Structure from Motion by Fast Alternation[1)]

Levente Hajder, Ákos Pernek, Csaba Kazó
Computer and Automation Research Institute, Hungary
and Budapest University of Technology and Economics
hajder@sztaki.hu, http://visual.ipan.sztaki.hu

Abstract:
We address the problem of moving object reconstruction. Several methods have been published in the past 20 years including stereo reconstruction as well as multi-view factorization methods. In general, reconstruction algorithms estimate the 3D structure of the object and the camera parameters in a non-optimal way and then a nonlinear optimization method refines the estimated camera parameters and 3D object coordinates.

In this paper, an adjustment method is proposed which is the fast version of the well-known down-hill alternation method. The novelty which yields the high speed of the algorithm is that the steps of the alternation give optimal solution to the subproblems by closed-form formulas. The proposed algorithm is discussed here and it is compared to the widely used bundle adjustment method.

1 Introduction

3D motion-based object reconstruction is a challenging and developing area of computer vision. Many of the published rigid SfM methods are based on various extensions of the well-known factorization method [6] by Tomasi and Kanade. The original method assumes orthographic projection, while its extensions can cope with weak-perspective [8] as well as real projection [5].

It is well known, that there is no optimal closed-form solution to the problem. For this reason, researchers have proposed non-optimal methods and then they try to minimize the so-called reprojection error (or other cost function) by adjustment algorithms such as bundle adjustment [2]. It is also well known that these classical algorithms [6, 8, 5] cannot cope with realistic situations: they assume that all features points are tracked over all frames. For this reason, several methods have been published to solve the case when data are missing or uncertain. (The Reader is referred to [3] for an excellent review.) However, the full matrix factorization is also an important issue, because Monte Carlo type robust algorithms [7, 4]

[1)] This work was supported by EU Network of Excellence MUSCLE (FP6-507752)

can use it to compute the camera motion from few points.

The proposed method assumes the weak-perspective camera model which is applicable if the depth of the object is much smaller than the object-camera distance, but it can be extended easily to the perspective case by using the so-called rescaled measurement matrix [5]. The proposed down-hill alternation method is very similar to the one published by us in 2006 [4]. The main difference is that the proposed method is faster, because the computation steps are given in closed forms.

2 Summary of the Tomasi-Kanade factorization

Given P feature points of a rigid object tracked across F frames, $x_{fp} = (u_{fp}, v_{fp})^T$, $f = 1, \ldots, F$, $p = 1, \ldots, P$, the goal of SfM is to recover the structure of the object.

If the origins of 2D coordinate system (at the centroid) is substracted from the trajectories in all images, the 2D coordinates are calculated as

$$x_{fp} = q_f R_f s_p, \tag{1}$$

where $R_f = [r_{f1}, r_{f2}]^T$ is the orthonormal rotation matrix, s_p the 3D coordinates of the point, q_f is the nonzero scale factor of weak perspective. For all points in the f-th image, the above equations can be rewritten as

$$\underbrace{W_f}_{2\times P} = (x_{f1} \ldots x_{fP}) = \underbrace{M_f}_{2\times 3} \cdot \underbrace{S}_{3\times P} \tag{2}$$

where M_f is called the motion matrix, $S = (s_1, \ldots, s_P)$ the structure matrix. Under orthography $M_f = R_f$, under weak respective $M_f = q_f R_f$. For all frames, the equations (2) form

$$\underbrace{W}_{2F\times P} = \underbrace{M}_{2F\times 3} \cdot \underbrace{S}_{3\times P}, \tag{3}$$

where $W^T = [W_1^T, W_2^T, \ldots, W_F^T]$ and $M^T = [M_1^T, M_2^T, \ldots, M_F^T]$.

The Tomasi-Kanade factorization method [6] has two main steps:
Step 1. In the noise-free case, W has rank 3. When noise is present, the method reduces the rank of the noisy measurement matrix by SVD. In this step, W is factorized into affine motion and affine structure: $W = M_{aff} S_{aff}$.
Step 2. In the reduced space, the method applies a linear transformation to obtain the real (Euclidean) motion and structure data. The transformation is represented by a matrix, Q, and the estimated motion and structure data are written as $M = M_{aff} Q$ and $S = Q^{-1} S_{aff}$, respectively. The matrix Q can be optimally determined by imposing orthographic [6] or weak-perspective [8] motion constraints. The final factorization is written as $W = MS$, where

the matrix S consists of the 3D coordinates of the moving object, and the matrix M contains the 3D directions of base vectors of the camera planes.

If few feature points are tracked, the quality of the Tomasi-Kanade factorization is not sufficient, and this is mainly caused by the SVD-based rank reduction. This reduction technique needs a relatively large number of feature points to obtain a good estimation. The SVD method determines a 3D subspace, and, despite the corrective transformation Q, the result cannot leave this subspace.

3 Proposed method: Fast Alternation

As it is discussed in the introduction, the Tomasi-Kanade factorization does not produce optimal solution. We propose a down-hill alternation method to minimize the reprojection error

$$||W - MS||_F^2 \tag{4}$$

subject to $M_f M_f^T = q_f I, \forall f$.

Our method guarantees that a local optimum will be reached starting from the initial matrices $M^{(0)}$ and $S^{(0)}$ yielded by the Tomasi-Kanade factorization. Unfortunately, global optimum can not be guaranteed.

The proposed method is an iterative one consisting of three main steps: the completion step, the S-step and the M-step. It runs until the difference between the subsequent error values drops below a given limit ϵ. The whole algorithm is overviewed in Alg. 1.

We call this algorithm **Fast Alteration**, because its speed is (relatively) high despite the alternation: all the parameter estimation of the substeps can be written in closed forms. The slowest calculation is the pseudo-inverse computation of a $2F \times 3$ motion matrix.

Completion step. The motion submatrix M_f consists of two rows: $M_f = [m_{f,1}^T m_{f,2}^T]^T$. Let us denote the completed matrix with $\tilde{M}_f = [m_{f,1}^T m_{f,2}^T m_{f,3}^T]^T$ where the direction of $m_{f,3}^T$ is parallel to the cross product of $m_{f,1}^T$ and $m_{f,2}^T$, its length is the average length of the vectors $m_{f,1}^T$ and $m_{f,2}^T$.

Based on the above strategy, the measurement matrix W is also completed. $\tilde{W}_f$ is the completed version of W_f if its third row is calculated as $m_{f,3}^T S$.

S-step. The goal of the S-step is to determine the structure matrix S optimally. The structure matrix is not constrained: its value can take arbitrary real values. The optimal value of $S^{(k)}$ is provided by the well known least-squares optimization:

$$S^{(k)} = M^{(k-1)^\dagger} \tilde{W}_f^{(k-1)} \tag{5}$$

where $M^{(k-1)^\dagger}$ denotes the Moore-Penrose pseudo-inverse of matrix $M^{(k-1)}$.

M-step. At the M-step, $M^{(k)}$ is estimated from $\tilde{W}^{(k)}$ and $S^{(k)}$.

It is trivial that the value of M_i is independent from that of M_j, if $i \neq j$. The main idea is that the calculation of M_f from W_f and S is a quasi-registration problem: the only difficulty is that S consists of 3-dimensional vectors, while W_f only contains 2-dimensional vectors. This problem can be eliminated if W_f and M_f are completed as it is described earlier in this section.

After completion, both the completed measurement matrix $\tilde{W}$ and the structure matrix S consist of p piece of 3D vectors. The estimated (complete) motion matrix $\tilde{M}$ can be decomposed into the product of a scale parameter q_f and an orthonormal matrix R_f.

According to [1], the rotation is optimally calculated as $R_f = V_f U_f^T$ if $H_f = U_f \Lambda_f V_f^T$ is the singular value decomposition (SVD) of

$$H_f = \sum_{p=1}^{P} s_p \tilde{w}_{fp}^T, \tag{6}$$

where s_p is the p^{th} columns of S and $\tilde{w}_{fp}$ is that of $\tilde{W}^{(k)}$.

The optimal scale is given by the following formula:

$$q_f = \frac{\sum_{p=1}^{P} \tilde{w}_{fp}^T R_f s_p}{\sum_{p=1}^{P} s_p^T s_p}. \tag{7}$$

4 Tests on synthesized data

Several experiments with synthetic data have been carried out to study the properties of the proposed method. The main examination have been done on synthesized data. Our method is compared with the original Tomasi-Kanade factorization[6] and with the Bundle Adjustment [2]. These methods are denoted by 'TK' and 'BA' in the figures, respectively. The proposed method is denoted by 'FA' which is the abbreviation of Fast Alternation.

In order to support the methods with testing data, trajectories are generated in the following way:

1. Random three-dimensional coordinates are generated by a zero-mean Gaussian random number generator with variance σ_{3D}

Algorithm 1 SfM by Fast Alternation

$M^{(0)}$,$S^{(0)} \leftarrow$ Tomasi-Kanade(W)
$\tilde{W}^{(0)}$, $\tilde{M}^{(0)} \leftarrow$ Complete(W,$M^{(0)}$,$S^{(0)}$)
$k \leftarrow 0$
repeat
  $k \leftarrow k+1$
  $S^{(k)} \leftarrow$ S-Step($\tilde{W}^{(k-1)}$,$\tilde{M}^{(k-1)}$)
  $\tilde{W}^{(k)} \leftarrow$ Complete(W,$\tilde{M}^{(k-1)}$,$S^{(k)}$)
  $\tilde{M}^{(k)} \leftarrow$ M-Step($\tilde{W}^{(k)}$,$S^{(k)}$)
  $\tilde{W}^{(k)} \leftarrow$ Complete(W,$\tilde{M}^{(k)}$,$S^{(k)}$)
until $\left(\left|\left| \tilde{W}^{(k-1)} - \tilde{M}^{(k-1)} S^{(k-1)} \right|\right|_F^2 - \left|\left| \tilde{W}^{(k)} - M^{(k)} S^{(k)} \right|\right|_F^2 \right) < \epsilon$
$M \leftarrow \tilde{M}^{(k)}$
$S \leftarrow S^{(k)}$

2. The generated 3D points are rotated several times by random angles.

3. Rotated points are projected into the image plane by weak perspective projection.

4. Noise is added to the projected coordinates. It is generated by a zero-mean Gaussian random number generator as well. Its variance is set to σ_{2D}.

5. Finally, the measurement matrix W is composed with the projected points.

To compare the methods, error values are needed. We know the original 3D structure of the generated objects as well as the camera motion. Then the reconstructed matrices are registered to the original ones. The registration is done by the method of Arun et al. [1] We use the norm of the registration error as error values. The registration error is computed both to the original and reconstructed structure matrices and to the corresponding motion matrices.

The time demand of the two adjustment methods is also compared. We have coded both method in Octave[1)] and run the scripts on an Intel P4 2.4 GHz PC with 512 MByte memory.

General remarks. The characteristics of each method is very similar as it is visualized in figures 1 – 4. It is obvious, that the improved methods ('BA' and 'FA') serve better results than the original Tomasi-Kanade factorization as it is expected. The difference is especially significant when the motion errors are compared.

Usually, the quality of the Fast Alternation is slightly better than that of the BA method, but

[1)] Octave is a Matlab-compatible interpreter. See www.octave.org.

the difference is very small in general. The main advantage of the Fast Alternation algorithm is its high speed: it is always faster than bundle adjustment.

Error versus the noise (Figure 1) The methods are run with gradually increasing noise level. The reprojection error is increasing approximately in a linear way for all the methods. The test sequence consists of 10 points and 5 frames. The noise level is calculated as $100\sigma_{2D}/\sigma_{3D}$.

Error versus number of frames (Figure 4). The plot shows that the error is decreasing with the number of frames. This is because the ratio of the number of the unknown parameters and known measurement matrix elements tends to a constant value w.r.t the number of frames. During the test, the noise level is 5%, and the model consists of 10 points.

Error versus number of points (Figures 2 – 3). We run two steps: in the first one, the number of points was set from 4 to 10, while in the second test from 10 to 300. The results are informative: one can recognize that the adjustment methods are effective for the structure data estimation if the number of points are very small: if the object to be reconstructed consists of more than 50–100 points, the original factorization serves very good structure data. This is because the quality of the factorization's first step (subspace estimation) is very sensitive to the number of points. If the subspace is estimated well, the original method finds the correct solution as well as the adjustment methods. It is also surprising that the camera motion estimation can be significantly adjusted in all cases.

During the test, the noise level is 5%, the sequence consists of 5 frames.

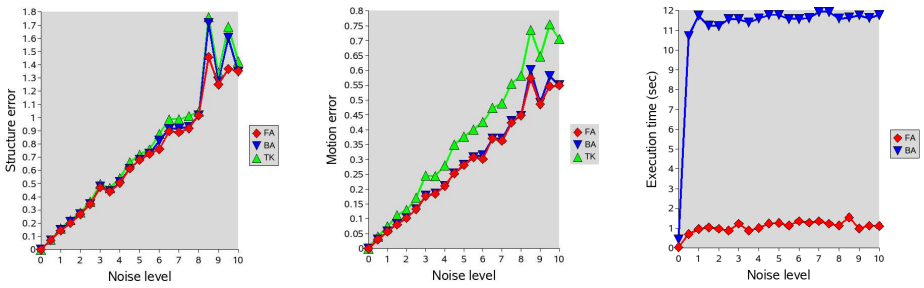

Figure 1: Structure and motion errors and time demand w.r.t. noise level.

5 Tests on real data

We have tested the proposed algorithms on real data as well. The weak-perspective reconstruction of a human face is visualized in fig 5. The input of the reconstructions was the tracked feature points of 3 frames taken of a face.

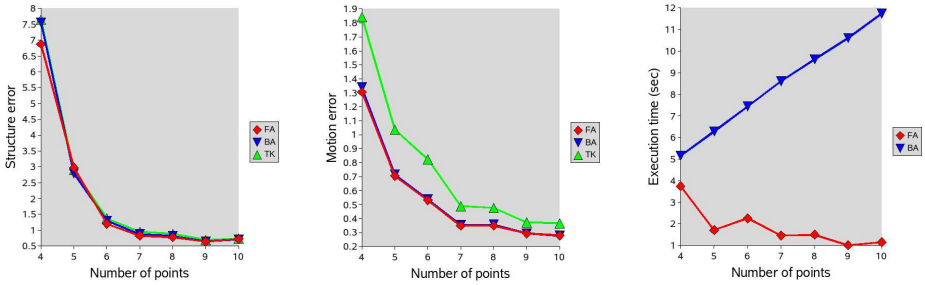

Figure 2: Structure and motion errors and time demand w.r.t. the number of points.

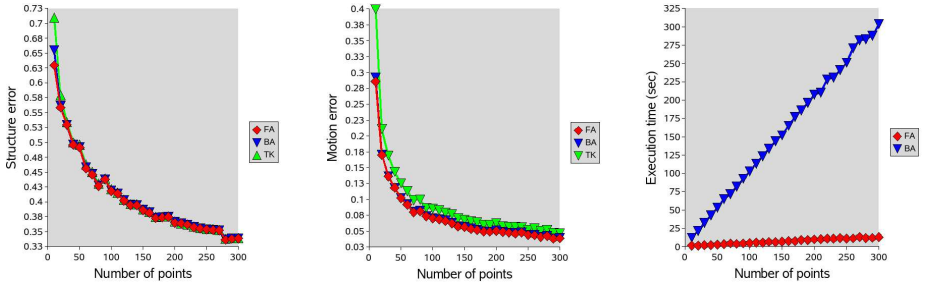

Figure 3: Structure and motion errors and time demand w.r.t. the number of points.

As it is mentioned in the introduction, adjustment methods can also be used for Monte Carlo type robust method where motion data are computed from few points. We tried LMedS SfM method [4] with and without the proposed Fast Alternation. In all cases, the separation of inliers from outliers is better with the usage of the adjustment (see Table 1).

6 Conclusion and future work

We have addressed the problem of weak-perspective structure from motion in this paper. A fast alternation method has been proposed and examined here. It has been diagnosed that our novel algorithm gives better results than the original factorization [6] and it is slightly better and significantly faster than the widely used bundle adjustment. It has been discussed that the adjustment methods are efficient if the 3D model to be reconstructed consists of few points. Therefore, we suggest the usage of the proposed adjustment method for Monte Carlo type robust algorithm (e.g. [4, 7]) where initial reconstruction is computed from $4 \dots 8$ points.

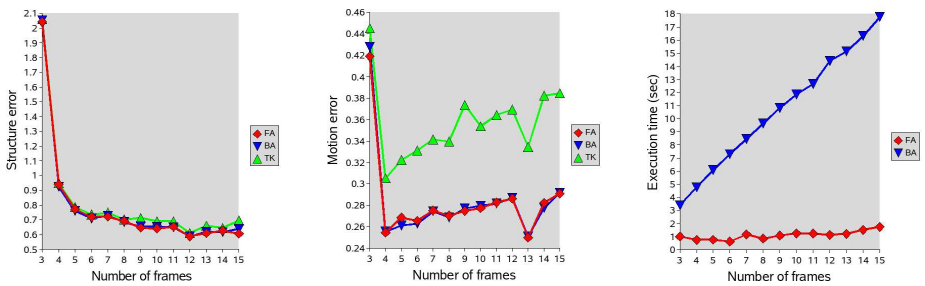

Figure 4: Structure and motion errors and time demand w.r.t. the number of frames.

Noise Level (%)	0	1	2	3	4	5	6
LMedS without Fast Alternation	0	1	52	82	97	133	301
LMedS with Fast Alternation	0	1	23	19	21	122	279

Table 1: Separation errors of the tested robust methods.

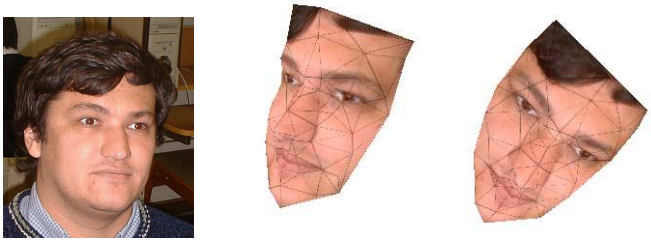

Figure 5: Left: one of original images. Center, right: reconstructed 3D face.

The proposed method can be extended to the perspective case as well as to the missing data problem.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. on PAMI*, 9(5):698–700, 1987.

[2] Bill Triggs and Philip McLauchlan and Richard Hartley and Andrew Fitzgibbon. Bundle Adjustment – A Modern Synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.

[3] A. M. Buchanan. Investigation into matrix factorization when elements are unknown. Technical report, University of Oxford, `http://www.robots.ox.ac.uk/~amb`, 2004.

[4] L. Hajder and D. Chetverikov. Weak-perspective structure from motion for strongly contaminated data. *Pattern Recognition Letters*, 27:1581–1589, 2006.

[5] P. Sturm and B. Triggs. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *ECCV*, volume 2, pages 709–720, 1996.

[6] Tomasi, C. and Kanade, T. Shape and Motion from Image Streams under orthography: A factorization approach. *Intl. Journal Computer Vision*, 9:137–154, 1992.

[7] Trajković, M. and Hedley, M. Robust Recursive Structure and Motion Recovery under Affine Projection. In *Proc. British Machine Vision Conference*, 1997.

[8] D. Weinshall and C. Tomasi. Linear and Incremental Acquisition of Invariant Shape Models From Image Sequences. *IEEE Trans. on PAMI*, 17(5):512–517, 1995.