal equations followed by normalization of each triple $\tilde{\boldsymbol{x}}_{\kappa\alpha}$ in $\boldsymbol{y}_\alpha$ to unit length.

The complete algorithm is summarized in Algorithm 1[1] .

[1]The sources will be made available for download upon acceptance of this work.

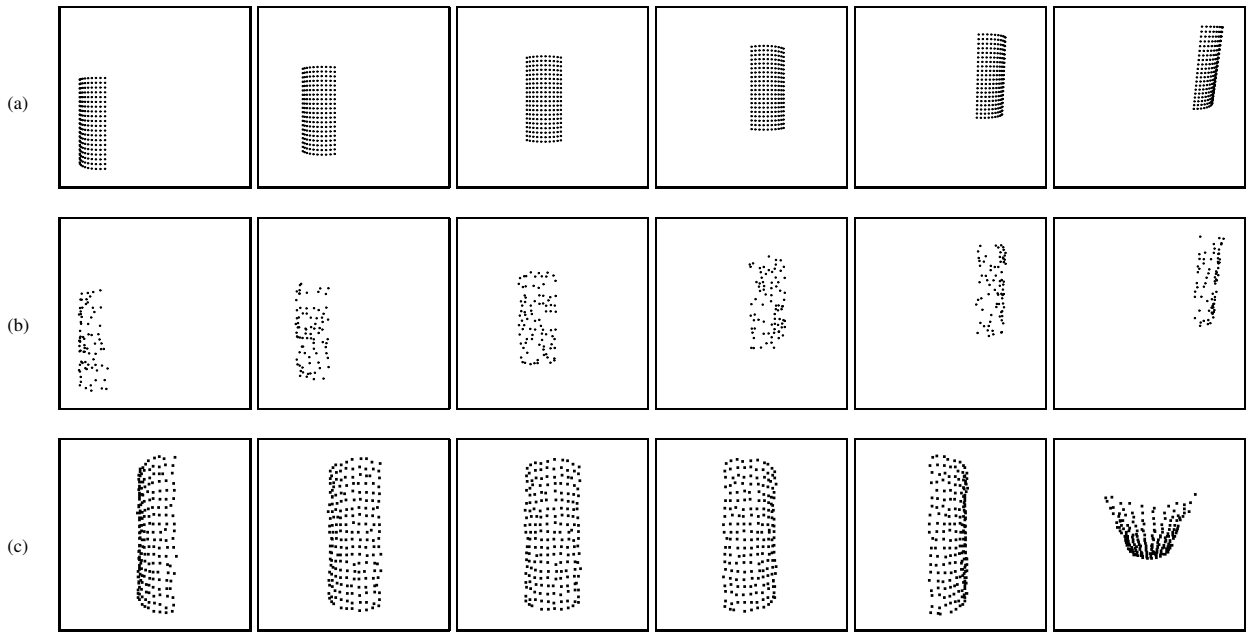


Figure 2: (a) Simulated image sequence of 200 points on a cylindrical surface through 20 frames (six images are shown here). (b) 6 images of the same sequence with 60% unobserved points and normally distributed noise with standard deviation $\sigma = 3.0$. (c) The 3D-reconstruction of the shape using the sequence shown in Fig. 2(b).

4 Experimental Results

In this Section we evaluate the algorithm introduced in this paper. In Sect. 4.1 , the performance is measured using synthetic image sequences. Let the reprojection error be

$$\epsilon = \frac{1}{s_0}\sqrt{\frac{1}{MN}\sum_{\kappa=1}^{M}\sum_{\alpha=1}^{N}\left\|\mathcal{N}\left[\boldsymbol{x}_{\kappa\alpha}\right]-\mathcal{N}\left[P_\kappa X_\alpha\right]\right\|^2} \qquad (18)$$

where $\mathcal{N}[\cdot]$ denotes normalization of the third coordinate to 1. The scale factor $s_0 = 1/600$ which we applied to all observed points $\boldsymbol{x}_{\kappa\alpha}$ needs to be considered in Eq. (18).

In Sec. 4.1, the proposed method is evaluated using a synthetic sequence. In Sec. 4.2, we present the results of our method using different indoor and outdoor real image sequences.

4.1 Synthetic Sequence

We created a synthetic image sequence of 200 points on a cylindrical surface shown in Fig. 2(a). The points were projected into images of size 512×512 pixels by a generally moving perspective camera with focal length 600 pixels. The sequence consists of 20 images.

The fundamental matrices were computed offline using RANSAC with the normalized 8-point algorithm [8]. Incomplete trajectories were simulated by randomly removing a certain percentage of points. This simulates trajectories as might be created using a SIFT tracker [14], where feature points are missing in one frame, but may reappear in subsequent images.

The accuracy of the algorithm was measured by computing the reprojection error as defined by Eq. (18). For the synthetic sequence, we computed the reprojection error for **all** points not just the observed ones. The 3D-error is measured by computing the homography which optimally projects the reconstructed shape to the ground-truth shape. The average Euclidean distance between the homography-projected 3D-points and the ground truth shape was adopted as measure.

The proposed method is evaluated with and without epipolar constraints. Utilizing only the rank-4 constraint makes this method similar to the approach of Tang and Hung [27][2]. The results of this algorithm are compared with the those of the algorithm proposed by Martinec and Pajdla [18, 26][3]. Since the proposed algorithm does not require an initial solution, we only compare it with [18] which also satisfies this requirement. The estimated solution can be used to initialize a bundle adjustment.

To measure the effect of missing data, the percentage of missing points was gradually increased from 10% until 60% in steps of 10%. We considered the algorithms to have converged successfully if the reprojection error was lower than 10 pixels.

The results of the algorithm of Martinec and Pajdla and the proposed method are shown in Fig. 3. The x-axis indicates increasing amounts of missing points. Results are shown for three different levels of normally distributed noise added to the points with standard deviations σ equal to 1.0, 2.0, and

[2]The method of Tang and Hung [27] employs a row-space approach whereas the proposed algorithm fits the subspace to the columns of W.

[3]The code is graciously provided by T. Svoboda in `cmp.felk.cvut.cz/~svoboda` [26].

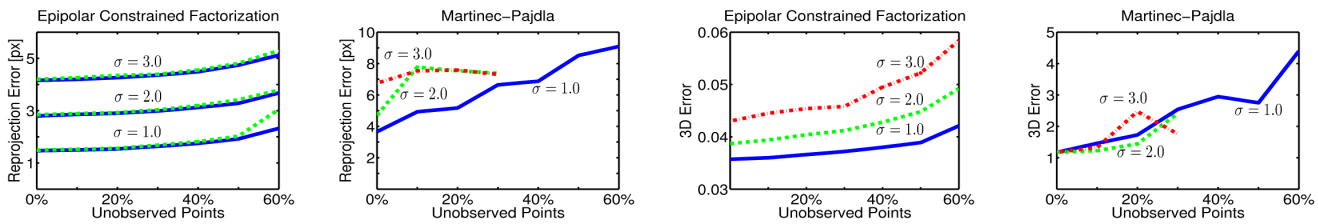


Figure 3: (Leftmost two images) Reprojection errors for the proposed method (left plot) and the method of Martinec and Pajdla [18] (right plot), respectively for three different noise levels with standard deviations $\sigma = 1.0,\ 2.0,\ 3.0$. For the right plot, the solid blue lines denote the mean error of 10 trials and the dashed green lines the maximum error. For the right plot, the solid blue line corresponds to the noise level $\sigma = 1.0$, the dashed green line to $\sigma = 2.0$, and the dash-dotted red line to $\sigma = 3.0$. The algorithm of Martinec and Pajdla failed to converge for any trial if the amount of missing points was larger than 30% and the noise $\sigma \geq 2.0$.
(Rightmost two images) Average 3D errors for the proposed algorithm (left plot) and the one of Martinec and Pajdla [18] (right plot). The solid blue lines indicate the noise level $\sigma = 1.0$, the dashed green line $\sigma = 2.0$, and the dash-dotted red line $\sigma = 3.0$. The algorithm of Martinec and Pajdla failed to converge for any trial if the amount of missing points was larger than 30% and the noise $\sigma \geq 2.0$.

3.0 pixels. The left two images in Fig. 3 show the minimal reprojection errors (in pixels). For our method (left plot), the solid blue lines indicate the mean errors of ten trials, and the dashed green lines the maximum errors. For the method of Martinec and Pajdla [18], the solid blue lines indicates the error corresponding to a noise level with $\sigma = 1.0$, the dashed green line $\sigma = 2.0$, and the dash-dotted red line $\sigma = 3.0$. Apparently, the proposed method achieves lower reprojection errors for almost all levels of noise and amounts of unobserved points (observe the different scales of the y-axes of the leftmost two plots of Fig. 3). The reprojection error almost linearly increases with noise, and our method is hardly affected by missing points. The algorithm of Martinec and Pajdla fails to converged for any of the ten trials if $\sigma \geq 2.0$ and more than 30% unobserved points, whereas the proposed method successfully converges for *all* ten trials for each noise level and amount of unobserved points. The right two images of Fig. 3 show the average 3D errors for both methods (left and right plots, respectively). The solid blue lines indicate the errors corresponding to a noise level with $\sigma = 1.0$, the dashed green lines $\sigma = 2.0$, and the dash-dotted red lines $\sigma = 3.0$. The proposed method is much more accurate than the one of Martinec and Pajdla (observe that the scales the y-axes in the two plots are different).

The algorithm using only the rank-4 constraint achieves approximately the same errors as the proposed algorithm. However, it fails to converge increasingly often with increasing noise and amount of unobserved points. Figure 4 shows how frequently the algorithm using only the rank-4 constraint and the one of Martinec and Pajdla converge for the three noise levels and increasing amounts of unobserved points. The proposed algorithm, i.e. factorization with epipolar constraints, always converges.

An example how the synthetic sequence of Fig. 2(a) looks like is shown in Fig. 2(b). Here, 60% of all points are not visible and normally distributed noise with standard deviation $\sigma = 3.0$ was added. For this sequences it is very difficult to perceive the structure of the object for a human observer. The 3D-reconstruction of the shape using the sequence of Fig. 2(b) is shown in Fig. 2(c). Apparently, the reconstructed shape looks reasonable although many points are missing and the noise is very strong.

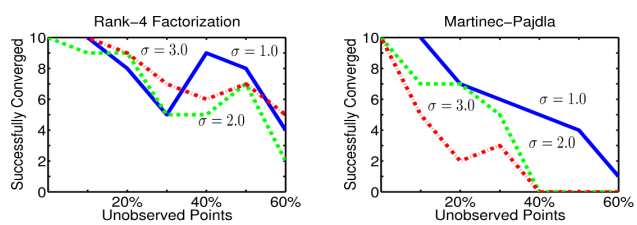


Figure 4: The left plot shows how often (out of 10 trials) the subspace method converged* without *epipolar constraints for each sampling and noise level. The right plot shows how often the algorithm of Martinec and Pajdla [18] converged. We defined success if the reprojection error for all points (not just the observed ones) was smaller than 10. If Factorization is employed* with *epipolar constraints, the algorithm always converges.

4.2 Real Image Sequences

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

First, we used the sequence of the model of a tea box consisting of 61 images of size 640×480 taken by the video mode of a Canon PowerShot A530 photo camera. One of the images is shown in the leftmost plot of Fig. 6(a). The publicly available *voodoo camera tracker*[4] was utilized for establishing a total of 430 correspondences. The amount of unobserved points

[4] www.digilab.uni-hannover.de/docs/manual.html

is approximately 52%. The leftmost diagram in Fig. 5 shows which feature was found in which image (white indicates that the feature was observed). All feature points visible in at least twenty images were utilized. A metric reconstruction was obtained using the algorithm from [12]. The algorithm converged to a reprojection error of 1.06 pixels. Three images of the reconstructed shape are shown in the rightmost plot in Fig. 6(a). We assigned each 3D point its color in the image it was first observed. The overall quality of the shape is satisfactory.

The second sequence shows an industrial dredger. A total of 564 points were tracked over 51 images of size 720×576 by a Canon X1-1 video camera . The first image of the sequence is shown in the leftmost plot of Fig. 6(b). The correspondences were again established using the *voodoo camera tracker*. In this sequence approximately 22% of the feature points were not 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 middle plot of Fig. 5. The algorithm converged to a reprojection error of 1.28 pixels. Three images of the reconstructed 3D-shape using the self-calibration algorithm from [12] are shown in the rightmost plots of Fig. 6(b). We assigned each 3D point the average color over all images it was observed in. As can be seen the reconstruction looks reasonable.

Lastly, we used a sequence showing a Lego spaceship. It consists of 90 images of size 720×576 taken by the Canon X1-1. One of the images is shown in the leftmost plot of Fig. 6(c). We used all of the provided correspondences which are visible in at least twenty images resulting in 495 trajectories. The unobserved correspondences amount to approximately 61%. The bottom diagram in Fig. 5 shows which feature was found in which image. The algorithm converged to a reprojection error of 2.79 pixels. We assigned each 3D point its color in the image it was first observed. The reconstructed 3D-shape using the self-calibration algorithm from [12] is displayed in the three rightmost plots of Fig. 6(c).

5 Summary

In this paper we deal with projective shape and motion reconstruction using subspace iteration methods. For such approaches it is a prerequisite that all feature points are observable in all images, a condition which is unrealistic in real videos.

We address this problem by using a geometrically meaningful measure to estimate unknown correspondences. The proposed algorithm uniformly utilizes all available data and finds a global solution without merging piece-wise reconstructions. Additionally to a global constraint – based on a subspace – we propose local constraints – based on epipolar geometries – to regularize and stabilize the estimations. We show that both types of constraints can be optimized simultaneously and allow for scene reconstruction in presence of strong noise and missing data. The proposed algorithm does not require any particular initialization.

Not many methods manage to perform a projective 3D-reconstruction without initialization. Algorithms known to us ([18, 27]) are based on the factorization approach. These methods fail on long sequences shown in Sec. 4.2. Neither does affine 3D-reconstruction ([2]) succeed on such scenes as the depth variation is large.

Other constraints, for instance homographies of a plane, can be easily included as well. The solution of this algorithm can be further used to initialize a statistical optimization technique such as bundle adjustment. It is possible to integrate our method into a RANSAC approach. The source code is available at [1].

References

[1] H. Ackermann and B. Rosenhahn. Sources of projective factorization with epipolar constraints: [removed for review], 2009.

[2] H. Ackermann and B. Rosenhahn. Trajectory Reconstruction for Affine Structure-from-Motion by Global and Local Constraints. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, USA, June 2009.

[3] S. Brandt. Closed-Form Solutions for Affine Reconstruction under Missing Data. In *7th European Conference on Computer Vision (ECCV)*, pages 109–114, Copenhagen, Denmark, 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, Washington, DC, USA, 2005.

[5] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[6] S. Christy and R. Horaud. Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 18(11):1098–1104, 1996.

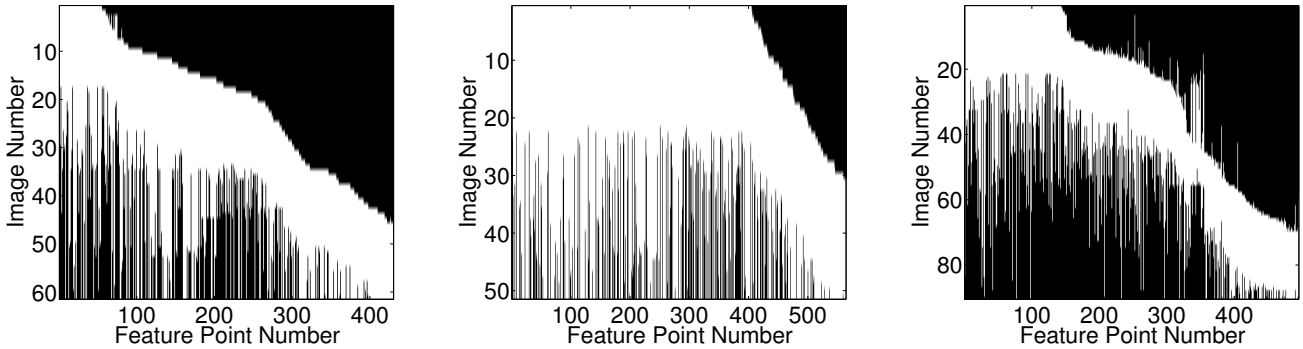
[7] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Dept. Electrical Engineering, Stanford University, March 2002.

[8] R. Hartley. In Defense of the Eight-Point Algorithm. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 19(6):580–593, June 1997.

[9] R. Hartley and F. Schaffalizky. PowerFactorization: 3D Reconstruction with Missing or Uncertain Data. In *Australia-Japan Advanced Workshop on Computer Vision*, June 2002.

[10] A. Heyden, R. Berthilsson, and G. Sparr. An Iterative Factorization Method for Projective Structure and Motion from Image Sequences. *Image and Vision Computing*, 17(13):981–991, November 1999.

[11] D. W. Jacobs. Linear Fitting with Missing Data for Structure-from-Motion. *Computer Vision and Image Understanding*, 82(1):57–81, April 2001.



(a) Tea box Sequence (b) Dredger Sequence (c) Space-Lego Sequence

Figure 5: Visibility of the tracked feature points, i.e. which point (indicated by the x-axis) was observed in which image (y-axis). The white color indicates that the feature could be observed, black that it was not observed. The upper figure corresponds to the tee-box sequence shown in Fig. 6(a) (missing data ratio of approximately 51%, the middle one to the dredger sequence shown in Fig. 6(b) (missing data ratio of approximately 22%) and the bottom diagram to the space-Lego sequence shown in Fig. 6(c) (missing data ratio of approximately 61%).

[12] K. Kanatani. Latest Progress of 3D-Reconstruction from Multiple Camera Images. *Robotics Research Trends*, pages 33–75, 2008.

[13] M. Lourakis and A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Transactions on Mathematical Software*, 36(1):1–30, 2009.

[14] D. G. Lowe. Object Recognition from Local Scale-Invariant Features. In *7th International Conference on Computer Vision (ICCV)*, pages 1150–1157, Corfu, Greece, 1999.

[15] S. Mahamud and M. Hebert. Iterative Projective Reconstruction from Multiple Views. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 430–437, Hilton Head, SC, USA, June 2000.

[16] S. Mahamud, M. Hebert, Y. Omori, and J. Ponce. Provably-Convergent Iterative Methods for Projective Structure from Motion. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1018–1025, Kauai Island, Hawaii, USA, December 2001.

[17] 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.

[18] 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.

[19] 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.

[20] J. Oliensis and R. Hartley. Iterative extensions of the Sturm/Triggs algorithm: Convergence and nonconvergence. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 29(12):2217–2233, 2007.

[21] A. Ruhe and P. Wedin. Algorithms for separable nonlinear least squares problems. *Society for Industrial and Applied Mathematics Review*, 22(3):318–337, 1980.

[22] H.-Y. Shum, K. Ikeuchi, and R. Reddy. Principal Component Analysis with Missing Data and its Application to Polyhedral Object Modeling. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 17(9):854–867, 1995.

[23] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM SIGGRAPH 2006*, SIGGRAPH 2006, pages 835–846, New York, NY, USA, 2006. ACM.

[24] P. Sturm and B. Triggs. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *4th European Conference on Computer Vision (ECCV)*, pages 709–720, Cambridge, UK, April 1996.

[25] 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.

[26] T. Svoboda, D. Martinec, and T. Pajdla. A Convenient Multi-Camera Self-Calibration for Virtual Environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4):407–422, August 2005.

[27] W. Tang and Y. Hung. A Subspace Method for Projective Reconstruction from Multiple Images with Missing Data. *Image and Vision Computing*, 24(5):515–524, May 2006.

[28] 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.

[29] 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.

[30] 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.

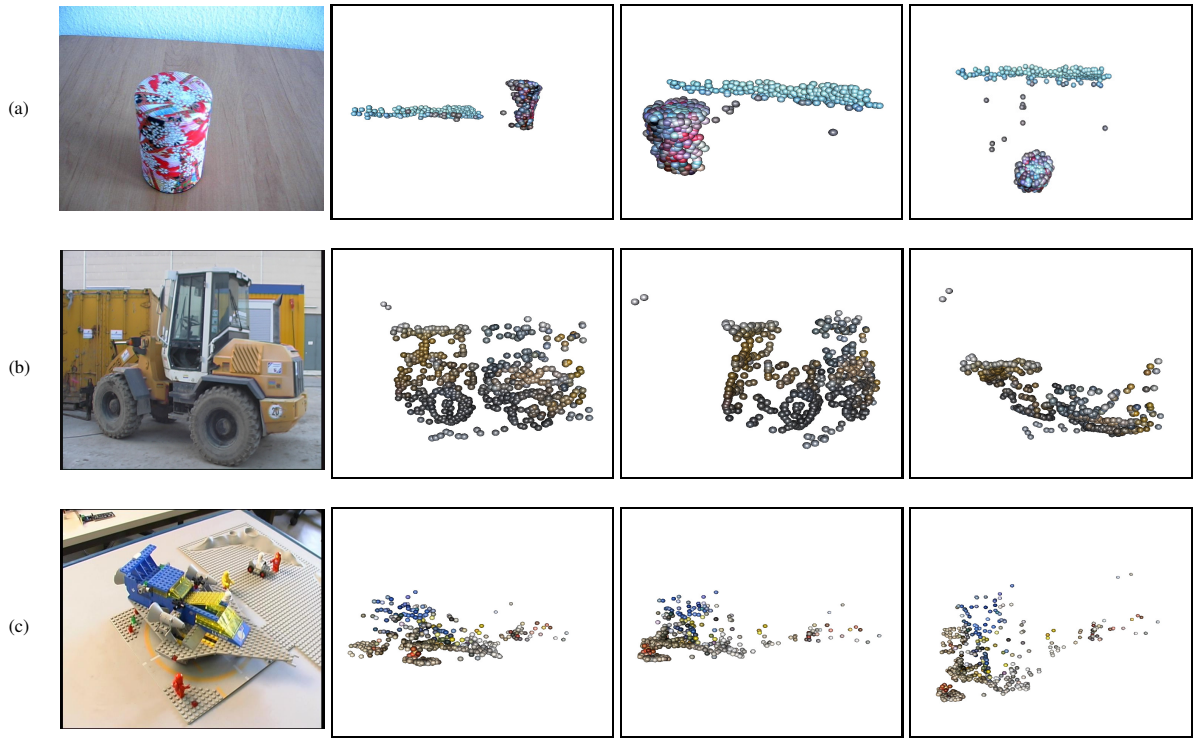


Figure 6: (a) Leftmost plot: One image of the 61***-frame sequence of a tea-box. The sequence consists of*** 430 ***trajectories. Rightmost three plots: reconstructed*** 3***D-shape using the algorithm from [12]. (b) Leftmost plot: One image of the*** 50***-frame sequence of an industrial dredger. The sequence consists of*** 564 ***trajectories. Approximately*** 22% ***of all points are not observed. The reconstructed*** 3***D-shape is shown in the rightmost three plots. (c) Leftmost plot: One image of the*** 90***-frame sequence of a Lego model consisting of*** 495 ***trajectories. The reconstructed*** 3***D-shape is shown in the rightmost three plots.***

[31] T. Ueshiba and F. Tomita. A Factorization Method for Projective and Euclidean Reconstruction from Multiple Perspective Views via Iterative Depth Estimation. In *5th European Conference on Computer Vision (ECCV)*, pages 296–310, Freiburg, Germany, June 1998.

[32] T. Wiberg. Computation of principal components when data are missing. In *Second Symp. on Computational Statistics*, pages 229–236, Berlin, Germany, 1976.

[33] H. Wold. Estimation of principal components and related models by iterative least squares. In Krishnaiah, editor, *Multivariate Analysis*, pages 391–420, 1966.