

# Optimal shape from motion estimation with missing and degenerate data

Manuel Marques and João Costeira  
Institute for Systems and Robotics - Instituto Superior Técnico  
Av. Rovisco Pais - 1049-001 Lisboa PORTUGAL  
{manuel,jpc}@isr.ist.utl.pt

## Abstract

*Reconstructing a 3D scene from a moving camera is one of the most important issues in the field of computer vision. In this scenario, not all points are known in all images (e.g. due to occlusion), thus generating missing data. The state of the art handles the missing points in this context by enforcing rank constraints on the point track matrix. However, quite frequently, close up views tend to capture planar surfaces producing degenerate data. If one single frame is degenerate, the whole sequence will produce high errors on the shape reconstruction, even though the observation matrix verifies the rank 4 constraint. In this paper, we propose to solve the structure from motion problem with degenerate data, introducing a new factorization algorithm that imposes the full scaled orthographic model in one single optimization procedure. By imposing all model constraints, a unique (correct) 3D shape is estimated regardless of the data degeneracies. Experiments show that remarkably good reconstructions are obtained with an approximate models such as orthography.*

## 1. Introduction

One of the most important issues in computer vision is definitely the structure from motion problem, where the object's structure and motion are obtained from image measurements. Considering the orthographic camera model, this task can be solved by the Tomasi-Kanade method [12]. More complex cameras have been considered in [8, 10]. Note that the referred algorithms are adequate only if all features points are visible in each image. Problems occur when some measurements are missing and this happens nearly always in real situations.

The problem which we propose to solve in this paper, is the 3D Reconstruction of an object's shape from an image stream with missing data. The assumed camera model

is the orthographic model. To minimize perspective effect, the image sequence should be produced by close-up views of the objects (these views quite frequently are of planar surfaces).

The missing data problem can be formulated as an optimization given by:

$$\mathbf{Problem\ 1} \quad (\hat{\mathbf{A}}, \hat{\mathbf{B}})^* = \arg \min_{\hat{\mathbf{A}}, \hat{\mathbf{B}}} \left\| (\mathbf{Z} - \hat{\mathbf{A}}\hat{\mathbf{B}}) \odot \mathbf{D} \right\|_F$$

where  $\mathbf{Z}$  and  $\mathbf{D}$  are the measurements and mask matrices, respectively. The mask matrix is a binary matrix identifying the known data with 1 and unknown data with 0.

According to the orthographic camera model, matrix  $\mathbf{Z}$  has rank 4 and this fact is used to estimate the unknown data. In [12], the missing data is sequentially replaced using complete subsets of the data. But, this first approach does not solve Problem 1 for generic configurations of  $\mathbf{D}$ , as proved by Jacobs [7]. The proposed approach in [7] is a non-iterative and sub-optimal algorithm where the measurement matrix verify the rank constraint. In the same way of Jacobs' approach, there are several algorithms [5, 11, 13] known as *batch algorithms*, because the solution (sub-optimal in presence of noise) is found in one global step. Due to this reason, this type of methods can be used to obtain an initialization to *iterative algorithm* where *alternation algorithms* play an important role. These last algorithms are based in the fact that if  $\mathbf{A}$  or  $\mathbf{B}$  are known, there is a closed-form solution for the other such that (Problem 1) is minimized.

Guerreiro and Aguiar approach [4] is similar to Aanaes et al [1], both algorithms project the data in a subspace in each iteration. The convergence of the referred methods is initially good but it is very susceptible to flatlining. Then, Buchanan [2] presented a Newton method to improve convergence.

The constraint used by these algorithms, the rank constraint, is not enough to obtain a correct estimate when there are images in the sequence where the 3D points of the known projections belongs to 1D or 2D subspaces. This happens because the optimization problem (1) has infinite *minimae*.

---

This work was partially supported by the Fundacao para a Ciencia e Tecnologia (ISR/IST pluriannual funding) through the POSC Program that includes FEDER funds, and project POCTI/AUR/48123/2002.

There are works that try to solve this ambiguity with more *priori* information. In [9], the planar surfaces of the object are known and due to this they impose these constraints to the object's shape and in [5, 11] a smooth camera trajectory is considered. But these last strategies can not be used to solve the problem presented in this paper, because we want to recover the 3D shape of an unknown object and without any constraints in the camera trajectory. We will show that this objective can be attained by solving problem (1) but constraining camera motion to comply with the rigidity constraints. To obtain this result, an intermediate step is required, leading to a new rigid factorization algorithm.

## 2. Rigid Factorization

In this section we will introduce a new factorization algorithm that computes shape and motion under scaled orthography in one single optimization step, and imposing the correct camera model upfront. We will show that this constraint is unavoidable if the known data is degenerate and because of this fact the affine factorization of Tomasi-Kanade does not provide the correct answer. Assume for now that there are no missing points on our data, and that the *known data matrix*  $\mathbf{Z}$  is given by

$$\mathbf{Z} = \mathbf{W} \quad (1)$$

where  $\mathbf{W}$  is called the *data matrix*. If one object is represented by  $P$  points view in  $F$  frames, this matrix is composed by the projections of the object's points in all images, such as

$$\mathbf{W} = \begin{bmatrix} u_1^1 & \dots & u_1^P \\ v_1^1 & \dots & v_1^P \\ \vdots & \ddots & \vdots \\ u_F^1 & \dots & u_F^P \\ v_F^1 & \dots & v_F^P \end{bmatrix} \quad (2)$$

where  $u_f^p$  and  $v_f^p$  are the  $p$  point projection in the frame  $f$ . Considering the scaled-orthographic camera model, (2) can be factorized as

$$\mathbf{Z} = \mathbf{M}\mathbf{S} + \mathbf{t}^T \mathbf{1} \quad (3)$$

where  $\mathbf{M}_{[2F \times 3]}$  is the *motion matrix*,  $\mathbf{S}_{[3 \times P]}$  the *shape matrix* and  $\mathbf{t} = [t_u^1 t_v^1 \dots t_u^F t_v^F]^T$  the *translation vector*. The *motion matrix* is composed by independent matrices like the next expression indicates.

$$\mathbf{M} = [\mathbf{M}_1^T \dots \mathbf{M}_F^T]^T$$

Each matrix  $M_f$  (also called the *motion matrix in frame f*) is composed by the two perpendicular vectors

$$\mathbf{M}_f = \begin{bmatrix} \mathbf{i}_x & \mathbf{i}_y & \mathbf{i}_z \\ \mathbf{j}_x & \mathbf{j}_y & \mathbf{j}_z \end{bmatrix}^f = \begin{bmatrix} \mathbf{i}^f \\ \mathbf{j}^f \end{bmatrix} \quad (4)$$

It is important to refer that the vectors  $\mathbf{i}^f$  and  $\mathbf{j}^f$  are not orthonormal vectors, as usual, but perpendicular with the

same norm (the scale factor). Since we are estimating the model's parameters without missing data, considering that the shape coordinate system is placed in the shape's centroid, the *translation vector* has a closed-form solution,

$$\mathbf{t} = \left[ \frac{1}{P} \sum_i \mathbf{Z}_{1i} \quad \dots \quad \frac{1}{P} \sum_i \mathbf{Z}_{[2f,i]} \right] \quad (5)$$

Replacing it in (3) we obtain the so called *centered data matrix*

$$\mathbf{Z}_c = \mathbf{M}\mathbf{S} \quad (6)$$

In the orthographic model (the considered model for the data), there are not constraints for  $\mathbf{S}$ . Instead  $\mathbf{M}$  has to verify the following *motion constraints*

$$\mathbf{i}^f \mathbf{i}^f = \alpha^f \quad (7)$$

$$\mathbf{j}^f \mathbf{j}^f = \alpha^f \quad (8)$$

$$\mathbf{i}^f \mathbf{j}^f = 0 \quad (9)$$

$$\forall f \in \{1, \dots, P\}$$

where  $\alpha^f$  is the scale factor of the frame  $f$  relatively to the reference one. Due to the Least Squares (LS) formulation of Tomasi-Kanade method, in the noisy data case, the obtained solution does not satisfy exactly these constraints (7,8,9).

### 2.1. Imposing Rigid Motion Constraints

To estimate the missing data that will be presented in section 3, the *motion constraints* must be satisfied. To do this, we replace the unconstrained optimization problem [12] by the following one

#### Problem 2

$$\begin{aligned} (\widehat{\mathbf{M}}, \widehat{\mathbf{S}})^* &= \arg \min_{\widehat{\mathbf{M}}, \widehat{\mathbf{S}}} \sum_f \|(\mathbf{Z}_c)_f - \widehat{\mathbf{M}}_f \widehat{\mathbf{S}}\|_F^2 \\ &\quad \widehat{\mathbf{M}}_1 \widehat{\mathbf{M}}_1^T = \alpha^1 I_{2 \times 2} \\ &\quad \vdots \\ &\quad \widehat{\mathbf{M}}_F \widehat{\mathbf{M}}_F^T = \alpha^F I_{2 \times 2} \\ &\quad \alpha^f \in \mathbb{R}^+, \forall f \end{aligned} \quad s.t.$$

The natural way of solving the Problem 2 is to search the motion matrix  $\widehat{\mathbf{M}}$  in the *motion manifold*, defined by the constraints of the referred problem. However, this is still a very hard problem. Instead we propose an iterative algorithm in  $\mathbb{R}^{2F \times P}$  where the solution found in each iteration (for the optimization problem without constraints) is projected in the *motion manifold*. It can be seen as a version of the power method where in each iteration a *motion matrix* is calculated (step 2) independently. The procedure of step 2, which projects the left factor of  $\mathbf{Z}_c$  ( $[\mathbf{R}_1^T \dots \mathbf{R}_F^T]^T$ ) onto the manifold of motion matrices, has a similar derivation to the Procrustes problem [3]. The solution of this problem for each matrix  $\mathbf{M}_f$  (4) in each iteration  $k$  is given by:

$$\begin{aligned} \widehat{\mathbf{M}}_f^k &= \alpha_f U_f V_f^T, \text{ where } R_f = U_f \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} V_f^T \\ &\text{and } \alpha_f = (\sigma_1 + \sigma_2)/2 \end{aligned}$$

---

**Algorithm 1** Rigid Factorization
 

---

**1. Initializations:**

 (factorize  $\mathbf{Z}_c$  using any factorization (e.g. SVD))

$$\mathbf{Z}_c = \mathbf{A}\mathbf{B}, \quad \mathbf{R} = \mathbf{A}, \quad \widehat{\mathbf{M}}^0 = \mathbf{A}, \quad \widehat{\mathbf{S}}^0 = \mathbf{B}$$

$$k = 1$$

**2. Project R into the manifold of motion matrices**

$$\widehat{\mathbf{M}}^k = \arg \min_{\mathbf{X}} \sum_f \|\mathbf{R}_f - \mathbf{X}_f\|_F^2$$

$$\text{s. t. } \mathbf{X}_f \mathbf{X}_f^T = \alpha_f \mathbf{I}_{2 \times 2} \quad \forall f$$

$$\alpha \in \mathbb{R}^+$$

**3.  $\widehat{\mathbf{S}}^k = \widehat{\mathbf{M}}^k{}^+ \mathbf{Z}_c$ ,  $\widehat{\mathbf{M}}^k{}^+$  - Moore-Penrose pseudoinverse**

**4.  $\mathbf{R} = \mathbf{Z}_c \widehat{\mathbf{S}}_k^+$**

**5. Verify if  $\|\widehat{\mathbf{M}}_k - \widehat{\mathbf{M}}_{k-1}\| < \epsilon$ .**

 If not, go to step 2 and  $k = k + 1$ .

**6.  $\widehat{\mathbf{M}} = \widehat{\mathbf{M}}_k$  and  $\widehat{\mathbf{S}} = \widehat{\mathbf{S}}_k$** 


---

Even though step 2 is solved  $F$  times in each iteration, the computational cost of this is irrelevant because as equation (10) shows, it has closed-form solution and is unique. In step 3, the estimate of the object's shape is given by a LS solution.

In terms of complexity, we need to compute two pseudo-inverses and a set of eigen vectors in closed-form (in  $\mathbb{R}^3$ ).

### 3. Estimating Shape With Missing Data

To represent the missing data, we introduce the matrix  $\mathbf{D}$ . Elements  $\mathbf{D}_{[2*i-1,j]}$  and  $\mathbf{D}_{[2*i,j]}$  of this matrix are 1 if point  $j$  is known in frame  $i$  and 0, otherwise. Due to missing data, the equations presented in section 2 are not valid in this one. In this way, (3) is replaced by

$$\mathbf{Z} = \mathbf{W} \odot \mathbf{D} \quad (10)$$

where  $\odot$  represents the Schur (elementwise) product. According to (10) and the orthographic model, we have

$$\mathbf{Z} = (\mathbf{M}\mathbf{S} + \mathbf{t}^T) \odot \mathbf{D} \quad (11)$$

In this section,  $\mathbf{t}$  is impossible to calculate such as (5) because this expression requires that all data are known. If the point  $p$  is known in all frames, the *translation vector* is computed as

$$\mathbf{t}^T = -\mathbf{Z} \underbrace{[0 \dots 0]_{p-1}}_{p-1} \underbrace{[1 \ 0 \dots 0]_{P-p}}_{P-p}^T = -[z_{1p} \quad \dots \quad z_{[2f,p]}] \quad (12)$$

and all measurements are registered to this point. If not this vector is another parameter to estimate.

#### 3.1. Using the rank constraint

The rank constraint can be used to calculate the unknown values of  $\mathbf{Z}$ . This fact is true because, by checking the model (11) and the dimensions of  $\mathbf{M}$  and  $\mathbf{S}$ , we can easily conclude that  $\mathbf{Z}$  has at most rank 4. Mathematically, this problem corresponds to solve the following optimization problem

**Problem 3**

$$(\widehat{\mathbf{W}})^* = \arg \min_{\widehat{\mathbf{W}}} \left\| \left( \mathbf{W} - \widehat{\mathbf{W}} \right) \odot \mathbf{D} \right\|_F^2$$

$$\text{s.t. } \widehat{\mathbf{W}} \in \mathcal{S}_4$$

where  $\mathcal{S}_4$  is the space of matrices with rank equal to 4. This approach allows us to obtain the correct solution for the missing data if the object is a **full 3D object** and the known points of each image of the stream are **not coplanar**. The solution is not the correct one if it exists, at least, one image where the known points are in the same 1D or 2D subspace (degenerate images). This happens because the problem 3 has infinite solutions.

To verify the described fact, consider that the *translation vector* is computed and, due to this we know  $\mathbf{Z}_c$  given by

$$\mathbf{Z}_c = \mathbf{M}\mathbf{S} \odot \mathbf{D}$$

where  $\mathbf{Z}_c \in \mathcal{S}_3$ . Suppose that the first  $m$  points belong to a 2D flat surface and the known points of the image  $f$  are these points, precisely. The object's shape is known and we can calculate a matrix  $\mathbf{A}$  such that the plane  $z = 0$  contains the first  $m$  points.

$$\underbrace{\begin{bmatrix} (\mathbf{Z}_c)_{[2f-1,1]} & \dots & (\mathbf{Z}_c)_{[2f-1,m]} & ? \dots ? \\ (\mathbf{Z}_c)_{[2f,1]} & \dots & (\mathbf{Z}_c)_{[2f,m]} & ? \dots ? \end{bmatrix}}_{\widehat{\mathbf{W}}_f} = \quad (13)$$

$$\underbrace{\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}}_{\widehat{\mathbf{M}}_f} \mathbf{A} \mathbf{A}^{-1} \underbrace{\begin{bmatrix} \widehat{\mathbf{S}}_{11} & \dots & \widehat{\mathbf{S}}_{1m} & \widehat{\mathbf{S}}_{[1,m+1]} \dots \widehat{\mathbf{S}}_{1P} \\ \widehat{\mathbf{S}}_{21} & \dots & \widehat{\mathbf{S}}_{2m} & \widehat{\mathbf{S}}_{[2,m+1]} \dots \widehat{\mathbf{S}}_{2P} \\ 0 & \dots & 0 & \widehat{\mathbf{S}}_{[3,m+1]} \dots \widehat{\mathbf{S}}_{3P} \end{bmatrix}}_{\widehat{\mathbf{S}}}$$

This particular parametrization just makes completely clear that the submatrix of the first  $m$  columns of  $\mathbf{S}$  is singular (the planar surface). Equation (13) is equivalent to the system of equations given by

$$\begin{cases} (\mathbf{Z}_c)_{[2f-1,1]} & = a\mathbf{S}_{11} + b\mathbf{S}_{21} \\ (\mathbf{Z}_c)_{[2f,1]} & = d\mathbf{S}_{11} + e\mathbf{S}_{21} \\ \vdots & = \vdots \\ (\mathbf{Z}_c)_{[2f-1,m]} & = a\mathbf{S}_{1m} + b\mathbf{S}_{2m} \\ (\mathbf{Z}_c)_{[2f,m]} & = d\mathbf{S}_{1m} + e\mathbf{S}_{2m} \end{cases} \quad (14)$$

In (14) we can clearly verify that the Problem 3 has two degrees of freedom, variables  $c$  and  $f$ . Then, the unknown image points projection in frame  $f$  have not two solutions, but infinite. This fact does not allow us to recover the original *motion* and *shape matrix* because the method can add an error to the known points' projections in this frame. In real situations with noise and other distortions these estimated projections can be quite far from reality. If, as it is common in real situations, there are large sets of degenerate images the total error can be significant.

### 3.2. Imposing the orthogonality constraints

Considering the orthographic camera model to explain the data, we can verify, in section 3.1, the rank constraint is not enough to estimate the missing data. To solve this problem, we propose the following optimization problem

**Problem 4**

$$\begin{aligned}
 (\widehat{\mathbf{M}}, \widehat{\mathbf{S}})^* &= \arg \min_{\widehat{\mathbf{M}}\widehat{\mathbf{S}}} \sum_f \left\| \left( \mathbf{W}_f - \widehat{\mathbf{M}}_f \widehat{\mathbf{S}} + \widehat{\mathbf{t}}_f \mathbf{1}_{[2,P]} \right) \odot \mathbf{D} \right\|_F^2 \\
 &\quad \widehat{\mathbf{M}}_1 \widehat{\mathbf{M}}_1^T = \alpha^1 I \\
 &\quad \vdots \\
 &\quad \widehat{\mathbf{M}}_F \widehat{\mathbf{M}}_F^T = \alpha^F I \\
 &\quad \alpha^f \in \mathbb{R}^+, \quad \forall_f
 \end{aligned}$$

With this new approach (Problem 4), the *motion constraints* (7,8,9) are added to (14). Then, the variables  $c$  and  $f$  have not an infinite number of solutions but only the two solutions.

---

#### Algorithm 2 Rigid Factorization with Missing Data

---

1. Initializations:  $\widehat{\mathbf{Z}}_0 = \mathbf{Z}, \quad k = 0$

2. Estimate translation (centroid).

$$\widehat{\mathbf{t}}_k = \left[ \frac{1}{P} \sum_i \widehat{\mathbf{Z}}_{1ik} \quad \dots \quad \frac{1}{P} \sum_i \widehat{\mathbf{Z}}_{[2f,i]k} \right] \quad (12)$$

$$\widehat{\mathbf{Z}}_{c_k} = \widehat{\mathbf{Z}}_k - \widehat{\mathbf{t}}_k \quad \text{Remove translation}$$

$$k = k + 1$$

3. Estimate  $\widehat{\mathbf{M}}_k$  e  $\widehat{\mathbf{S}}_k$  Using Rigid Factorization

4. Update data matrix

$$\widehat{\mathbf{Z}}_k = \underbrace{(\widehat{\mathbf{M}}_k \widehat{\mathbf{S}}_k + \widehat{\mathbf{t}}_{k-1} \mathbf{1}_{[2F,P]}) \odot \bar{\mathbf{D}}}_{\text{Missing data estimate}} + \underbrace{\mathbf{Z} \odot \mathbf{D}}_{\text{Known data}}$$

$$\bar{\mathbf{D}} - 2\text{'s complement of } \mathbf{D} \text{ i.e. } \bar{\mathbf{D}} = \mathbf{1}_{[2F,P]} - \mathbf{D}$$

5. Verify if  $\|\widehat{\mathbf{Z}}_k - \widehat{\mathbf{Z}}_{k-1}\| < \epsilon$ .

If not verify go to step 2 and  $k = k + 1$ .

---

The two solutions correspond to the correct one, and the reflection of the camera over the plane (shape). This reflection produces the same image (it is intrinsic to orthography). Even though motion can be "reflected" the computed shape will always be correct. Note that in case known data is over a line (rank 1 known shape) there will be infinite solutions for motion too, however this will not affect the shape either (orthography is valid). Then, to compute the structure from motion it is proposed the iterative **Algorithm 2**.

In step 3, it is used an algorithm which solution verifies (7,8,9), such as algorithm 1 described previously.

## 4. Experiments

### 4.1. Hotel Experiment

Benchmark tests were performed against the state-of-the-art, using Buchanan&Fitzgibbon's matlab code<sup>1</sup>. We

<sup>1</sup>This package contains several high performance methods and was made generously available in [www.robots.ox.ac.uk/~abm](http://www.robots.ox.ac.uk/~abm)

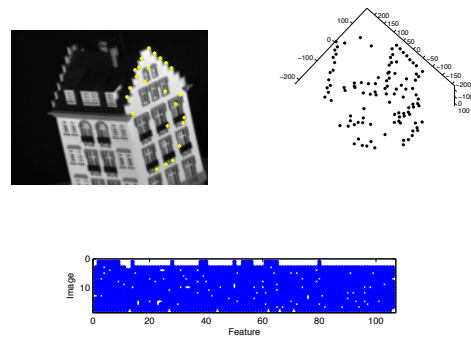


Figure 1. Up Left - Image feature tracks. Up Right - The reference 3D shape, computed with Tomasi-Kanade's factorization. Bottom - Patte:

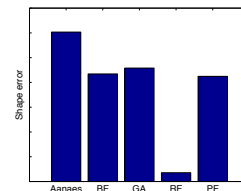
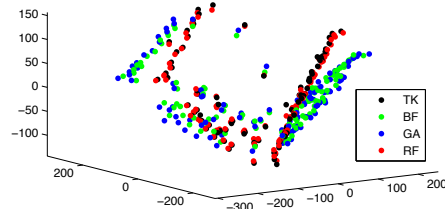


Figure 2. The top figure compares several methods by showing a top view of the hotel's 3D shape. Besides the reference shape (TK), we show the proposed method (RF) and two other (BF and GA). The graph in the bottom figure plots the relative error of several methods showing that none of them can compute shape adequately

modified the known hotel sequence<sup>2</sup> selecting all 106 feature points and 18 equally spaced frames from the total of 180. The matrix measurement has only 14% of missing features and two of those frames were artificially made degenerate with only 24 features visible, all lying on a planar surface (the rightmost wall of the hotel, shown in figure 1 top right). Since there is no ground-truth, we used the object's shape computed using Tomasi-Kanade's factorization method with full observation matrix (no missing data), as the reference shape (figure 1 middle).

In figure 2 we show the performance of our (rigid factorization) algorithm against 4 state-of-the-art methods: BF- Buchanan&Fitzgibbon's Damped Newton method[2], PF- Vidal&Hartley's power factorization [14], GA- Guerreiro&Aguiar EM (alternate) algorithm [4], Aanaes- Aanaes et al [1]. To avoid graphical clutter, in the top plot of figure 2 we show the reconstruction of two methods - GA and

<sup>2</sup><http://vasc.ri.cmu.edu/idb/html/motion/index.html>

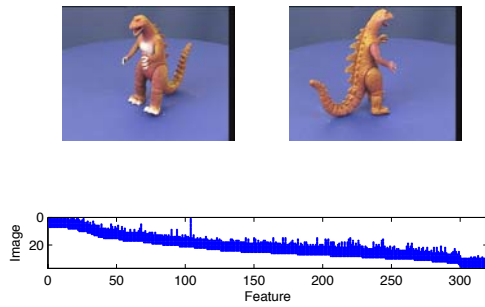


Figure 3. Up - Two images of the sequence Bottom - Pattern of missing points

BF (an alternate algorithm vs. a Newton algorithm) -, the "ground-truth" (TK) and our method (RF). The figure was generated from a top view so that the 3 planes of the hotel walls can be perceived clearly. Since both methods seek the best rank 4 matrix, like all other methods, they reach similar solutions, both inadequate (figure 2). As expected by enforcing rigidity constraints the reconstruction is quite close to the reference. In figure 2 (bottom), each bar represents the shape error  $\sum |\hat{S}_i - S_i^{tk}|$ , where  $\hat{S}_i$  is the shape estimate for point  $i$  and  $S_i^{tk}$  is the reference shape for that point. Of course the absolute value of the error depends on the particular shape, but the relevant aspect is that all other methods obtain estimates that are at least one order of magnitude above our method, and this happens with only 14% of data missing.

#### 4.2. Dinosaur Experiment

In the second experiment with real data, the proposed algorithm was tested with the dinosaur sequence<sup>3</sup> (figure 3). The sequence is composed by 36 images and does not contain any image with degenerate data. The measurement matrix is sparse with only 28% of known data (see figure 3).

Since there is no full matrix in this experiment, in opposite to the previous one, the "ground-truth" is not available and due to this reason the numerical evaluation of the results becomes very hard. In figures 4, we can see two different views of the dinosaur's shape which was obtained by the new method. Checking these figures, the obtained shape seems to be correct.

#### 4.3. Full reconstruction with largely scaled images

Finally we present a real life example using the rigid factorization with missing data. The aim is to produce a 3D reconstruction of one building from a set of uncalibrated images. We searched on Google for images depicting the building from a quite diverse set of viewpoints. The image scale is also quite diverse, since there are images taken

<sup>3</sup>www.robots.ox.ac.uk/~abm

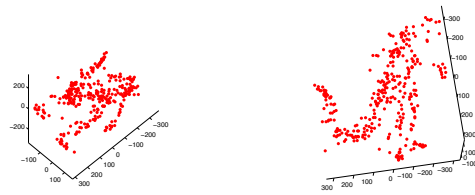


Figure 4. Two views of the dinosaur's shape obtained by the proposed method

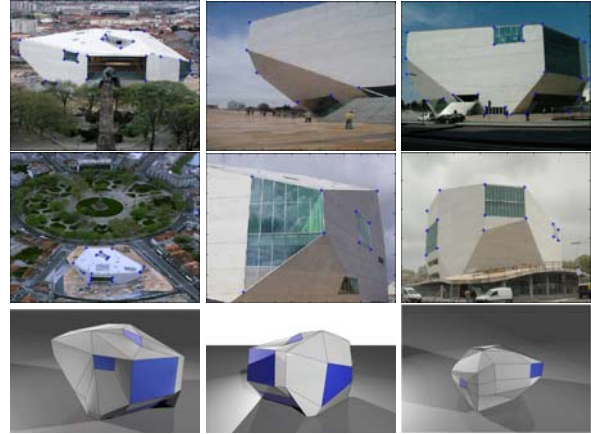


Figure 5. Reconstruction of a famous Rem Koolhaas "piecewise planar" building.

from a pedestrian capturing one window together with far off aerial views from an airplane. Resolution ranged from 3 Megapixel to a 60 Kilopixel and perspective effects were quite large in some of them. Image features were tracked by hand and were basically the vertices of the building and windows' corners. In figure 5 we show some of the 19 images with the features superimosed. Close up views with a large depth range produce strong perspective, therefore we just inserted points that were in a small depth range compared to the distance to the camera. Even though it favors the camera model adequacy it generates high volumes of (possibly degenerate) missing data. A triangulation was computed to convey the shape in a more natural way. The blue surfaces are windows. As figure 5 shows, the reconstruction is quite faithful. However perspective effects are noticeable, specially in the large front window. Nevertheless, in our opinion, this is a quite a hard set of images to any 3D reconstruction algorithm and remember that no prior knowledge was used. For more precise applications this can be a good starting point to perspective factorization algorithms such as [6, 10].

#### 4.4. Convergence

To analyze the algorithm's performance, several synthetic experiments were done. Convergence was evaluated as a function of the total number of points and the number of

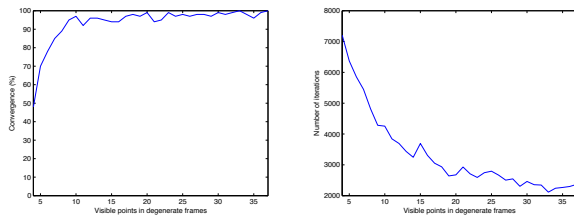


Figure 6. Left - Convergence: % of converged trials vs. Number of visible points in degenerate frames. Right - Number of iterations vs. the same parameters as before

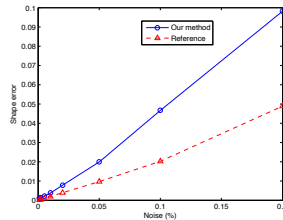


Figure 7. Shape error as a function of noise and number of visible points

visible points. The tested object was composed of 3 faces of a cube and 37 features were in each of them. In each experiment, a total of 21 images were generated and 15 of them were degenerate. The missing data in the remaining 6 (non-degenerate) images was set to 30%.

Figure 6 plots graphs of the ratio and speed of convergence. One hundred random experiments were made for each parameter (#visible points). Checking figure 6 (left) we can verify that to obtain good estimations (rates of convergence above 97%), we must know at least 8 points per each degenerate image. The convergence's velocity changes in accordance to the known projections in degenerate frames (figure 6 - right): the new algorithm is faster when more projections are known.

The graphs of figure 7 illustrate the algorithm's behavior with noise. In this experiment, the parameter value was chosen such that a high convergence rate is achieved: the number of known projections in degenerate images (15 in 21 also) is 13. The figure of merit to the algorithm's evaluation is shape error. These results are compared with the reference shape, computed by TK factorization with full matrix. Observing figure 7, we can verify that the error shape is proportional to that of the reference shape.

In summary, the relevant conclusion here is that the proposed algorithm converges to the correct shape, when there are at least 8 known points in each degenerate image.

## 5. Conclusions

We have presented a new factorization algorithm that computes the optimal motion shape and scale estimates (scaled orthography) from a feature track matrix. We also

introduced an iterative algorithm that produces the same estimates when feature points are missing and the known points are degenerate. To our knowledge no other algorithm can solve this problem. We have presented a solution of the structure from motion problem to the most general and realistic situation.

## References

- [1] H. Aanaes, R. Fisker, K. Astrom, and J. M. Carstensen. Robust factorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1215–1225, 2002.
- [2] A. Buchanan and A. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 316–322, 2005.
- [3] G. Golub and V. Loan. Matrix computations.
- [4] R. Guerreiro and P. Aguiar. Estimation of rank deficient matrices from partial observations: Two-step iterative algorithms. *Proc. Conf. Energy Minimization Methods in Computer Vision and Pattern Recognition.*, 2003.
- [5] N. Guilbert, A. Bartoli, and A. Heyden. Affine approximation for direct batch recovery of euclidean motion from sparse data. *Int. Journal of Computer Vision*, 69(3):317–333, September 2006.
- [6] A. Heyden, R. Berthilsson, and G. Sparr. An iterative factorization method for projective structure and motion from image sequences. *Image Vision Comput.*, 17(13):981–991, 1999.
- [7] D. Jacobs. Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images. In *IEEE International Conference on Computer Vision*, pages 206–212, 1997.
- [8] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(3):206–218, 1997.
- [9] G. Sparr. Euclidean and affine structure/motion for uncalibrated cameras from affine shape and subsidiary information. In *SMILE Workshop on Structure from Multiple Images, Freiburg*, 1998.
- [10] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion, 1996.
- [11] J.-P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert, and S. Roy. Algorithms for batch matrix factorization with application to structure-from-motion. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2007.
- [12] C. Tomasi and T. Kanade. Shape and motion from image stream under orthography: a factorization method. *Int. Journal of Computer Vision*, 9(2):137–154, 1992.
- [13] B. Triggs. Linear projective reconstruction from matching tensors. *Image and Vision Computing*, 15(8):617–625, 1997.
- [14] R. Vidal and R. Hartley. Motion segmentation with missing data using powerfactorization and gpca. *Proceedings CVPR*, 2:310–316, 2004.