Rank estimation in 3D multibody motion segmentation

C. Julià, A.D. Sappa, F. Lumbreras, J. Serrat and A. López

A novel technique for rank estimation in 3D multibody motion segmentation is proposed. It is based on the study of the frequency spectra of moving rigid objects and does not use or assume a prior knowledge of the objects contained in the scene (i.e. number of objects and motion). The significance of rank estimation on multibody motion segmentation results is shown by using two motion segmentation algorithms over both synthetic and real data.

Introduction: Given a matrix of trajectories $W_{2f\times p}$, where f and p are the number of frames and tracked feature points, respectively, the goal of motion segmentation approaches is to cluster feature points into different motion subspaces (i.e. features belonging to the same object). Some of the proposed techniques are formulated under the framework of factorisation methods, where the knowledge of the rank of W becomes an essential piece of information for solving the problem. In most of the cases it is assumed that tracked feature points are visible along the whole video sequence, giving rise to a full matrix of trajectories. However, trajectories are often incomplete or split owing to objects occlusions, missing on the tracking or simply because they exit the camera field of view. In this case two different problems are faced: first, the unknown entries in the matrix of trajectories should be filled in using some imputation technique, and, secondly, once W has been filled in, feature trajectories corresponding to the same object should be clustered together.

Imputation techniques allow recovery of missing entries by imposing a predefined rank value or assuming some prior knowledge about the number of objects contained in the scene. For instance, [1] proposes an approach able to deal with missing data in the multibody case. It consists in fixing the rank of W to five and applying a factorisation technique to project the point trajectories from R^{2f} to R^5. The problem, as is presented in this work, is that motion segmentation results are sensitive to the assumed rank value. In this Letter, a novel technique for estimating the rank value of a missing data trajectory matrix, in 3D multibody motion segmentation, is proposed. Two different multibody motion segmentation algorithms are used to show the importance of a proper rank estimation.

Rank estimation: A direct technique to estimate the rank (r) of a matrix of trajectories with missing data is presented, instead of assuming a predefined value or a prior knowledge of the scene. The intuition behind the proposed approach is that, since feature points belong to surfaces of rigid objects, the frequency content of W should be preserved after filling in missed entries. In other words, the filled matrix should contain a frequency spectra similar to the one of the input matrix. It works as follows:

1. Select a range of possible ranks (see '*Note*' below), $r \in \{r_0, r_1, \ldots, r_m\}$.
2. For each rank, fill in the given input matrix by using an imputation technique (e.g. the *alternation* technique [2]).
3. Compute the error:

$$e = \|F - F_r\|_F \tag{1}$$

where F is the modulus of the fast Fourier transform (FFT) of the initial matrix W and F_r is the modulus of the FFT of the filled matrix corresponding to the current r.

4. Repeat until a local minimum of (1) is found.

Note: In the implemented version the lowest bound of this range is set to five ($r_0 = 5$) and the highest bound is automatically defined by the first local minima in the error value (1).

Multibody motion segmentation algorithms: Once a full trajectory matrix is computed, feature points belonging to the same object are clustered together. In this Letter, the code of two 3D motion segmentation algorithms [3], publicly available at http://www.vision.jhu.edu/db/ have been used: (a) generalised principal component analysis (GPCA) [1]; and (b) local subspace affinity (LSA) [4]. As mentioned above, [1] proposes to apply a factorisation technique to project the point trajectories from R^{2f} to R^5, independently of the number of objects; then GPCA is used to cluster trajectories. In addition to GPCA, [3] proposes to use LSA but by assuming two ranks; one choice is to fix r = 5, which at the same time is the dimension used by GPCA. The other choice is r = 4n, which implicitly assumes both the number of objects contained in the scene is given (n) and their motions are independent. In the experiments presented in the following Section, four different variants are considered to compute the multibody motion segmentation, when the given matrix contains missing data. GPCA (r = 5) is applied to the data projected into R^5. LSA is applied to a filled in matrix, obtained with an imputation technique (the alternation technique has been used) assuming different rank values: (a) r = 5; (b) r = 4n; and (c) $r = r_{FFT}$ (where r_{FFT} is the value estimated with the proposed technique).

Experimental results: Synthetic and real data are used to empirically validate the proposed rank estimation approach. Additionally, a study of the robustness of motion segmentation results with respect to rank values is performed by using two public motion segmentation algorithms.

From a given full matrix, missing data are automatically generated by removing parts of random columns, simulating the behaviour of tracked features. Considering different percentages of missing data, five attempts are repeated and the percentage of bad-clustered features over the total of features in W gives a measure of error in the clustering.

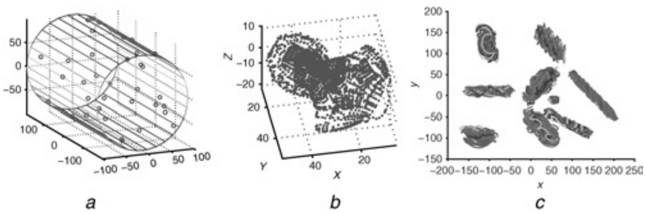

Fig. 1 *Synthetic objects*

a Cylinder
b Beethoven bust
c Feature point trajectories of sequence with nine objects

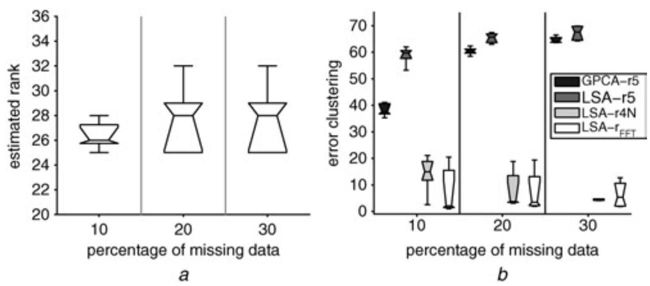

Fig. 2 *Sequence with nine synthetic objects*

a Rank values estimated with proposed approach for different percentages of missing data
b Error obtained in clustering considering different percentages of missing data and for different variants of LSA ($r = r_{FFT}$, r = 4n and r = 5) and GPCA

Synthetic data sets are generated by randomly distributing 3D feature points over the surfaces of a cylinder (Fig. 1*a*) and a Beethoven bust (Fig. 1*b*). Taking these two objects, different sequences are obtained by performing rotations and translations over both of them. At the same time, the camera also rotates and translates. Different numbers of cylinders and Beethoven busts and different motions are considered in order to generate sequences with several objects. Owing to lack of space, only a sequence of nine objects (four cylinders and five Beethoven busts) is presented in the synthetic data experiments. Obtained feature point trajectories are plotted in Fig. 1*c*. The rank of the initially full matrix is 25. Fig. 2*a* shows the rank values obtained with the proposed approach. It can be seen that the rank is quite well estimated even with 30% of missing data, which is a very remarkable performance considering the large number of objects contained in the scene. The error obtained in the clustering with the different studied variants is plotted in Fig. 2*b*. It can be seen that LSA ($r = r_{FFT}$ and r = 4n) gives smaller error than GPCA, for any percentage of missing data. In the case of LSA, the error obtained when the rank is estimated with the proposed technique ($r = r_{FFT}$) is in general similar to r = 4n. However, it should be highlighted that with the proposed approach, no prior knowledge about the number of objects contained in the scene, nor about their motion, is used. The error obtained when the rank is fixed to 5 is very high.

a b

Fig. 3 *First frame of two sequences (feature points are marked)*

a Checkerboard sequence example
b Traffic sequence example

From the benchmark presented in [3], two sets of video sequences, containing three motions, have been considered: 13 video sequences from checkerboard sequences; and five video sequences from traffic sequences. Fig. 3 shows first frames of two sequences, one from each set. The estimated rank values obtained with the proposed approach are plotted in Fig. 4*a*. As can be appreciated, the rank takes different values. This is because the motion and shape are different in each sequence. Therefore, it is expected that the results will be more accurate if the rank is estimated than if it is set to 5 or 4n (12 in this case) for every sequence. Fig. 4*b* shows the error obtained in the clustering with different variants. It can be seen that, as in the synthetic experiment, LSA ($r = r_{FFT}$ and $r = 4n$) gives smaller error than GPCA, as was pointed out in [3]. Focusing on the LSA, error values considering ($r = r_{FFT}$) or setting ($r = 4n$) are similar. However, if the rank is set to 5, the error obtained in the clustering is very high.

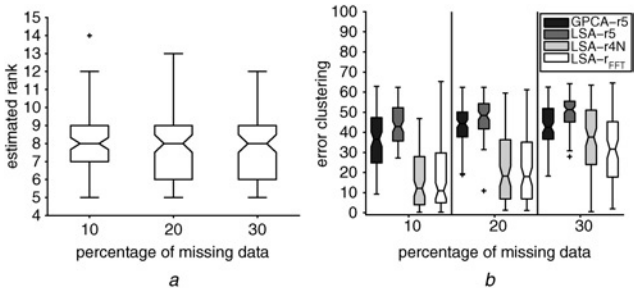

Fig. 4 *Real sequences with three objects*

a Rank values estimated with proposed approach for different percentages of missing data and for every tested sequence
b Error obtained in clustering considering different percentages of missing data and for different variants of LSA ($r = r_{FFT}$, $r = 4n$, $r = 5$) and GPCA

Conclusion: This Letter presents a novel technique for estimating the rank of a multibody trajectory matrix, which contains missing data, as well as an empirical study about the relationship between motion segmentation results and estimated or assumed rank values. It is shown that the error obtained in the clustering when the rank is estimated with the proposed approach is in general smaller than that obtained when the rank is set beforehand. Moreover, it should be noticed that with the proposed approach no prior knowledge about the scene is used (such as number of objects in the scene or kind of motion).

Acknowledgments: The authors thank R. Vidal for providing them with the datasets and code. This work was partially supported by the Government of Spain under the MEC project TRA2007-62526/AUT and research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018). A.D. Sappa was supported by The Ramón y Cajal Program.

17 September 2007
Electronics Letters online no: 20082503
doi: 10.1049/el:20082503

C. Julià, A.D. Sappa, F. Lumbreras, J. Serrat and A. López (*Computer Vision Center, Campus UAB, Bellaterra 08193, Barcelona, Spain*)

E-mail: cjulia@cvc.uab.es

F. Lumbreras, J. Serrat and A. López: Also with the Dep. Ciències de la Computaciò-UAB, Campus UAB, Bellaterra 08193 Barcelona, Spain

References

1 Vidal, R., and Hartley, R.: 'Motion segmentation with missing data using Powerfactorisation and GPCA'. IEEE Int. Conf. on Computer Vision and Pattern Recognition, Washington, DC, USA, June 2004, Vol. II, pp. 310–316

2 Buchanan, A., and Fitzgibbon, A.W.: 'Damped Newton algorithms for matrix factorisation with missing data'. IEEE Int. Conf. on Computer Vision and Pattern Recognition, San Diego, CA, USA, June 2005, Vol. II, pp. 316–322

3 Tron, R., and Vidal, R.: 'A benchmark for the comparison of 3-D motion segmentation algorithms'. IEEE Int. Conf. on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, June 2007, pp. 1–8

4 Yan, J., and Pollefeys, M.: 'A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate and non-degenerate'. European Conf. on Computer Vision, Austria, May 2006, pp. 94–106