An Iterative Multiresolution Scheme for SFM

Carme Julià, Angel Sappa, Felipe Lumbreras, Joan Serrat, and Antonio López

Computer Vision Center and Computer Science Department,
Universitat Autònoma de Barcelona,
08193 Bellaterra, Spain
{cjulia, asappa, felipe, joans, antonio}@cvc.uab.es

Abstract. Several factorization techniques have been proposed for tackling the *Structure from Motion* problem. Most of them provide a good solution, while the amount of missing and noisy data is within an acceptable ratio. Focussing on this problem, we propose to use an incremenal multiresolution scheme, with classical factorization techniques. Information recovered following a coarse-to-fine strategy is used for both, filling in the missing entries of the input matrix and denoising original data. An evaluation study, by using two different factorization techniques–the *Alternation* and the *Damped Newton*–is presented for both synthetic data and real video sequences. [1]

1 Introduction

Structure From Motion (SFM) consists in extracting the 3D shape of a scene as well as the camera motion from trajectories of tracked features. Factorization is a method addressing to this problem. The central idea is to express a matrix of trajectories W as the product of two unknown matrices, namely, the 3D object's shape S and the relative camera pose at each frame M: $W_{2f \times p} = M_{2f \times r} S_{r \times p}$, where f, p are the number of frames and feature points respectively and r the rank of W. The goal is to find the factors M and S that minimize

$$\|W - MS\|_F^2 \tag{1}$$

where $\| \cdot \|$ is the Frobenius matrix norm [1].

The *Singular Value Decomposition* (SVD) gives the best solution for this problem when there are not missing entries. Unfortunately, in most of the real cases not all the data points are available, hence other methods need to be used. With missing data, the expression to minimize is the following

$$\|W - MS\|_F^2 = \sum_{i,j} |W_{ij} - (MS)_{ij}|^2 \tag{2}$$

where i and j correspond to the index pairs where W_{ij} is defined.

[1] This work has been supported by the Government of Spain under the CICYT project TRA2004-06702/AUT. The second author has been supported by The Ramón y Cajal Program.

A. Campilho and M. Kamel (Eds.): ICIAR 2006, LNCS 4141, pp. 804–815, 2006.

In the seminal approach Tomasi and Kanade [2] propose an initialization method in which they first decompose the largest full submatrix by the factorization method and then the initial solution grows by one row or by one column at a time, unveiling missing data. The problem is that finding the largest full submatrix is a NP-hard problem. Jacobs [3] treats each column with missing entries as an affine subspace and shows that for every r-tuple of columns the space spanned by all possible completions of them must contain the column space of the completely filled matrix. Unknown entries are recovered by finding the least squares regression onto that subspace. However it is strongly affected by noise on the data. An incremental SVD scheme of incomplete data is proposed by Brand [4]. The main drawback of that technique is that the final result depends on the order in which the data are given. Brandt [5] proposes a different technique based on the expectation maximization algorithm (EM) and although the feature points do not have to be visible in all views, the affine projection matrices in each image must be known. A method for recovering the most reliable imputation, addressing the SFM problem, is provided by Suter and Chen [6]. They propose an iterative algorithm to employ this criterion to the problem of missing data. Their aim to obtain the projection of W onto a low rank matrix to reduce noise and to fill in missing data.

A different approach to address the factorization with missing data is the *Alternation technique* [7]. One of the advantages of this method is that it converges quickly. The algorithm starts with an initial random S_0 or M_0 and solves one factor at each iteration k, until the product $M_k S_k$ converges to W. The key point of this 2-step algorithm is that, since the updates of S given M (analogously in the case of M given S) can be independently done for each row of S, missing entries in W correspond to omitted equations. Due to that fact, with a large amount of missing data the method would fail to converge.

In [7], Buchanan and Fitzgibbon summarize factorization approaches with missing data and proposes the *Alternation/Damped Newton Hybrid*, which combines the *Alternation* strategy with the *Damped Newton* method. The latter is fast in valleys, but not effective when far from the minima. The goal of introducing this hybrid scheme is to give a method that has fast initial convergence and, at the same time, has the power of non-linear optimization.

One disadvantage of the above methods is that. They give a good factorization while the amount of missing points is low, which is not common in real image sequences, unfortunately. Addressing to this problem, we recently presented an iterative multiresolution scheme [8], which incrementally fill in missing data. At the same time noisy data are filtered. The key point of that approach is to work with a reduced set of feature points along a few number of consecutive frames. Thus, the 3D reconstruction corresponding to the selected feature points and the camera motion of the used frames are obtained. The missing entries of the trajectory matrix can be recovered just multiplying the shape and motion matrices. The amount of missing data is reduced after each iteration; at the same time it increases the chances of having a better result. In the current paper we propose improvements over the original approach, as well as the use of two different techniques under this incremental multiresolution scheme.

This paper is organized as follows. Section 2 briefly introduces the incremental multiresolution scheme. Improvements on the original version are emphasized. Section 3 presents an evaluation study of the use of two factorization techniques with the proposed scheme. Conclusions and future work are given in section 4.

2 Iterative Multiresolution Scheme

In this section the iterative multiresolution scheme, which incrementally fill in missing data, is presented. Essentially, the basic idea is to generate sub-matrices with a reduced density of missing points. Thus, any classical factorization technique could be used for factoring these sub-matrices, recovering their corresponding 3D shape and motion; at the same time missed data on the original matrix will be filled with their resulting product. Additionally, it is expected that noisy data are filtered. The technique consists of two stages, which are briefly described below. Improvements on the original version are highlighted, more details of the original technique can be found in [8].

2.1 Observation Matrix Splitting

Let $W_{2f \times p}$ be the observation matrix (also referred through the paper as input matrix) of p feature points tracked over f frames containing missing entries; it will be denoted as W. Let k be the index indicating the current iteration number.

In a first step, the input matrix W is split in a uniformly distributed set of $k \times k$ non-overlapped sub-matrices, W_k, with a size of $\lfloor \frac{2f}{k} \rfloor \times \lfloor \frac{p}{k} \rfloor$.

Then, in a second step, a multiresolution approach is followed. It consists in computing four W_{2k} overlapped sub-matrices with twice the size of W_k (only for $k > 2$). The idea of this enlargement process is to study the behavior of feature points contained in W_k when a bigger region is considered (see Fig. 1).

Since generating four W_{2k}, for every W_k, is a computationally expensive task, a simple and more direct approach is followed. It consists in splitting the input matrix W in four different ways, by shifting W_{2k} half of its size (i.e., W_k) through rows, columns or both at the same time. Fig. 2 illustrates the five partitions of matrix W—i.e., W_k and W_{2k} sub-matrices generated at the sixth iteration. When all these matrices are considered together, the overlap between the different areas is obtained, see textured cell in Fig. 1 and Fig. 2. As it can be appreciated in Fig. 2, corners and border cells are considered only twice and three times, respectively, at each iteration.

2.2 Sub-matrices Processing

At this stage, the objective is to recover missing data by applying a factorization technique at every single sub-matrix. Independently of their size hereinafter sub-matrices will be referred as W_s.

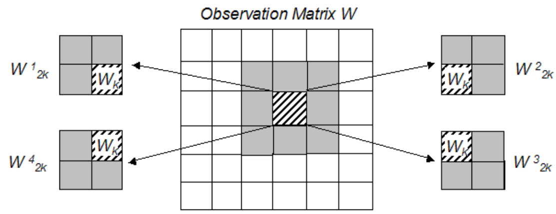

Fig. 1. W_k and W_{2k} overlapped matrices of the observation matrix W, computed during the first stage (section 2.1), at iteration $k = 6$

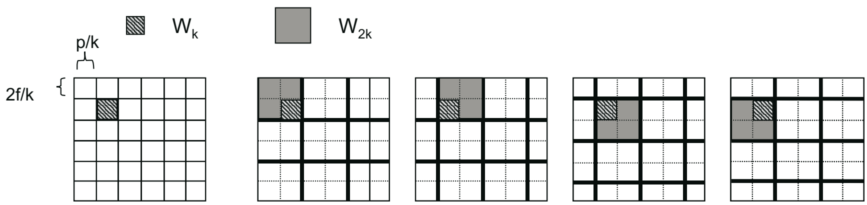

Fig. 2. Five partitions of matrix W. Note the overlap between a W_k sub-matrix with its corresponding four W_{2k} sub-matrices, computed during the first stage (section 2.1)

Given a sub-matrix W_s, its factorization using a particular technique gives its corresponding M_s and S_s matrices. Their product could be used for computing an approximation error ε_s such as equation (2). Actually, in this paper we use the root mean squared (rms) of this error per image point:

$$rms_s = \sqrt{\frac{\sum_{i,j} |(W_s)_{ij} - (M_s S_s)_{ij}|^2}{\frac{n}{2}}} \tag{3}$$

where i and j correspond to the index pairs where $(W_s)_{ij}$ is defined and n the amount of those points in W_s.

The main advantage of using rms is that it does not depend on the number of known entries. Therefore, it is a normalized measure that gives a better idea of the goodness of the solution than the error defined at equation (2), which could be confusing or provide erroneous results. For instance, by using the latter, a big error value could be obtained only due to the fact of working with a great amount of known entries. On the contrary, a small error value could correspond to a case with a few known entries.

After processing the current W_s, its corresponding rms_s is compared with a user defined threshold σ. In case the resulting rms_s is smaller than σ, every point in W_s is kept in order to be merged with overlapped values after finishing the current iteration. Additionally, every point of W_s is associated with a weighting factor, defined as $\frac{1}{rms_s}$, in order to measure the goodness of that value. These weighting factors are later on used for merging data on overlapped areas. Otherwise, the resulting rms_s is higher than σ, computed data are discarded. With the rms_s error measure, a unique threshold, valid for every matrix, is defined. As mentioned above, this is the main advantage over the previous version.

Finally, when every sub-matrix W_s has been processed, recovered missing data are used for filling in the input matrix W. In case a missing datum has been recovered from more than one sub-matrix (overlapped regions), those recovered data are merged by using their corresponding normalized weighting factors. On the contrary, when a missing datum has been recovered from only one sub-matrix, this value is directly used for filling in that position. Already known entries in W could be also modified. In this case, instead of taking the new computed data directly as in [8], the mean between the initial value and the new one, obtained after the merging process, is assigned.

Once recovered missing data have been used for filling in the input matrix W, the iterative process starts again (section 2.1) splitting the new matrix W either by increasing k one unit or, in case the size of sub-matrices W_k at the new iteration stage is quite small, by setting $k = 2$. This iterative process is applied until one of the following conditions is true: a) the matrix of trajectories is totally filled; b) at the current iteration no missing data were recovered; c) a maximum number of iterations is reached.

3 Evaluation Study

Assuming both the filling missing entries and denoising capabilities, as it was presented in [8], in this section a study of using the iterative multiresolution scheme with different factorization techniques is presented. In particular, the work is focussed on the use of: Alternation and Damped Newton [7]. Experiments using both synthetic and real data are presented below. The methodology proposed to evaluate the obtained results consists in applying:

- A factorization technique over the input matrix W.
- The same factorization technique with the proposed multiresolution scheme.

3.1 Synthetic Object

Synthetic data are randomly generated by distributing 35 3D feature points over the whole surface of a cylinder, see Fig. 3 (left). The cylinder is defined by a radius of 100 and a height of 400; it rotates over its principal axis; the camera also moves. An input matrix W with 20% of known data, directly obtained taking 35 frames, is used for the evaluation, see Fig. 3 (middle). Notice that W has a banded structure; it is symmetric due to the way it has been generated. Elements of the matrix W are represented by means of a grey level scale. Feature point trajectories are plotted in Fig. 3 (right).

Factorization Using Alternation. Fig. 4 (left) shows the recovered trajectories obtained by applying the Alternation technique to the input matrix W for this synthetic example. In this case, the resulting rms is 6.43.

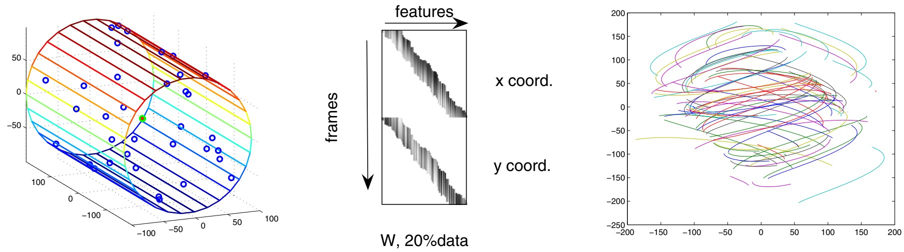

Fig. 3. (left) Synthetic cylinder. (middle) Input matrix of trajectories, with 20% of known data. (right) The same feature point trajectories represented in the image plane.

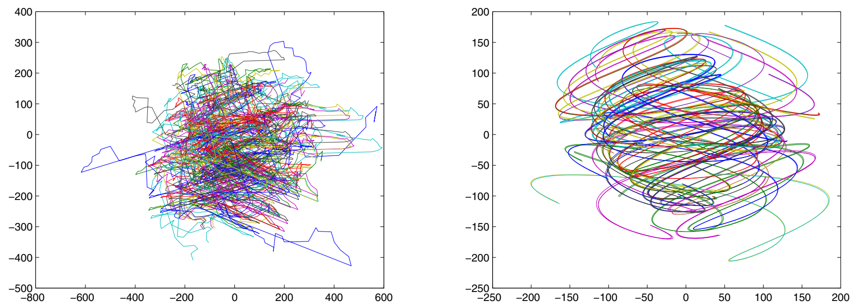

Fig. 4. (left) Recovered feature points trajectories applying Alternation to W. (right) The same, but applying Alternation with the iterative mutiresolution scheme.

Fig. 5 shows intermediate results obtained after applying the Alternation technique with the multiresolution scheme to W. In order to illustrate the process, the amount of recovered data at the third, fifth and last iterations of the multiresolution scheme are presented. While the input matrix has 20% of data, the final one has 63% of data. The trajectories plotted in Fig. 4 (right) are obtained after applying the Alternation technique to this final matrix (Fig. 5 (right)). The rms value is 0.29. Notice that these trajectories do not form a cylinder, as one might expected, due to the fact that the camera moves.

Fig. 6 (left) and (middle) shows the recovered $\mathbf{x}$ and $\mathbf{y}$ camera motion axes, from both strategies and at each frame. The 3D plot corresponds to each component of the vector axes: $\mathbf{x}(x, y, z)$, and $\mathbf{y}(x, y, z)$ at each frame. The obtained 3D reconstructions of the cylinder are plotted in Fig. 6 (right). In order to show the goodness of the reconstruction, the cylinder that best fits the final data is also plotted. Fig. 6 (top) corresponds to results computed by applying Alterantion to W, while Fig. 6 (bottom) presents the results obtained by applying Alternation with the proposed multiresolution scheme.

It can be seen that the results are considerably improved with the proposed multiresolution scheme, both for the recovered feature points trajectories and for the obtained shape and motion.

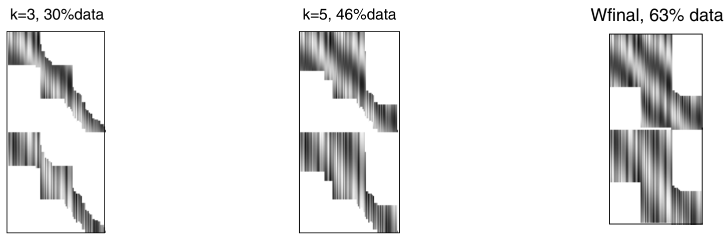

Fig. 5. Third, fifth and last iterations of the multiresolution scheme when the Alternation technique is used. A final matrix with 63% of data is obtained.

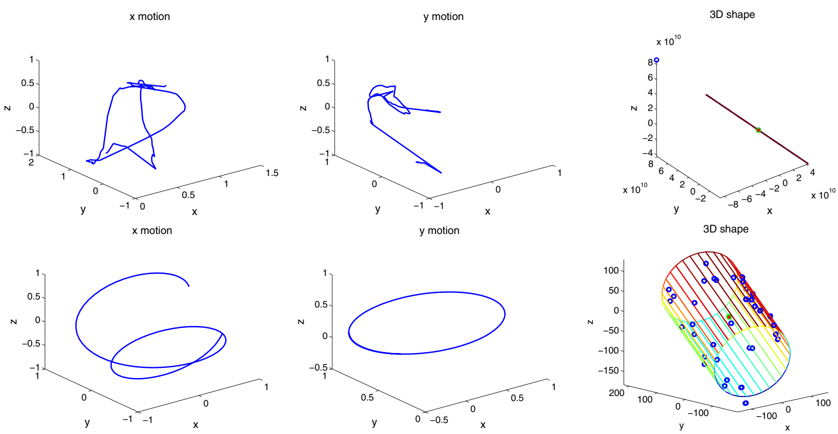

Fig. 6. (left) and (middle) Plots of the recovered **x** and **y** camera motion axes, at each frame. (right) 3D reconstructions of the cylinder. (top) Applying Alternation to W. (bottom) Applying Alternation with the multiresolution scheme.

Factorization Using Damped Newton. The observation matrix W presented in Fig. 3 (middle) has been also used as input for the Damped Newton factorization technique. Fig. 7 (left) plots the recovered trajectories applying the Damped Newton to W. The rms is 0.76.

Fig. 7 (right) shows the obtained trajectories when Damped Newton is applied with the proposed multiresolution scheme; in this case the rms is 22.22. As pointed out in [6], the measure of error taking only the known entries of W could be ambiguous. In this particular case, since this is a synthetic sequence, the missing entries are available. Hence, the rms taking into account all those entries has been computed for both results. Their values are 77.0 applying Damped Newton to W and 69.1, with the multiresolution scheme. Again, better results, not only visual but also numerical, are obtained with the multiresolution scheme.

The recovered **x** and **y** camera motion axes from both strategies are plotted in Fig. 8 (left) and (middle) respectively. In Fig. 8 (right) the obtained 3D reconstructions of the cylinder are shown.

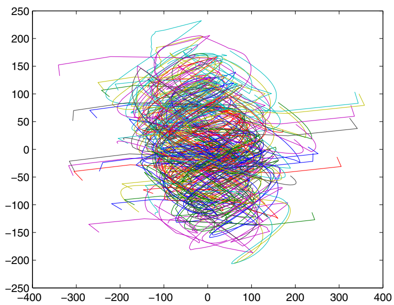

Fig. 7. (left) Recovered trajectories applying Damped Newton to W. (right) The same, but applying Damped Newton with the proposed scheme.

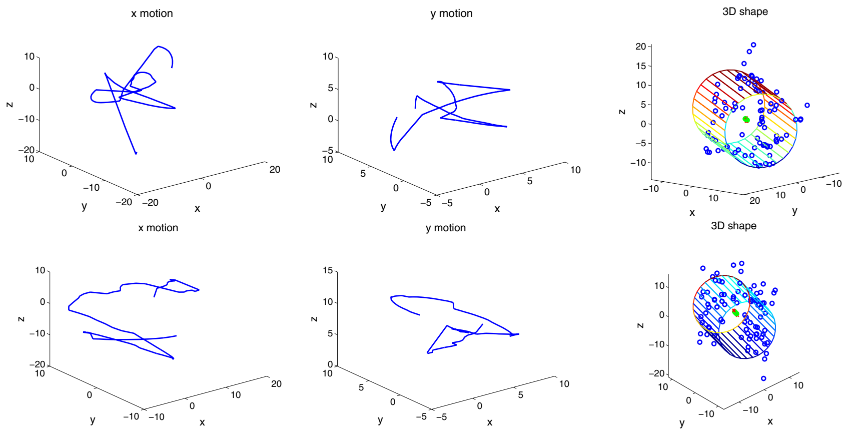

Fig. 8. (left) and (middle) Plots of the recovered **x** and **y** camera motion axes, at each frame. (right) 3D reconstructions of the cylinder. (top) Applying Damped Newton to W. (bottom) Applying Damped Newton with the multiresolution scheme.

The results obtained with the Damped Newton are not as good as the ones obtained with the Alternation, both applying it to W and with the proposed multiresolution scheme. Notice that, inspite of that, results with the proposed scheme are better than using Damped Newton directly over W.

3.2 Real Object

A real video sequence of 101 frames with a resolution of 640×480 pixels is used. The studied object is shown in Fig. 9 (left). A single rotation around a vertical axis was performed. Feature points are selected by means of a corner detector algorithm and 87 points distributed over the squared-face-box are considered. An iterative feature tracking algorithm has been used. More details about corner detection and tracking algorithm can be found in [9]. Missing data are obtained by removing data randomly. As in the previous case, an input matrix with 20%

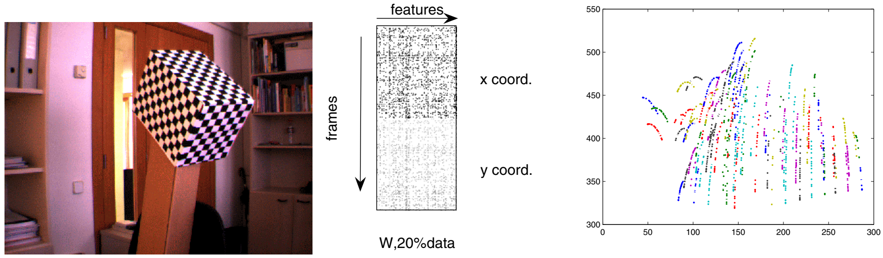

Fig. 9. (left) Object used for the real case. (middle) Input matrix of trajectories, with 20% of known data. (right) Feature point trajectories represented in the image plane.

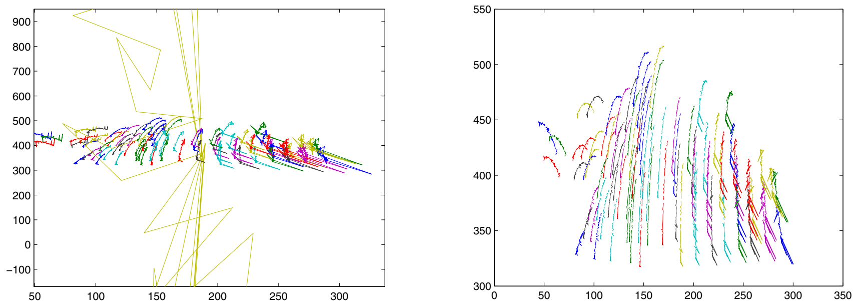

Fig. 10. (left) Recovered feature points trajectories applying Alternation to W, Fig. 9 (middle). (right) The same, but applying Alternation with the proposed scheme.

of data is considered, see Fig. 9 (middle). Notice that the input matrix W has not a band structure like in the synthetic case, but a structure that unveil the random nature of missing entries. The feature point trajectories are plotted in Fig. 9 (right). The methodology applied over the previous W is presented below.

Factorization Using Alternation. The Alternation technique is applied to the input matrix W and the obtained rms is 2.41. Fig. 10 (left) plots an enlargement of the recovered trajectories, in order to avoid a plot such as Fig. 11 (top-right).

On the contrary, when Alternation is applied with the proposed multiresolution scheme, results improve considerably. For instance, about 95% of data are contained in matrix W during last iteration (recall that the input matrix only contains about 20%), and the final resulting rms is 0.69. Fig. 10 (right) shows the trajectories recovered with the proposed scheme.

Fig. 11 (left) and (middle) show the recovered **x** and **y** camera motion axes, from both strategies. The obtained 3D reconstructions of the input object are presented in Fig. 11 (right). As in the synthetic experiment, results are improved applying Alternation with the proposed iterative multiresolution scheme.

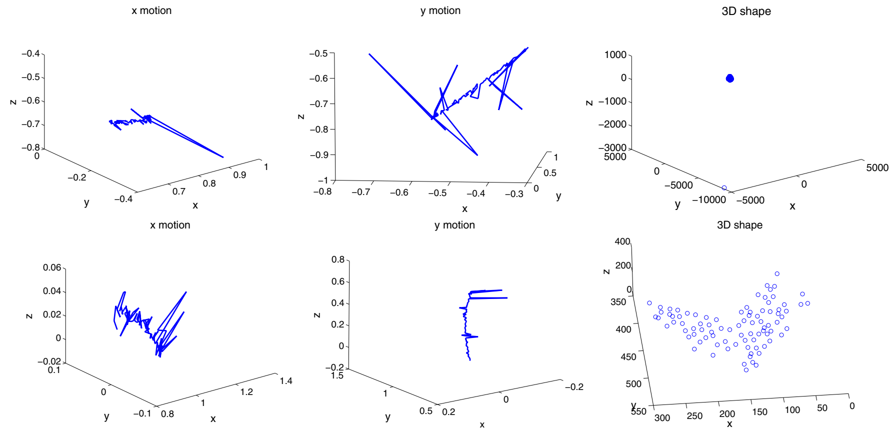

Fig. 11. (left) and (middle) Plot of the recovered **x** and **y** camera motion axes, at each frame. (right) 3D reconstruction of the input object. (top) Results obtained applying Alternation to W. (bottom) Applying Alternation with the proposed scheme.

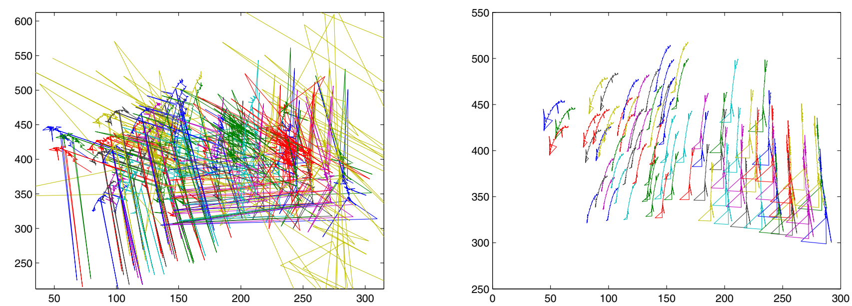

Fig. 12. (left) Recovered trajectories applying Damped Newton to W, Fig. 9 (middle). (right) The same, but applying Damped Newton with the proposed scheme.

Factorization Using Damped Newton. The recovered trajectories applying Damped Newton are plotted in Fig. 12 (left) and the *rms* is 0.87. Again, an enlargement has been performed in order to obtain a better visualization.

Fig. 12 (right) shows the obtained trajectories applying Damped Newton with the proposed multiresolution scheme. In this case, the *rms* is 7.38. Since the missing points have been obtained randomly from a full matrix, all the entries are available. Therefore, as in the synthetic case, all the points have been taken into account for computing the *rms*. The resulting rms errors are 29.54 applying Damped Newton to W and 8.15 with the multiresolution scheme.

Finally, the recovered **x** and **y** camera motion axes, from both strategies, are plotted in Fig. 13 (left) and (middle), respectively. The obtained 3D reconstructions of the input object are shown in Fig. 13 (right).

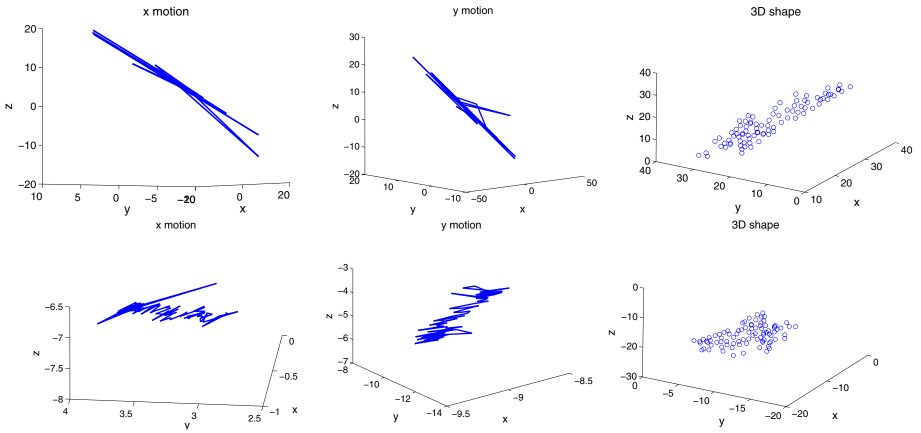

Fig. 13. (left) and (middle) Plots of the recovered $\mathbf{x}$ and $\mathbf{y}$ camera motion axes, at each frame. (right) 3D reconstructions of the input object. (top) Applying Damped Newton to W. (bottom) Applying Damped Newton with the multiresolution scheme.

As in the synthetic case, the obtained results using Damped Newton are worse than the ones obtained with Alternation. Again, results with the proposed scheme are better than the ones obtained using Damped Newton directly over W.

4 Conclusions and Future Work

This paper presents improvements on the original iterative multiresolution approach. The main contribution is the modified definition of the error. Here, the root mean of the previous error per image point is considered. As mentioned before, this is a normalized measure that gives a better idea of the goodness of the result. A less important improvement is the way in which the recovered missing data are merged in the case of data known a priori. The average between the input entries and the new computed data obtained after the merging process is assigned. Moreover, in the current paper, the attention is not only focused on the error value, but also on the obtained M and S. However, we would like to define a function to measure the goodness of these recovered factors.

Additionally, when the original multiresolution scheme was presented, its validation was done using only the Alternation technique. In this work, the Damped Newton technique is also studied and compared with the Alternation. Although any factorization technique can be applied with this multiresolution scheme, the Damped Newton method seems to be more appropriated when the input matrix contains only a few missing data. Otherwise it takes a lot of time and may be wrong results are obtained. As a future work the use of an hybrid technique will be considered.

Acknowledgments

The authors would like to thanks Aeron Buchanan for providing us with the Damped Newton code and interesting insights.

References

1. Golub, G., Van Loan, C., eds.: Matrix Computations. The Johns Hopkins Univ. Press (1989)
2. Tomasi, C., Kanade, T.: Shape and motion from image streams: a factorization method. Full report on the orthographic case (1992)
3. Jacobs, D.: Linear fitting with missing data for structure-from-motion. Computer vision and image understanding, CVIU (2001) 7–81
4. Brand, M.: Incremental singular value decomposition of uncertain data with missing values. In: Proceedings, ECCV. (2002) 707–720
5. Brandt, S.: Closed-form solutions for affine reconstruction under missing data. In: Proceedings Statistical Methods for Video Processing Workshop, in conjunction with ECCV. (2002) 109–114
6. Chen, P., Suter, D.: Recovering the missing components in a large noisy low-rank matrix: Application to sfm. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004)
7. Buchanan, A., Fitzgibbon, A.: Damped newton algorithms for matrix factorization with missing data. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) **2** (2005) 316–322
8. Julià, C., Sappa, A., Lumbreras, F., Serrat, J., López, A.: Factorization with missing and noisy data. In: Computational Science–ICCS 2006: 6th International Conference, Proceedings, Part I. (2006) 555 – 562
9. Ma, Y., Soatto, J., Koseck, J., Sastry, S.: An invitation to 3d vision: From images to geometric models. Springer-Verlang, New York (2004)